VOLUME 5

HANDBOOK OF ALGEBRA

M. HAZEWINKEL

EDITOR

N·H

NORTH-HOLLAND

HANDBOOK OF ALGEBRA
VOLUME 5

HANDBOOK OF ALGEBRA

Volume 5

edited by

M. HAZEWINKEL

CWI, Amsterdam

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

North-Holland is an imprint of Elsevier

N·H

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

First edition 2008

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-444-53101-8
ISSN: 1570-7954

Printed and bound in Hungary

08 09 10 11 10 9 8 7 6 5 4 3 2 1

Preface

Basic philosophy

Algebra, as we know it today (2007), consists of a great many ideas, concepts and results. And this was also the case in 1995 when this Handbook started (which does not mean that nothing has happened in those 12 years; on the contrary, the field of algebra and its applications has developed at a furious pace.)

A reasonable estimate of the number of all the various different concepts, ideas, definitions, constructions, results, ... would be somewhere between 50 000 and 200 000. Many of these have been named and many more could (and perhaps should) have a "name", or other convenient designation. Even a nonspecialist is quite likely to encounter most of these, either somewhere in the published literature in the form of an idea, definition, theorem, algorithm, ... somewhere, or to hear about them, often in somewhat vague terms, and to feel the need for more information. In such a case, if the concept relates to algebra, then one should be able to find something in this handbook; at least enough to judge whether it is worth the trouble to try to find out more. In addition to the primary information the numerous references to important articles, books, or lecture notes should help the reader find out as much as desired.

As a further tool the index is perhaps more extensive than usual, and is definitely not limited to definitions, (famous) named theorems and the like.

For the purposes of this Handbook, "algebra" is more or less defined as the union of the following areas of the Mathematics Subject Classification Scheme:

- 20 (Group theory)
- 19 (K-theory; this will be treated at an intermediate level; a separate Handbook of K-theory which goes into far more detail than the section planned for this Handbook of Algebra is under consideration)
- 18 (Category theory and homological algebra; including some of the uses of categories in computer science, often classified somewhere in section 68)
- 17 (Nonassociative rings and algebras; especially Lie algebras)
- 16 (Associative rings and algebras)
- 15 (Linear and multilinear algebra. Matrix theory)
- 13 (Commutative rings and algebras; here there is a fine line to tread between commutative algebras and algebraic geometry; algebraic geometry is definitely not a topic that will be dealt with in any detail in this Handbook; there will, hopefully, one day be a separate Handbook on that topic)
- 12 (Field theory and polynomials)
- 11 (Number theory, the part that also used to be classified under 12 (Algebraic number theory))

- 08 (General algebraic systems)
- 06 (Order, lattices, ordered algebraic structures; certain parts; but not topics specific to Boolean algebras as there is a separate three-volume Handbook of Boolean Algebras)

Planning

Originally (1992), we expected to cover the whole field in a systematic way. Volume 1 would be devoted to what is now called Section 1 (see below), Volume 2 to Section 2, and so on. A quite detailed and comprehensive plan was made in terms of topics that needed to be covered and authors to be invited. That turned out to be an inefficient approach. Different authors have different priorities and to wait for the last contribution to a volume, as planned originally, would have resulted in long delays. Instead there is now a dynamic evolving plan. This also permits to take new developments into account.

Chapters are still by invitation only according to the then current version of the plan, but the various chapters are published as they arrive, allowing for faster publication. Thus in this Volume 5 of the Handbook of Algebra the reader will find contributions from 5 sections.

As the plan is dynamic, suggestions from users, both as to topics that could or should be covered, and authors, are most welcome and will be given serious consideration by the board and editor.

The list of sections looks as follows:

Section 1: Linear algebra. Fields. Algebraic number theory
Section 2: Category theory. Homological and homotopical algebra. Methods from logic (algebraic model theory)
Section 3: Commutative and associative rings and algebras
Section 4: Other algebraic structures. Nonassociative rings and algebras. Commutative and associative rings and algebras with extra structure
Section 5: Groups and semigroups
Section 6: Representations and invariant theory
Section 7: Machine computation. Algorithms. Tables
Section 8: Applied algebra
Section 9: History of algebra

For the detailed plan (2007 version), the reader is referred to the Outline of the Series following this preface.

The individual chapters

It is not the intention that the handbook as a whole can also be a substitute undergraduate or even graduate, textbook. Indeed, the treatments of the various topics will be much too dense and professional for that. Basically, the level should be graduate and up, and such material as can be found in P.M. Cohn’s three volume textbook ‘Algebra’ (Wiley) should, as a rule, be assumed known. The most important function of the chapters in this Handbook is to provide professional mathematicians working in a different area with a sufficiency of information on the topic in question if and when it is needed.

Each of the chapters combines some of the features of both a graduate level textbook and a research-level survey. Not all of the ingredients mentioned below will be appropriate in each case, but authors have been asked to include the following:

- Introduction (including motivation and historical remarks)
- Outline of the chapter
- Basic concepts, definitions, and results. (These may be accompanied by proofs or (usually better) ideas/sketches of the proofs when space permits)
- Comments on the relevance of the results, relations to other results, and applications
- Review of the relevant literature; possibly complete with the opinions of the author on recent developments and future directions
- Extensive bibliography (several hundred items will not be exceptional)

The present

Volume 1 appeared in 1995 (copyright 1996), Volume 2 in 2000, Volume 3 in 2003, Volume 4 in 2005 (copyright 2006). Volume 6 is planned for 2008. Thereafter, we aim at one volume every two years (or better).

The future

Of course, ideally, a comprehensive series of books like this should be interactive and have a hypertext structure to make finding material and navigation through it immediate and intuitive. It should also incorporate the various algorithms in implemented form as well as permit a certain amount of dialogue with the reader. Plans for such an interactive, hypertext, CDROM (DVD)-based (or web-based) version certainly exist but the realization is still a non-trivial number of years in the future.

Kvoseliai, August 2007 Michiel Hazewinkel

Kaum nennt man die Dinge beim richtigen Namen
so verlieren sie ihren gefährlichen Zauber

(You have but to know an object by its proper name
for it to lose its dangerous magic)

Elias Canetti

This page intentionally left blank

Outline of the Series

(as of June 2007)

Philosophy and principles of the Handbook of Algebra

Compared to the outline in Volume 1 this version differs in several aspects.

First there is a major shift in emphasis away from completeness as far as the more elementary material is concerned and towards more emphasis on recent developments and active areas. Second the plan is now more dynamic in that there is no longer a fixed list of topics to be covered, determined long in advance. Instead there is a more flexible nonrigid list that can and does change in response to new developments and availability of authors.

The new policy, starting with Volume 2, is to work with a dynamic list of topics that should be covered, to arrange these in sections and larger groups according to the major divisions into which algebra falls, and to publish collections of contributions (i.e. chapters) as they become available from the invited authors.

The coding below is by style and is as follows.

- **Author(s) in bold**, followed by chapter title: articles (chapters) that have been received and are published or are being published in this volume.
- *Chapter* title in *italic*: chapters that are being written.
- Chapter title in plain text: topics that should be covered but for which no author has yet been definitely contracted.

Chapters that are included in Volumes 1–5 have a (x; yy pp.) after them, where ‘x’ is the volume number and ‘yy’ is the number of pages.

Compared to the plan that appeared in Volume 1 the section on “Representation and invariant theory” has been thoroughly revised from Volume 2 on.

Compared to the plan that appeared in Volume 4, Section 4H (Rings and algebras with additional structure) has been split into two parts: 4H (Hopf algebras and related structures) and 4I (Other rings and algebras with additional structure). The old Section 4I (Witt vectors) has been absorbed into the section on Hopf algebras.

There also a few more changes; mostly addition of some more topics.

Editorial set-up

Managing editor: M Hazewinkel.

Editorial board: M. Artin, M. Nagata, C. Procesi, O. Tausky-Todd†, R.G. Swan, P.M. Cohn, A. Dress, J. Tits, N.J.A. Sloane, C. Faith, S.I. A’dyan, Y. Ihara, L. Small, E. Manes, I.G. Macdonald, M. Marcus, L.A. Bokut’, Eliezer (Louis Halle) Rowen, John S. Wilson, Vlastimil Dlab. Note that three editors have been added starting with Volume 5.

Planned publishing schedule (as of July 2007)

1996: Volume 1 (published)
2001: Volume 2 (published)
2003: Volume 3 (published)
2005: Volume 4 (published)
2007: Volume 5 (last quarter)
Further volumes at the rate of one every year.

Section 1. Linear algebra. Fields. Algebraic number theory

A. *Linear Algebra*

G.P. Egorychev, Van der Waerden conjecture and applications (1; 22 pp.)
V.L. Girko, Random matrices (1; 52 pp.)
A.N. Malyshev, Matrix equations. Factorization of matrices (1; 38 pp.)
L. Rodman, Matrix functions (1; 38 pp.)
Correction to the chapter by **L. Rodman**, Matrix functions (3; 1 p.)
J.A. Hermida-Alonso, Linear algebra over commutative rings (3; 49 pp.)
Linear inequalities (also involving matrices)
Orderings (partial and total) on vectors and matrices
Positive matrices
Structured matrices such as Toeplitz and Hankel
Integral matrices. Matrices over other rings and fields.
Quasideterminants, and determinants over noncommutative fields.
Nonnegative matrices, positive definite matrices, and doubly nonnegative matrices.
Linear algebra over skew fields

B. *Linear (In)dependence*

J.P.S. Kung, Matroids (1; 28 pp.)

C. *Algebras Arising from Vector Spaces*

Clifford algebras, related algebras, and applications
Other algebras arising from vector spaces (working title only)

D. *Fields, Galois Theory, and Algebraic Number Theory*

(There is also a chapter on ordered fields in Section 4)
J.K. Deveney, **J.N. Mordeson**, Higher derivation Galois theory of inseparable field extensions (1; 34 pp.)
I. Fesenko, Complete discrete valuation fields. Abelian local class field theories (1; 48 pp.)
M. Jarden, Infinite Galois theory (1; 52 pp.)
R. Lidl, **H. Niederreiter**, Finite fields and their applications (1; 44 pp.)
W. Narkiewicz, Global class field theory (1; 30 pp.)
H. van Tilborg, Finite fields and error correcting codes (1; 28 pp.)
Skew fields and division rings. Brauer group

Topological and valued fields. Valuation theory
Zeta and L-functions of fields and related topics
Structure of Galois modules
Constructive Galois theory (realizations of groups as Galois groups)
Dessins d'enfants
Hopf Galois theory
T. Albu, From field theoretic to abstract co-Galois theory (5; 81 pp.)

E. *Nonabelian Class Field Theory and the Langlands Program*

(to be arranged in several chapters by Y. Ihara)

F. *Generalizations of Fields and Related Objects*

U. Hebisch, **H.J. Weinert**, Semi-rings and semi-fields (1; 38 pp.)
G. Pilz, Near rings and near fields (1; 36 pp.)

Section 2. Category theory. Homological and homotopical algebra. Methods from logic

A. *Category Theory*

S. MacLane, **I. Moerdijk**, Topos theory (1; 28 pp.)
R. Street, Categorical structures (1; 50 pp.)
B.I. Plotkin, Algebra, categories and databases (2; 71 pp.)
P.S. Scott, Some aspects of categories in computer science (2; 71 pp.)
E. Manes, Monads of sets (3; 48 pp.)
M. Markl, Operads and PROPs (5; 54 pp.)

B. *Homological Algebra. Cohomology. Cohomological Methods in Algebra. Homotopical Algebra*

J.F. Carlson, The cohomology of groups (1; 30 pp.)
A. Generalov, Relative homological algebra. Cohomology of categories, posets, and coalgebras (1; 28 pp.)
J.F. Jardine, Homotopy and homotopical algebra (1; 32 pp.)
B. Keller, Derived categories and their uses (1; 32 pp.)
A.Ya. Helemskii, Homology for the algebras of analysis (2; 143 pp.))
Galois cohomology
Cohomology of commutative and associative algebras
Cohomology of Lie algebras
Cohomology of group schemes
V. Lyubashenko, **O. Manzyuk**, $A_{\{\infty\}}$-algebras, $A_{\{\infty\}}$-categories, and $A_{\{\infty\}}$-functors (5; 46 pp.)
B.V. Novikov, 0-Cohomology of semigroups (5; 21 pp.)

C. *Algebraic K-theory*

A. Kuku, Classical algebraic K-theory: the functors K_0, K_1, K_2 (3; 55 pp.)

A. Kuku, Algebraic K-theory: the higher K-functors (4; 122 pp.)
Grothendieck groups
K_2 and symbols
KK-theory and EXT
Hilbert C^*-modules
Index theory for elliptic operators over C^* algebras
Simplicial algebraic K-theory
Chern character in algebraic K-theory
Noncommutative differential geometry
K-theory of noncommutative rings
Algebraic L-theory
Cyclic cohomology
Asymptotic morphisms and E-theory
Hirzebruch formulae

D. Model Theoretic Algebra

(See also P.C. Eklof, Whitehead modules, in Section 3B)
M. Prest, Model theory for algebra (3; 31 pp.)
M. Prest, Model theory and modules (3; 34 pp.)
Logical properties of fields and applications
Recursive algebras
Logical properties of Boolean algebras
F.O. Wagner, Stable groups (2; 36 pp.)
The Ax–Ershov–Kochen theorem and its relatives and applications

E. Rings up to Homotopy

Rings up to homotopy
Simplicial algebras

Section 3. Commutative and associative rings and algebras

A. Commutative Rings and Algebras

(See also C. Faith, Coherent rings and annihilator conditions in matrix and polynomial rings, in Section 3B)
(See also Freeness theorems for group rings and Lie algebras in Section 5A)
J.P. Lafon, Ideals and modules (1; 24 pp.)
General theory. Radicals, prime ideals, etc. Local rings (general). Finiteness and chain conditions
Extensions. Galois theory of rings
Modules with quadratic form
Homological algebra and commutative rings. Ext, Tor, etc. Special properties (p.i.d., factorial, Gorenstein, Cohen–Macauley, Bezout, Fatou, Japanese, excellent, Ore, Prüfer, Dedekind, . . . and their interrelations)
D. Popescu, Artin approximation (2; 45 pp.)

Finite commutative rings and algebras. (Absorbed in the Chapter A.A. Nechaev, Finite rings with applications, in Section 3B)

Localization. Local–global theory

Rings associated to combinatorial and partial order structures (straightening laws, Hodge algebras, shellability, ...)

Witt rings, real spectra

R.H. Villareal, Monomial algebras and polyhedral geometry (3; 62 pp.)

B. Associative Rings and Algebras

P.M. Cohn, Polynomial and power series rings. Free algebras, firs and semifirs (1; 30 pp.)

Classification of Artinian algebras and rings

V.K. Kharchenko, Simple, prime, and semi-prime rings (1; 52 pp.)

A. van den Essen, Algebraic microlocalization and modules with regular singularities over filtered rings (1; 28 pp.)

F. Van Oystaeyen, Separable algebras (2; 66 pp.)

K. Yamagata, Frobenius rings (1; 48 pp.)

V.K. Kharchenko, Fixed rings and noncommutative invariant theory (2; 27 pp.)

General theory of associative rings and algebras

Rings of quotients. Noncommutative localization. Torsion theories

von Neumann regular rings

Semi-regular and pi-regular rings

Lattices of submodules

A.A. Tuganbaev, Modules with distributive submodule lattice (2; 25 pp.)

A.A. Tuganbaev, Serial and distributive modules and rings (2; 25 pp.)

PI rings

Generalized identities

Endomorphism rings, rings of linear transformations, matrix rings

Homological classification of (noncommutative) rings

S.K. Sehgal, Group rings and algebras (3; 96 pp.)

Dimension theory

V.V. Bavula, Filter dimension (4; 28 pp.)

A. Facchini, The Krull–Schmidt theorem (3; 42 pp.)

Duality. Morita-duality

Commutants of differential operators

E.E. Enochs, Flat covers (3; 21 pp.)

C. Faith, Coherent rings and annihilation conditions in matrix and polynomial rings (3; 31 pp.)

Rings of differential operators

Graded and filtered rings and modules (also commutative)

P.C. Eklof, Whitehead modules (3; 23 pp.)

Goldie's theorem, Noetherian rings and related rings

Sheaves in ring theory

A.A. Tuganbaev, Modules with the exchange property and exchange rings (2; 25 pp.)

A.A. Nechaev, Finite rings with applications (5; 116 pp.)
T.Y. Lam, Hamilton's quaternions (3; 26 pp.)
Koszul algebras
A.A. Tuganbaev, Semiregular, weakly regular, and π-regular rings (3; 25 pp.)
Hamiltonian algebras
A.A. Tuganbaev, Max rings and V-rings (3; 23 pp.)
Algebraic asymptotics
Anti-automorphisms

C. Coalgebras

W. Michaelis, Coalgebras (3; 120 pp.)
Co-Lie-algebras
Comodules and corings

D. Deformation Theory of Rings and Algebras (*Including Lie Algebras*)

Deformation theory of rings and algebras (general)
Yu. Khakimdzanov, Varieties of Lie algebras (2; 35 pp.)
Deformation theoretic quantization

Section 4. Other algebraic structures. Nonassociative rings and algebras. Commutative and associative algebras with extra structure

A. Lattices and Partially Ordered Sets

Lattices and partially ordered sets
A. Pultr, Frames (3; 63 pp.)
D. Kruml, **J. Paseka**, Algebraic and categorical aspects of quantales (5; 40 pp.)

B. Boolean Algebras

C. Universal Algebra

Universal algebra

D. Varieties of Algebras, Groups, ...

(See also Yu. Khakimdzanov, Varieties of Lie algebras, in Section 3D)
V.A. Artamonov, Varieties of algebras (2; 30 pp.)
Varieties of groups
V.A. Artamonov, Quasi-varieties (3; 19 pp.)
Varieties of semigroups

E. Lie Algebras

Yu.A. Bahturin, **M.V. Zaitsev**, **A.A. Mikhailov**, Infinite-dimensional Lie superalgebras (2; 30 pp.)
General structure theory
Ch. Reutenauer, Free Lie algebras (3; 21 pp.)
Classification theory of semisimple Lie algebras over **R** and **C**

The exceptional Lie algebras
M. Goze, **Y. Khakimdjanov**, Nilpotent and solvable Lie algebras (2; 51 pp.)
Universal enveloping algebras
Modular (ss) Lie algebras (including classification)
Infinite-dimensional Lie algebras (general)
Kac–Moody Lie algebras
Affine Lie algebras and Lie super algebras and their representations
Finitary Lie algebras
Standard bases
A.I. Molev, Gelfand–Tsetlin bases for classical Lie algebras (4; 69 pp.)
Kostka polynomials

F. Jordan Algebras (finite and infinite-dimensional and including their cohomology theory)

G. Other Nonassociative Algebras (Mal'tsev, alternative, Lie admissible, ...)

Mal'tsev algebras
Alternative algebras

H. Hopf Algebras and Related Structures

(See also "Hopf-Galois theory" in Section 1D)
(See also "Co-Galois theory" in Section 1D)
(See also "Algebraic structures on braided categories" in Section 2A)
(See also "Representation theory of semi-simple Hopf algebras" in Section 6D)
M. Cohen, **S. Gelakov**, **S. Westreich**, Hopf algebras (4; 87 pp.)
Classification of pointed Hopf algebras
Recursive sequences from the Hopf algebra and coalgebra points of view
Quantum groups (*general*)
Crystal bases
A.I. Molev, Yangians and their applications (3; 54 pp.)
Formal groups
p-divisible groups
Combinatorial Hopf algebras
Symmetric functions
Special functions and q-special functions, one and two variable case
Quantum groups and multiparameter q-special functions
D. Manchon, Hopf algebras in renormalisation (5; 63 pp.)
Noncommutative geometry à la Connes
Noncommutative geometry from the algebraic point of view
Noncommutative geometry from the categorical point of view
Hopf algebras and operads
Noncommutative symmetric functions and quasi-symmetric functions
Solomon descent algebras
Witt vectors and symmetric function
Picard–Vessiot theory and Hopf algebras

Hopf-algebroids

Trees, dendriform algebras and dialgebras

A. Masuoka, Classification of semisimple Hopf algebras (5; 27 pp.)

Quantum differential geometry, quantum calculus and the quantum approach to noncommutative geometry

Connes–Baum theory

I. Other Rings and Algebras with Additional Structure

Graded and super algebras (commutative, associative; for Lie superalgebras, see Section 4E)

Topological rings

F. Patras, Lambda-rings (3; 34 pp.)

Ordered and lattice-ordered groups, rings and algebras

Rings and algebras with involution. C^*-algebras

A. Levin, Difference algebra (4; 100 pp.)

Differential algebra

Ordered fields

Hypergroups

Stratified algebras

Section 5. Groups and semigroups

A. Groups

(See also "Groups and semigroups of automata transformations" in Section 5B)

A.V. Mikhalev, **A.P. Mishina**, Infinite Abelian groups: Methods and results (2; 48 pp.)

Simple groups, sporadic groups

Representations of the finite simple groups

Diagram methods in group theory

Abstract (finite) groups. Structure theory. Special subgroups. Extensions and decompositions.

Solvable groups, nilpotent groups, p-groups

Infinite soluble groups

Word problems

Burnside problem

Combinatorial group theory

Free groups (including actions on trees)

Formations

Infinite groups. Local properties

Algebraic groups. The classical groups. Chevalley groups

Chevalley groups over rings

The infinite-dimensional classical groups

Other groups of matrices. Discrete subgroups.

M. Geck, **G. Malle**, Reflection groups. Coxeter groups (4; 38 pp.)

M.C. Tamburini, **M. Vsemirnov**, Hurwitz groups and Hurwitz generation (4; 38 pp.)
Groups with BN-pair, Tits buildings, . . .
Groups and (finite combinatorial) geometry
"Additive" group theory
Probabilistic techniques and results in group theory
V.V. Vershinin, Survey on braids (4; 24 pp.)
L. Bartholdi, **R.I. Grigorchuk**, **Z. Šuniḱ**, Branch groups (3; 129 pp.)
Frobenius groups
Just infinite groups
V.I. Senashov, Groups with finiteness conditions (4, 27 pp.)
Automorphism groups of groups
Automorphism groups of algebras and rings
Freeness theorems in groups and rings and Lie algebras
Groups with prescribed systems of subgroups
Automatic groups
Groups with minimality and maximality conditions (school of Chernikov)
Lattice-ordered groups
Linearly and totally ordered groups
Finitary groups
Random groups
Hyperbolic groups
Probabilistic techniques in group theory
Infinite dimensional groups

B. Semigroups

(See also B.V. Novikov, 0-cohomology of semigroups, in Section 2B)
Semigroup theory. Ideals, radicals, structure theory
Semigroups and automata theory and linguistics
Groups and semigroups of automata transformations

C. Algebraic Formal Language Theory. Combinatorics of Words

D. Loops, Quasigroups, Heaps, . . .

Quasigroups in combinatorics

E. Combinatorial Group Theory and Topology

(See also "Diagram methods in group theory" in Section 5A)

Section 6. Representation and invariant theory

A. Representation Theory. General

Representation theory of rings, groups, algebras (general)
Modular representation theory (general)
Representations of Lie groups and Lie algebras (general)
Multiplicity free representations

B. Representation Theory of Finite and Discrete Groups (and Algebras)

Representation theory of finite groups in characteristic zero
Modular representation theory of finite groups. Blocks
Representation theory of the symmetric groups (both in characteristic zero and modular)
Representation theory of the finite Chevalley groups (both in characteristic zero and modular
Modular representation theory of Lie algebras

C. Representation Theory of 'Continuous Groups' (linear algebraic groups, Lie groups, loop groups, ...) and the Corresponding Algebras

(See also A.I. Molev, Gelfand–Tsetlin bases for classical Lie algebras, in Section 4E)
Representation theory of compact topological groups
Representation theory of locally compact topological groups
Representation theory of $SL_2(\mathbf{R})$, ...
Representation theory of the classical groups. Classical invariant theory
Classical and transcendental invariant theory
Reductive groups and their representation theory
Unitary representation theory of Lie groups
Finite-dimensional representation theory of the ss Lie algebras (in characteristic zero); structure theory of semi-simple Lie algebras
Infinite-dimensional representation theory of ss Lie algebras. Verma modules
Representation of Lie algebras. Analytic methods
Representations of solvable and nilpotent Lie algebras. The Kirillov orbit method
Orbit method, Dixmier map, ... for ss Lie algebras
Representation theory of the exceptional Lie groups and Lie algebras
Representation theory of 'classical' quantum groups
A.U. Klimyk, Infinite-dimensional representations of quantum algebras (2; 26 pp.)
Duality in representation theory
Representation theory of loop groups and higher-dimensional analogues, gauge groups, and current algebras
Representation theory of Kac–Moody algebras
Invariants of nonlinear representations of Lie groups
Representation theory of infinite-dimensional groups like GL_∞
Metaplectic representation theory

D. Representation Theory of Algebras

Representations of rings and algebras by sections of sheafs
Representation theory of algebras (Quivers, Auslander–Reiten sequences, almost split sequences, ...)
Quivers and Their Representations
Tame algebras
Ringel–Hall algebras
Composition algebras

Quasi-heriditary algebras
Cellular algebras
Representation theory of (semi-simple) Hopf algebras

E. Abstract and Functorial Representation Theory

Abstract representation theory
S. Bouc, Burnside rings (2; 60 pp.)
P. Webb, A guide to Mackey functors (2; 32 pp.)

F. Representation Theory and Combinatorics

G. Representations of Semigroups

Representation of discrete semigroups
Representations of Lie semigroups

H. Hecke Algebras

Hecke–Iwahori algebras

I. Invariant theory

Section 7. Machine computation. Algorithms. Tables

Some notes on this volume: Besides some general article(s) on machine computation in algebra, this volume should contain specific articles on the computational aspects of the various larger topics occurring in the main volume, as well as the basic corresponding tables. There should also be a general survey on the various available symbolic algebra computation packages.

The CoCoA computer algebra system

G.P. Egorychev, **E.V. Zima**, Integral representation and algorithms for closed form summation (5; 71 pp.)

Groebner bases and their applications

Section 8. Applied algebra

Section 9. History of algebra

(See also K.T. Lam, Hamilton's quaternions, in Section 3B)
History of coalgebras and Hopf algebras
Development of algebra in the 19-th century

This page intentionally left blank

Contents

This page intentionally left blank

List of Contributors

T. Albu, Koç University, Department of Mathematics, Rumeli Feneri Yolu, 34450 Sariyer, Istanbul, Turkey; “Simion Stoilow”, Institute of Mathematics of the Romanian Academy, P.O. Box 1, 764, RO, 010145 Bucharest 1, Romania, *e-mail: toma.albu@imar.ro*

G.P. Egorychev, Krasnoyarsk State Technical University, Kirenskogo 26, Krasnoyarsk 660074, Russia, *e-mail: anott@scn.ru*

D. Kruml, Department of Mathematics, Masaryk University, Janáčkovo nám. 2a, 602 00 Brno, Czech Republic, *e-mail: kruml@math.muni.cz*

V. Lyubashenko, Institute of Mathematics, National Academy of Sciences of Ukraine, 3 Tereshchenskivska st., Kyiv-4, 01601 MSP, Ukraine, *e-mail: lub@imath.kiev.ua*

D. Manchon, Laboratoire de mathématiques (CNRS-UMR 6620), Université Blaise Pascal, F63177 Aubière cedex, France, *e-mail: manchon@math.univ-bpclermont.fr*

O. Manzyuk, Fachbereich Mathematik, Postfach 3049, 67653 Kaiserslautern, Germany, *e-mail: manzyuk@mathematik.uni-kl.de*

M. Markl, Mathematical Institute of the Academy, Zitná 25, 11567 Prague 1, Czech Republic, *e-mail: markl@math.cas.cz*

A. Masuoka, Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan, *e-mail: akira@math.tsukuba.ac.jp*

A. Nechaev, Center of New Information Technologies, Moscow State University, Vorobjevy Gory, Moscow, Russia, *e-mail: nechaev@cnit.msu.ru*

B.V. Novikov, Department of Mathematics, University of Kharkov, Ukraine, *e-mail: boris.v.novikov@univer.kharkov.ua*

J. Paseka, Department of Mathematics, Masaryk University, Janáčkovo nám. 2a, 602 00 Brno, Czech Republic, *e-mail: paseka@math.muni.cz*

E.V. Zima, Physics and Computer Science, Wilfrid Laurier University, Waterloo, Canada, *e-mail: ezima@wlu.ca*

This page intentionally left blank

Section 1D

Fields, Galois Theory, and Algebraic Number Theory

This page intentionally left blank

From Field Theoretic to Abstract Co-Galois Theory

Toma Albu[*]

"Simion Stoilow" Institute of Mathematics of the Romanian Academy, P.O. Box 1-764, *RO*-010145 *Bucharest* 1, *Romania*

E-mail: *Toma.Albu@imar.ro*

Contents

[*] The author gratefully acknowledges partial financial support from Contract MEdC 2-CEx 06-11-20/2006–2008 awarded by the Ministerul Educaţiei şi Cercetării, România and from Grant 1618/2007 awarded by the Consiliul Naţional al Cercetării Ştiinţifice în Învăţământul Superior, România.

HANDBOOK OF ALGEBRA, VOL. 5
Edited by M. Hazewinkel
© 2008 Elsevier B.V. All rights reserved
DOI: 10.1016/S1570-7954(07)05001-2

Abstract

The aim of this chapter is to present the basic notions and results of co-Galois theory, a fairly new area in field theory investigating field extensions, finite or not, that possess a co-Galois correspondence. The subject is somewhat dual to the very classical Galois theory dealing with field extensions possessing a Galois correspondence. There exists an abstract Galois theory for arbitrary profinite groups; an abstract co-Galois theory for such groups was recently invented, and presented here.

Keywords: field, field extension, radical extension, Kneser extension, strongly Kneser extension, Kneser criterion, Galois extension, co-Galois extension, G-co-Galois extension, purity criterion, Galois theory, co-Galois theory, Galois connection, co-Galois connection, Galois group, co-Galois group, Kneser group, Kummer theory, Abelian extension, Galois cohomology, profinite group, cohomology group, crossed homomorphism, Pontryagin duality, algebraic number field, group-graded algebra, Hopf algebra, abstract co-Galois theory, Kneser group of cocycles, co-Galois group of cocycles, Kummer group of cocycles

1. Introduction

An interesting but difficult problem in field theory is to describe in a satisfactory manner the set $\mathbb{I}(E/F)$ which, as a matter of fact is a lattice, of all intermediate fields of a given field extension E/F. If E/F is a finite Galois extension, then by the *fundamental theorem of finite Galois theory*, there exists a canonical one-to-one order-reversing correspondence, or equivalently, a *lattice anti-isomorphism* between the lattice $\mathbb{I}(E/F)$ and the lattice $\mathbb{L}(\Gamma)$ of all subgroups of a certain group Γ canonically associated with the extension E/F, namely the Galois group $\mathrm{Gal}(E/F)$; we say that such an E/F is an *extension with Γ-Galois correspondence*.

On the other hand, there exists a fairly large class of field extensions which are not necessarily Galois, but enjoy a property that is dual to the previous one. Namely, these are the extensions E/F for which there exists a canonical lattice isomorphism (and *not* a lattice anti-isomorphism as in the Galois case) between $\mathbb{I}(E/F)$ and $\mathbb{L}(\Delta)$, where Δ is a certain group canonically associated with the extension E/F. We call the members of this class, *extensions with Δ-co-Galois correspondence*. Their prototype is the field extension $\mathbb{Q}(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r})/\mathbb{Q}$, where $r, n_1, \ldots, n_r, a_1, \ldots, a_r$ are positive integers, and where $\sqrt[n_i]{a_i}$ is the positive real n_i-th root of a_i for each i, $1 \leqslant i \leqslant r$. For such an extension, the associated group Δ is the quotient group $\mathbb{Q}^*\langle \sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}\rangle/\mathbb{Q}^*$. Note that the finite *classical Kummer* extensions have a privileged position: they are at the same time extensions with Galois and with co-Galois correspondences, and the two groups appearing in this setting are isomorphic.

The purpose of this chapter is to present the basic concepts, results, and methods of studying field extensions, finite or not, which possess a co-Galois correspondence. This topic, called *co-Galois theory*, is dual to the very classical one known as *Galois theory* investigating field extensions possessing a Galois correspondence.

On the other hand, Galois theory can be also developed in an abstract group theoretic framework, namely for arbitrary profinite groups. Since the profinite groups are precisely those topological groups which arise as Galois groups of Galois extensions, an *abstract Galois theory* for arbitrary profinite groups was developed within *abstract class field theory* (see, e.g., Neukirch [63]). Therefore, a dual theory to abstract Galois theory, called *abstract co-Galois theory*, emerging from and generalizing (field theoretic) co-Galois theory, has been very recently invented. We present it in this chapter.

The field extensions possessing a Galois or co-Galois correspondence are very particular illustrations of the following general problem in mathematics: *Describe in a satisfactory manner the collection of all subobjects of a given object of a category $\mathcal{C}$.* In general, this is a difficult problem, but sometimes it can be reduced to describing the subobjects of an object in another more suitable category $\mathcal{D}$. For instance, let F be a given field and let $\mathcal{C}$ denote the category of all field extensions of F. If E is any object of $\mathcal{C}$, i.e., a field extension E/F, then the set $\mathbb{I}(E/F)$ of all subfields of E containing F is precisely the set of all subobjects of E in $\mathcal{C}$. This set is, in general, a complicated-to-conceive, potentially infinite set of hard-to-describe-and-identify objects. However, when E/F is a finite Galois extension, then, as we have already noted, the fundamental theorem of finite Galois theory establishes a lattice anti-isomorphism between the lattices $\mathbb{I}(E/F)$ and $\mathbb{L}(\mathrm{Gal}(E/F))$. In this way, the lattice of all subobjects of an object $E \in \mathcal{C}$, which has the additional property that is a finite Galois extension of F, can be described by the lattice of all subobjects of the object $\mathrm{Gal}(E/F)$ in the category of all finite groups; in principle, this category is more convenient than the category $\mathcal{C}$ of all field extensions of F, since the set of all subgroups of a finite group is a far more benign object. Thus, many questions concerning a field are best studied by transforming them into group theoretical questions in the group of automorphisms of the field.

If now $E \in \mathcal{C}$ is an infinite Galois extension of F, the above description does not work anymore, and in this case the Galois group $\mathrm{Gal}(E/F)$ is in fact a *profinite group*, that is, a projective limit of finite groups, or equivalently, a Hausdorff, compact, totally disconnected topological group, having as a fundamental system of open neighborhoods of the identity element the set of all subgroups of it of the form $\mathrm{Gal}(E/K)$ with $K \in \mathbb{I}(E/F)$ and K/F a finite Galois extension; this topology on $\mathrm{Gal}(E/F)$ is called the *Krull topology*, and it coincides with its *finite topology*. By the *fundamental theorem of infinite Galois theory*, there exists a canonical one-to-one order-reversing correspondence, or equivalently, a *lattice anti-isomorphism* between the lattice $\mathbb{I}(E/F)$ and the lattice $\overline{\mathbb{L}}(\Gamma)$ of all closed subgroups of the Galois group Γ of the extension E/F. Observe that the lattice $\overline{\mathbb{L}}(\Gamma)$ is nothing else than the lattice of all subobjects of Γ in the category of all profinite groups.

The Galois group of a given Galois field extension E/F, finite or not, is in general difficult to describe concretely; so, it will be desirable to impose additional conditions on E/F such that the lattice $\mathbb{I}(E/F)$ be isomorphic (or anti-isomorphic) to the lattice $\mathbb{L}(\Delta)$ of all subgroups of some other group Δ, easily computable and appearing explicitly in the data of the given Galois extension E/F. A class of such Galois extensions is that of the *classical Kummer extensions*. Recall that a field extension E/F is said to be a *classical n-Kummer extension*, where n is a positive integer, if $\gcd(n, e(F)) = 1, \zeta_n \in F$, and $E = F(\{\sqrt[n]{a_i} \mid i \in I\})$, where $e(F)$ is the characteristic exponent of F, ζ_n is a primitive n-th root of unity in a fixed algebraic closure $\overline{F}$ of F, I is an arbitrary set, finite or not, $a_i \in F^*$, and $\sqrt[n]{a_i}$ is a certain root in $\overline{F}$ of the polynomial $X^n - a_i$, $i \in I$ (see Section 2 for the notation used); note that the extension E/F is finite if and only if the set I can be chosen to be finite. For such a classical n-Kummer extension E/F, if one denotes by Δ the so called *Kummer group* $\mathrm{Kum}(E/F) := F^*\langle\{\sqrt[n]{a_i} \mid i \in I\}\rangle/F^*$ of E/F, then *Kummer theory* establishes a canonical lattice isomorphism $\mathbb{I}(E/F) \xrightarrow{\sim} \mathbb{L}(\Delta)$. Observe that the torsion Abelian group Δ is intrinsically given with the extension E/F and easily manageable as well. This group is isomorphic, but not canonically, with the *character group* $\widehat{\Gamma}$ of the Galois group Γ of E/F (see also Section 5); in particular, it follows that for E/F finite, the group Δ has exactly $[E : F]$ elements. Note also that E/F is a Galois extension, with Abelian Galois group of exponent a divisor of n.

In our opinion, there are six periods or moments in the history of co-Galois theory:

I. The classical period (∼1830–∼1930) was basically concerned with the description of the lattices of all intermediate fields of finite Galois extensions and Kummer extensions E/F in terms of the anti-isomorphic and isomorphic lattices of all subgroups of two different, but isomorphic, groups canonically associated with E/F, namely the Galois group and the Kummer group of E/F, respectively. Of course, E. Galois (1811–1832) and E.E. Kummer (1810–1893) were the main contributors in this era.

II. The pre-Kneser period (1930–1975) comprises mainly attempts made in the following two directions: (1) to see when is $[F(\sqrt[n]{a_1}, \dots, \sqrt[n]{a_r}) : F] = \prod_{1 \leqslant i \leqslant r} [F(\sqrt[n]{a_i}) : F]$; and (2) to weaken the condition $\zeta_n \in F$ above for Kummer extensions in order to compute effectively the degree of particular finite radical extensions, i.e., of extensions of type $F(\sqrt[n_1]{a_1}, \dots, \sqrt[n_r]{a_r})/F$, where F was mainly an algebraic number field. The question (1) was first answered by H. Hasse in 1930 (see the second edition [50] of his mimeographed lectures on class field theory). The case when the algebraic number field F does not necessarily contain a primitive n-th root ζ_n of unity was, surprisingly, first discussed fairly late, only in 1940 by A. Besicovitch [30] for $F = \mathbb{Q}$ and $\sqrt[n]{a_1}, \dots, \sqrt[n]{a_r}$ real roots of positive integers $a_1, \dots a_r$ satisfying certain additional conditions, and then, by L.J. Mordell [62] in 1953 for F any real number field and $\sqrt[n]{a_1}, \dots, \sqrt[n]{a_r} \in \mathbb{R}$. Later, in 1972, C.L. Siegel [73] shows that the degree $[F(\sqrt[n]{a_1}, \dots, \sqrt[n]{a_r}) : F]$ is the order of the quotient group $F^*\langle \sqrt[n]{a_1}, \dots, \sqrt[n]{a_r} \rangle / F^*$ for any real number field F and any real roots $\sqrt[n]{a_1}, \dots, \sqrt[n]{a_r}$. A particular case of Besicovitch's result was proved by I. Richards [67] in 1974 (see also L. Gaal's book [42], where Richards' proof is reproduced). A result of the same nature was established by H.D. Ursell [79]. All these results deal with a particular case of the following

PROBLEM 1. *Let F be a field, let $\overline{F}$ be an algebraic closure of F, and let $x_1, \dots, x_r \in \overline{F}$ be elements of degree $n_1, \dots, n_r$ over F, respectively. When does the field $F(x_1, \dots, x_r)$ have degree $n_1 \cdots n_r$ over F?*

A more general question is

PROBLEM 2. *With the same notation and hypotheses as in Problem 1, when can be found an explicit formula to compute $[F(x_1, \dots, x_r) : F]$?*

Partial answers to Problem 2 are given, as we already have seen above, by a well known result on classical finite Kummer extensions (see, e.g., E. Artin's book [22]), as well as by a result appearing in I. Kaplansky's book [53], and by another one of similar nature due to A. Baker and H.M. Stark [24].

III. The Kneser moment (1975) is represented by the appearance of the only two-page-paper of M. Kneser [55]. In this paper Kneser answered Problem 2 for a large class of extensions that have since been named *Kneser extensions* by Albu and Nicolae [16], honoring his nice and important result.

IV. The pre-Greither & Harrison period (1975–1986) contains the investigation of finite radical extensions, with or without the use of the Kneser criterion. The main work in this area was done, chronologically, by A. Schinzel [68–70], D. Gay and W.Y. Vélez [43,44], W. May [60], M. Norris and W.Y. Vélez [64], F. Halter-Koch [47,48], W.Y. Vélez [80–82], M. Acosta de Arozco and W.Y. Vélez [1,2], etc.

V. The Greither & Harrison moment (1986) represents in fact the birth year of the *co-Galois theory*. It gives a partial answer to the following natural

PROBLEM 3. *Find large classes $\mathcal{F}$ of (finite) field extensions, not necessarily Galois, for which the lattice of all intermediate fields of every $E/F \in \mathcal{F}$ can be described in terms of the lattice of all subgroups of a certain group canonically associated with E/F.*

To the best of our knowledge, the term of "co-Galois" appeared for the first time in the literature in 1986 in the fundamental paper of C. Greither and D.K. Harrison [45], where *co-Galois extensions* were introduced and investigated. They also considered other classes of finite field extensions possessing a co-Galois correspondence, including the so called *neat presentations*. The prototype of an extension with a Δ-co-Galois correspondence is any field extension $\mathbb{Q}(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r})/\mathbb{Q}$, where $r, n_1, \ldots, n_r, a_1, \ldots, a_r \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}$, and for every i, $1 \leqslant i \leqslant r$, $\sqrt[n_i]{a_i}$ is the positive real n_i-th root of a_i. For such an extension, the associated group Δ is the quotient group $\mathbb{Q}^*\langle \sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}\rangle/\mathbb{Q}^*$. It seems surprising it was stated and proved explicitly only fairly late, in 1986, by Greither and Harrison [45], that such extensions are extensions with a Δ-co-Galois correspondence. In particular, it follows that $[\mathbb{Q}(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}) : \mathbb{Q}] = |\mathbb{Q}^*\langle \sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}\rangle/\mathbb{Q}^*|$.

VI. The post-Greither & Harrison period (1986–present) begins in 1989 with A. Masuoka [59], who apparently used for the first time in the literature the term of "co-Galois theory". Besides the co-Galois extensions introduced by C. Greither and D.K. Harrison already in 1986, new important classes of finite radical field extensions co-Galois theory deals with are introduced and investigated: *G-Kneser extensions*, *strongly G-Kneser extensions*, *G-co-Galois extensions*. The setting of G-co-Galois extensions permits a simple and unified manner to study the classical Kummer extensions, the Kummer extensions with few roots of unity, co-Galois extensions, and neat presentations. In 2001 an *infinite co-Galois theory* investigating infinite radical extensions has been developed, in 2003 appeared the author's monograph "Cogalois Theory" [11], and in 2005 infinite co-Galois theory has been generalized to arbitrary profinite groups, leading to a so called *abstract co-Galois theory*. The main contributors of co-Galois theory in this period, in chronological order, are: D.S. Dummit [39], A. Masuoka [59], F. Barrera-Mora, M. Rzedowski-Calderón, and G. Villa-Salvador [26,27], T. Albu [3–13], F. Barrera-Mora and W.Y. Vélez [28], T. Albu and F. Nicolae [16–19], T. Albu, F. Nicolae, and M. Ţena [20], P. Lam-Estrada, F. Barrera-Mora, and G.D. Villa-Salvador [56], F. Barrera-Mora [25], D. Ştefan [74], M. Ţena [78], T. Albu and M. Ţena [21], T. Albu and Ş.A. Basarab [14,15], etc.

We are now going to describe the contents of this chapter. Section 2 contains the necessary basic notation and terminology which will be used throughout the chapter.

Section 3 introduces and investigates three basic concepts of co-Galois theory, namely that of *G-radical extension*, of *G-Kneser extension*, and of *co-Galois extension*. The concept of a *radical extension* is rather basic and well known in Galois theory. However, our terminology is somewhat different from that used in Galois theory (see, e.g., Kaplansky [53], Karpilovsky [54], Lang [57]), but coincides for simple extensions. Note that radical extensions are called *coseparable* by Greither and Harrison [45]. Roughly speaking, a radical extension is a field extension E/F such that E is obtained by adjoining to the base field F an arbitrary set of "radicals" over F, i.e., of elements $x \in E$ such that $x^n = a \in F$ for some $n \in \mathbb{N}^*$. Such an x is denoted by $\sqrt[n]{a}$ and is called an n-th radical of a. So, E/F is a radical extension when $E = F(R)$, where R is a set of radicals over F. Clearly, one can replace R by the subgroup $G = F^*\langle R\rangle$ of the multiplicative group E^* of E generated by F^* and R. Thus, any radical extension E/F has the form $E = F(G)$, where G is a subgroup of E^* containing F^*, with G/F^* a torsion group. Such an extension is called *G-radical*. A field extension E/F, which is not necessarily finite, is called *G-Kneser* if it is a G-radical extension such that there exists a set of representatives for the quotient group G/F^* which is linearly independent over F; in case the G-radical extension E/F is finite then the last condition can be expressed equivalently as $|G/F^*| = [E : F]$. The extension E/F is called *Kneser* if it is G-Kneser for some group G. These extensions were introduced by Albu and Nicolae [16] for finite extensions and by Albu and Ţena [21] for infinite extensions, honoring the nice criterion due to Kneser [55] evaluating the degrees of separable finite radical extensions. This criterion is a basic tool in co-Galois theory. *Co-Galois extensions*, which were introduced by Greither and Harrison in [45] for finite extensions and by Albu and Ţena [21] for infinite extensions are nothing else than the field extensions E/F which are $T(E/F)$-Kneser, where $T(E/F)$ is the subgroup of the multiplicative group E^* of the field E such that the quotient subgroup $T(E/F)/F^*$ is the torsion subgroup $t(E^*/F^*)$ of the quotient group E^*/F^*. The group $T(E/F)/F^*$, called by Greither and Harrison the *co-Galois group* of the extension E/F, is denoted by $\mathrm{Cog}(E/F)$. Note that the torsion group $t(E^*/F^*)$ of an extension E/F was intensively investigated by Acosta de Orozco and Vélez [2], Gay and Vélez [44]. To the best of our knowledge, the name of *co-Galois group* of E/F for the group $t(E^*/F^*)$ appeared for the first time in the literature in the fundamental paper [45] of Greither and Harrison. The term of "co-Galois group" was also used by Enochs, Rozas, and Oyonarte [40,41], but with a completely different meaning, involving the concept of an $\mathcal{F}$-cover of a module. In general, the concrete computation of the co-Galois group of an extension is not so easy. We completely describe $\mathrm{Cog}(E/F)$ for quadratic extensions, as well as for other classes of finite or infinite extensions. The main results of Section 3 were obtained, chronologically, by Kneser [55], Gay and Vélez [44], Greither and Harrison [45], Barrera-Mora, Rzedowski-Calderón, and Villa-Salvador [26,27], Albu and Nicolae [16], Albu, Nicolae, and Ţena [20], Albu and Ţena [21], Albu [5].

Section 4 contains some of the main results of this chapter. After presenting a very general discussion of Galois connections and co-Galois connections, we associate with any G-radical extension E/F, finite or not, a canonical co-Galois connection

$$\begin{tikzcd}\mathbb{I}(E/F) \arrow[r, shift left, "\varphi"] & \mathbb{L}(G/F^*) \arrow[l, shift left, "\psi"]\end{tikzcd}$$

between the lattice $\mathbb{I}(E/F)$ of all intermediate fields of the extension E/F and the lattice $\mathbb{L}(G/F^*)$ of all subgroups of the quotient group G/F^*, where $\varphi : \mathbb{I}(E/F) \to \mathbb{L}(G/F^*)$, $\varphi(K) = (K \cap G)/F^*$, and $\psi : \mathbb{L}(G/F^*) \to \mathbb{I}(E/F)$, $\psi(H/F^*) = F(H)$. Then, the basic notion of a *strongly G-Kneser extension*, introduced by Albu and Nicolae [16] for finite extensions and by Albu and Ţena [21] for infinite extensions is discussed: an extension E/F is said to be strongly G-Kneser if it is G-radical such that, for any intermediate field K of E/F, the extension K/F is $K^* \cap G$-Kneser, or equivalently, the extension E/K is K^*G-Kneser. These are precisely the G-Kneser extensions for which the maps φ and ψ defined above are isomorphisms of lattices, inverse to one another; in other words, the G-Kneser extensions E/F with *G/F^*-co-Galois correspondence*. In the theory of field extensions with a G/F^*-co-Galois correspondence the most interesting are those which additionally are separable. They were called *G-co-Galois extensions* by Albu and Nicolae [16], and are completely characterized within the class of finite G-radical extensions by means of a very useful *n-purity criterion*, where n is the exponent of the finite group G/F^*, and generalized by Albu [5] to infinite extensions. This allows to obtain in a simple and unified manner, and even in a more general setting, a series of results from classical Kummer theory, as well as results of Albu [3] concerning Kummer extensions with few roots of unity, and of Barrera-Mora, Rzedowski-Calderón, and Villa-Salvador [26] and Greither and Harrison [45] concerning co-Galois extensions and neat presentations. It is shown that a separable G-Kneser extension E/F is G-co-Galois if and only if the group G/F^* has a prescribed structure. As a consequence, the uniqueness of the group G is deduced; this means that if the extension E/F is simultaneously G-co-Galois and H-co-Galois, then necessarily $G = H$. Consequently, it makes sense to define the *Kneser group* of a G-co-Galois extension as the group G/F^*, which is denoted by $\mathrm{Kne}(E/F)$.

It is well known that for any finitely many elements $x_1, \dots, x_r$ in $\overline{F}$ which are separable over an infinite field F, there exist elements $c_1, \dots, c_r$ in F such that $\theta = \sum_{1 \leqslant i \leqslant r} c_i x_i$ is a primitive element of the finite separable extension $F(x_1, \dots, x_r)/F$, i.e., $F(x_1, \dots, x_r) = F(\theta)$. The following natural problem arises: *Let F be any field, and let $x_1, \dots, x_r$ in $\overline{F}$ be finitely many separable elements over F. When is $\sum_{1 \leqslant i \leqslant r} x_i$ a primitive element of the finite separable extension $F(x_1, \dots, x_r)/F$?* Partial answers to this problem are given, e.g., by Albu [3], Kaplansky [53], and Zhou [84]. These were extended by Albu and Nicolae in [17], where a general statement was proved for the large class of finite separable field extensions with co-Galois correspondence, namely: if E/F is a G-co-Galois extension, $n \in \mathbb{N}^*$, and $(x_i)_{1 \leqslant i \leqslant n}$ is a finite family of elements of G such that $x_i F^* \neq x_j F^*$ for every $i, j \in \{1, \dots, n\}, i \neq j$, then $x_1 + \cdots + x_n$ is a primitive element of E/F if and only if $G = K^* \langle x_1, \dots, x_n \rangle$. In particular, from this general approach it follows very easily that $\mathbb{Q}(\sqrt[n_1]{a_1}, \dots, \sqrt[n_r]{a_r}) = \mathbb{Q}(\sqrt[n_1]{a_1} + \cdots + \sqrt[n_r]{a_r})$, where $r, n_1, \dots, n_r, a_1, \dots, a_r \in \mathbb{N}^*$. The results of Section 4 were obtained, chronologically, by Vélez [80], Acosta de Orozco and Vélez [1], Albu and Nicolae [16,17], Lam-Estrada, Barrera-Mora, and Villa-Salvador [56], Albu and Ţena [21], Albu [5,8].

Section 5 is devoted to the investigation of Galois G-co-Galois extensions. The main result of this section is the description, due to Dummit [39] for extensions of algebraic number fields, and to Masuoka [59] and Barrera-Mora, Rzedowski-Calderón, and Villa-Salvador [26] for arbitrary (infinite) Galois extensions, of the co-Galois group $\mathrm{Cog}(E/F)$ of any Galois extension E/F by means of the canonically isomorphic group $Z_c^1(\mathrm{Gal}(E/F), \mu(E))$ of all continuous 1-cocycles of the profinite Galois group $\mathrm{Gal}(E/F)$ of the extension E/F with coefficients in the discrete group $\mu(E)$ of all roots of unity in E. This immediately implies a nice result of Greither and Harrison [45], that originally has been proved in a complicated manner, saying that the co-Galois group of any extension of algebraic number fields is finite. It is also shown that if E/F is a Galois n-bounded G-co-Galois extension then the Kneser group $\mathrm{Kne}(E/F)$ of E/F is isomorphic to the group $Z_c^1(\mathrm{Gal}(E/F), \mu_n(E))$ of all continuous crossed homomorphisms of $\mathrm{Gal}(E/F)$ with coefficients in the discrete group $\mu_n(E)$ of all n-th roots of unity contained in E. A similar result holds for an arbitrary Galois G-co-Galois extension, with $\mu_n(E)$ replaced by a certain subgroup $\mu_G(E)$ of the group $\mu(E)$ of all roots of unity contained in E. Then we investigate Galois extensions which are radical, Kneser, or G-co-Galois in terms of continuous crossed homomorphisms. Next, we present the basic terminology, notation, and facts concerning lattice-isomorphic groups, which will be needed in the investigation of Abelian G-co-Galois extensions. A prototype of these extensions is any classical n-Kummer extension. We are especially interested in finding the connection between the Kneser group $\mathrm{Kne}(E/F)$ and the Galois group $\mathrm{Gal}(E/F)$ of an arbitrary Abelian G-co-Galois extension E/F. We show that the first one is isomorphic to the group $\mathrm{Ch}(\mathrm{Gal}(E/F))$ of characters of the profinite group $\mathrm{Gal}(E/F)$. In particular, in the case of finite extensions these two groups are isomorphic, but not in a canonical way.

Further we show that classical *Kummer theory* can be immediately deduced from co-Galois theory using the n-purity criterion. Moreover, this criterion allows us to provide large classes of G-co-Galois extensions which generalize or are closely related to classical Kummer extensions: *generalized Kummer extensions*, *Kummer extensions with few roots of unity*, and *quasi-Kummer extensions*. The prototype of an (infinite) Kummer extension with few roots of unity is any subextension of $\mathbb{R}/\mathbb{Q}$ of the form $\mathbb{Q}(\{\sqrt[n]{a_i} \mid i \in I\})/\mathbb{Q}$, where $\{a_i \mid i \in I\}$ is an arbitrary nonempty set of strictly positive rational numbers. Notice that, in general, these extensions are not Galois if $n \geqslant 3$. In particular, we derive from co-Galois theory results on Kummer extensions with few roots of unity, which are very similar to the known ones for classical finite Kummer extensions. The idea to place the finite classical Kummer extensions in the framework of G-co-Galois extensions goes back to Albu and Nicolae [16], and was also exploited by Albu [7] and Albu and Ţena [21] for infinite Kummer extensions. The notion of a (finite) *neat presentation* is due to Greither and Harrison [45]. Albu and Nicolae [16] introduced the more general concept of (finite) *generalized neat presentation*. Infinite generalized neat presentations, called here *quasi-Kummer extensions* were introduced and investigated by Albu and Ţena [21]. Note that the fact that finite neat presentations have a co-Galois correspondence, i.e., they are G-co-Galois, was originally proved by Greither and Harrison [45] in a very complicated manner by using heavy cohomological machinery, which includes the Lyndon–Hochschild spectral sequence. The results of Section 5 were obtained, chronologically, by Baer [23], Greither and Harrison [45], Dummit [39], Masuoka [59], Barrera-Mora, Rzedowski-Calderón, and Villa-Salvador [26], Albu and Nicolae [16,19] Albu and Ţena [21], Albu [5–7,11], Albu and Basarab [14].

Section 6 presents some applications of co-Galois theory to elementary field arithmetic and algebraic number theory. We are specially interested in investigating when the extension $\mathbb{Q}(\sqrt{r + \sqrt{d}})/\mathbb{Q}$ with $r, d \in \mathbb{Q}$ is Galois, radical, Kneser, or co-Galois. It turns out that such extensions provide many interesting examples and counterexamples in co-Galois theory. The Kneser criterion has nice applications not only in investigating field extensions with co-Galois correspondence, but also in proving some results in algebraic number theory. We present in this section such applications. Thus, a series of classical results due to Hasse [50], Besicovitch [30], Mordell [62], and Siegel [73] concerning the computation of degrees of particular radical extensions of algebraic number fields, as well as the links between these results, are not only presented in Subsection 6.3, but their very easy proofs, based on the Kneser criterion, are also provided. Next we deal with a surprising application of co-Galois theory in proving some very classical results in algebraic number theory. More precisely, we apply our approach to establish very easily a classical result, related to the so-called *Hecke systems of ideal numbers*, claimed by Hecke (but not proved) in his book [52]. The results of this section were obtained, in a chronological order, by Hecke [51,52], Hasse [49,50], Besicovitch [30], Mordell [62], Baker and Stark [24], Siegel [73], Kaplansky [53], Ursell [79], Schinzel [70], Neukirch [63], Albu and Nicolae [17,18], Albu [4,9].

Section 7 contains links of co-Galois theory with graded algebras and Hopf algebras. We analyze first the basic concepts of co-Galois theory like G-radical, G-Kneser, and G-co-Galois field extension in terms of Clifford extensions and strongly group-graded algebras. The concepts of a *Clifford system* and a *Clifford extension* were invented in 1970 by Dade in his papers [35,36] devoted to so-called *Clifford theory*. This theory investigates when an absolutely irreducible character of a normal subgroup N of a finite group G, defined over an algebraically closed field of arbitrary characteristic, can be extended to a character of G. Dade also introduced ten years later in [37] the concept of a *strongly group-graded algebra*. The idea to use the Clifford systems and Clifford extensions in investigating finite G-Kneser and G-co-Galois extensions is due to Ştefan [74]. His results were generalized from finite to infinite field extensions by Albu [12]. A similar approach in investigating co-Galois extensions E/F, finite or not, is due to Masuoka [59] using the concepts of group-graded field extension and coring. Next, we describe the Kneser and co-Galois field extensions in terms of *Galois H-objects* appearing in *Hopf algebras*. Note that the connection between co-Galois extensions and Hopf algebras is mentioned in passing in Greither and Harrison [45]. The explicit connections provided in this section are due to Albu [12]. The main results of this section were obtained by Masuoka [59], Ştefan [74], Albu [12].

The efforts to generalize the famous Gauss' quadratic reciprocity law led to the theory of Abelian extensions of global and local fields, known as *class field theory*. This theory can be also developed in an abstract group theoretic framework, namely for arbitrary profinite groups. Since, according to a result of Leptin [58] (see also Ribes and Zaleskii [66], and/or Wilson [83]), the profinite groups are precisely those topological groups which arise as Galois groups of Galois extensions, an *abstract Galois theory* for arbitrary profinite groups was developed within the *abstract class field theory* (see, e.g., Neukirch [63]). The aim of Section 8 is to present a dual theory, we called *abstract co-Galois theory*, to the abstract Galois theory. The basic concepts of field theoretic co-Galois theory, namely that of G-Kneser and G-co-Galois field extensions, as well as their main properties are generalized to arbitrary profinite groups. More precisely, let Γ be an arbitrary profinite group, and let A be any subgroup of the Abelian group $\mathbb{Q}/\mathbb{Z}$ such that Γ acts continuously on the discrete group A. Then, one defines the concepts of a *Kneser* subgroup and *co-Galois* subgroup of the group $Z_c^1(\Gamma, A)$ of all continuous 1-cocycles of Γ with coefficients in A, and one establish their main properties, including an *abstract Kneser criterion* for Kneser groups of cocycles and an *abstract quasi-purity criterion* for co-Galois groups of cocycles. The proofs, involving cohomological as well as topological tools, are completely different from that of their field theoretic analogs. A natural dictionary relates the basic notions of (field theoretic) co-Galois theory to their correspondents in the abstract co-Galois theory, which permit us to retrieve easily most of the basic results of the former one from the corresponding results from the latter one.

The idea to involve the group $Z_c^1(\Gamma, A)$ in defining the abstract concepts mentioned above comes from the description presented in Section 5, via the Hilbert theorem 90, of the co-Galois group $\mathrm{Cog}(E/F)$ of an arbitrary Galois extension E/F as the group $Z_c^1(\mathrm{Gal}(E/F), \mu(E))$ of all continuous 1-cocycles of the profinite Galois group $\mathrm{Gal}(E/F)$ of the extension E/F with coefficients in the group $\mu(E)$ of all roots of unity in E. Note that the multiplicative group $\mu(E)$ is isomorphic (in a noncanonical way) to a subgroup of the additive group $\mathbb{Q}/\mathbb{Z}$, and that the basic groups appearing in the investigation of E/F from co-Galois theory perspective are subgroups of $\mathrm{Cog}(E/F)$. In this way, the above description of $\mathrm{Cog}(E/F)$ in terms of 1-cocycles naturally suggests to study the abstract setting of subgroups of groups of type $Z_c^1(\Gamma, A)$, with Γ an arbitrary profinite group and A any subgroup of $\mathbb{Q}/\mathbb{Z}$ such that Γ acts continuously on the discrete group A. Such a continuous action establishes through the evaluation map $\Gamma \times Z_c^1(\Gamma, A) \to A$, $(\sigma, g) \mapsto g(\sigma)$, a Galois connection between the lattice $\mathbb{L}(Z_c^1(\Gamma, A))$ of all subgroups of $Z_c^1(\Gamma, A)$ and the lattice $\overline{\mathbb{L}}(\Gamma)$ of all closed subgroups of Γ. On the other hand, the continuous action of Γ on A endows the character group $\widehat{Z_c^1(\Gamma, A)} = \mathrm{Hom}(Z_c^1(\Gamma, A), \mathbb{Q}/\mathbb{Z})$ with a natural structure of a topological Γ-module, related to Γ through a canonical continuous cocycle $\eta : \Gamma \to \widehat{Z_c^1(\Gamma, A)}$ which plays a key role in the study of Kneser and co-Galois groups of cocycles. Four types of *Kummer groups of cocycles*, that are precisely the abstract group theoretic correspondents of the various types of Kummer field extensions studied in Galois theory and co-Galois theory are introduced. As in field theoretic co-Galois theory, it turns out that all of them are co-Galois groups of cocycles. The results of this section are entirely due to Albu and Basarab [15] and Albu [13].

2. Basic notation and terminology

By $\mathbb{N}$ we denote the set $\{0, 1, 2, \ldots\}$ of all natural numbers, by $\mathbb{N}^*$ the set $\mathbb{N} \setminus \{0\}$ of all strictly positive natural numbers, by $\mathbb{Z}$ the ring of all rational integers, and by $\mathbb{Q}$ (respectively $\mathbb{R}, \mathbb{C}$) the field of all rational (respectively real, complex) numbers. $\mathbb{Z}_n$ denotes the ring of rational integers modulo $n \in \mathbb{N}^*$, and $\mathbb{F}_q$ the finite field with q elements. For any $\varnothing \neq A \subseteq \mathbb{C}$ (respectively $\varnothing \neq X \subseteq \mathbb{R}$) we denote $A^* = A \setminus \{0\}$ (respectively $X_+ = \{x \in X \mid x \geqslant 0\}$. Thus, $\mathbb{Q}_+^*$ means the set of all strictly positive rational numbers. If $a \in \mathbb{R}_+^*$ and $n \in \mathbb{N}^*$, then the unique positive real root of the equation $x^n - a = 0$ will be denoted by $\sqrt[n]{a}$ and called the n-th *radical of* a; however, if $a \in \mathbb{C} \setminus \mathbb{R}_+^*$, then $\sqrt[n]{a}$ will designate a root (which in general is not specified) in $\mathbb{C}$ of the polynomial $X^n - a \in \mathbb{C}[X]$. For any set M, not necessarily finite, $|M|$ will denote the cardinal number of M.

By an *overfield* of a field F we mean any field which includes F as a subfield. A *field extension* is a pair (F, E) of fields, where F is a subfield of E (or E is an overfield of F), and in this case we shall write E/F. Very often, instead of "field extension" we shall use the shorter term "extension". If E is an overfield of a field F we will also say that E is an extension of F. Throughout this chapter F always denotes a field, $\mathrm{Char}(F)$ its characteristic, $e(F)$ its characteristic exponent (that is, $e(F) = 1$ if F has characteristic 0, and $e(F) = p$ if F has characteristic $p > 0$), and Ω a fixed algebraically closed field containing F as a subfield. Any overfield of F considered is supposed to be a subfield of Ω.

For an arbitrary nonempty subset S of Ω and a number $n \in \mathbb{N}^*$ we denote throughout this chapter:

$$S^* = S \setminus \{0\}, \qquad S^n = \big\{x^n \mid x \in S\big\}, \qquad \mu_n(S) = \big\{x \in S \mid x^n = 1\big\},$$
$$\mu(S) = \big\{x \in S \mid x^k = 1 \text{ for some } k \in \mathbb{N}^*\big\}.$$

For any $x \in \Omega^*$, $\widehat{x}$ will denote the coset xF^* of x in the quotient group Ω^*/F^*. By a *primitive n-th root of unity* we mean any generator of the cyclic group $\mu_n(\Omega)$; ζ_n will always denote such an element. When $\Omega = \mathbb{C}$, then we can choose a canonical generator of the cyclic group $\mu_n(\mathbb{C})$ of order n, and in this case ζ_n will always mean the complex number $\cos(2\pi/n) + i\sin(2\pi/n)$.

For an arbitrary multiplicative group G, the notation $H \leqslant G$ means that H is a subgroup of G. The lattice of all subgroups of G will be denoted by $\mathbb{L}(G)$. For any subset M of G, $\langle M\rangle$ will denote the subgroup of G generated by M. For any $n \in \mathbb{N}, n \geqslant 2$ we denote by $\mathcal{D}_{2n}$ the dihedral group of order $2n$. Given an action of a group C on a group D, the semidirect product of C by D is denoted by $D \rtimes C$, with a suitable subscript, if necessary, to specify the action. If S and T are topological groups, then $\overline{\mathbb{L}}(T)$ will denote the lattice of all closed subgroups of T and $\mathrm{Hom}_c(S, T)$ will denote the set of all continuous group morphisms from S to T. We denote by $\mathrm{Ch}(T)$ or by $\widehat{T}$ the *character group* of T, that is, the group $\mathrm{Hom}_c(T, \mathbb{U})$ of all continuous group morphisms of T into the unit circle $\mathbb{U} = \{z \in \mathbb{C} \mid |z| = 1\}$.

For a field extension E/F we shall denote by $[E : F]$ the *degree*, and by $\mathrm{Gal}(E/F)$ the *Galois group* of E/F. For any subgroup Δ of $\mathrm{Gal}(E/F)$, $\mathrm{Fix}(\Delta)$ will denote the fixed field of Δ. If E/F is an extension and $A \subseteq E$, then $F[A]$ will denote the smallest subring of E containing both A and F as subsets, or equivalently, the smallest F-subalgebra of E containing A as a subset. We call $F[A]$ the subring of E obtained by adjoining to F the set A, or the F-subalgebra of E generated by A. We also denote by $F(A)$ the smallest subfield of E containing both A and F as subsets. We call $F(A)$ the subfield of E obtained by adjoining to F the set A, and the extension $F(A)/F$ is called the subextension of E/F generated by A. For all other undefined terms and notation concerning basic field theory the reader is referred to Bourbaki [31], Karpilovsky [54], and/or Lang [57].

We shall also use throughout this chapter the following notation:

$$\mathbb{P} = \big\{p \in \mathbb{N}^* \mid p \text{ prime}\big\},$$
$$\mathcal{P} = \big(\mathbb{P} \setminus \{2\}\big) \cup \{4\},$$
$$\mathbb{P}_n = \{p \in \mathbb{P} \mid p \mid n\} \quad \text{for any } n \in \mathbb{N}^*,$$
$$\mathbb{D}_n = \{m \in \mathbb{N}^* \mid m \mid n\} \quad \text{for any } n \in \mathbb{N}^*,$$
$$\mathcal{P}_n = \mathcal{P} \cap \mathbb{D}_n \quad \text{for any } n \in \mathbb{N}^*.$$

If E/F is an extension, then any subfield K of E with $F \subseteq K$ is called an *intermediate field* of the extension E/F, and $\mathbb{I}(E/F)$ will denote the set of all its intermediate fields. A *subextension* (respectively *quotient extension*) of the extension E/F is any extension of the form K/F (respectively E/K), where K is an intermediate field of the extension E/F. Note that $\mathbb{I}(E/F)$ is a *poset*, that is, a partially ordered set, with respect to the partial order given by inclusion. Actually, this poset is a complete lattice, where $\inf_{i\in I} K_i = \bigcap_{i\in I} K_i$ and $\sup_{i\in I} K_i = F(\bigcup_{i\in I} K_i)$.

3. Kneser and co-Galois extensions

3.1. *G-Radical and G-Kneser extensions*

This subsection introduces two basic concepts of co-Galois theory, namely that of *G-radical* and of *G-Kneser* field extension, and presents the *Kneser criterion* [55].

For any extension E/F we shall use throughout this chapter the following notation:

$$T(E/F) := \{x \in E^* \mid x^n \in F^* \text{ for some } n \in \mathbb{N}^*\}.$$

Observe that for every element $x \in T(E/F)$ there exists an $n \in \mathbb{N}^*$ such that $x^n = a \in F$, so x is an n-th *radical* of a. Thus, $T(E/F)$ is precisely the set of all "radicals" belonging to E of elements of F^*. This observation suggests the following

DEFINITION. An extension E/F is said to be *radical* (respectively *G-radical*) if there exists a set A with $A \subseteq T(E/F)$ (respectively a group G with $F^* \leqslant G \leqslant T(E/F)$) such that $E = F(A)$ (respectively $E = F(G)$).

Clearly, one can replace the set A in definition above by the subgroup $G = F^*\langle A\rangle$ of the multiplicative group E^* of E generated by F^* and A. So, any radical extension E/F has the form $E = F(G)$, where G is a subgroup of E^* containing F^*, with G/F^* a torsion group, i.e., it is G-radical.

Recall that by a *set of representatives* of quotient group G/H we mean any subset S of G consisting of precisely one representative of each (left) coset modulo H.

PROPOSITION 3.1. ([16,21].) *The following assertions are equivalent for a G-radical extension E/F.*

(1) *There exists a set of representatives of the quotient group G/F^* which is linearly independent over F.*
(2) *Every set of representatives of G/F^* is a vector space basis of E over F.*
(3) *Every finite subset $\{x_1, \dots, x_n\} \subseteq G$ such that $\widehat{x_i} \neq \widehat{x_j}$ for each $i, j \in \{1, \dots, n\}$, $i \neq j$, is linearly independent over F.*
(4) *For every subgroup H of G such that $F^* \leqslant H$ and H/F^* finite, $|H/F^*| \leqslant [F(H) : F]$.*

DEFINITION. An extension E/F is said to be *G-Kneser* if it is a G-radical extension satisfying one of the equivalent conditions from Proposition 3.1. The extension E/F is called *Kneser* if it is G-Kneser for some group G.

Observe that if E/F is an arbitrary G-radical extension, then any set of representatives of the quotient group G/F^* is a set of generators of the F-vector space E; in particular, one has $|G/F^*| \geqslant [E : F]$. This implies that a finite G-radical extension E/F is G-Kneser if and only if $|G/F^*| = [E : F]$, and that an arbitrary G-radical extension E/F is G-Kneser if and only if for every subgroup H of G such that $F^* \leqslant H$ and H/F^* is a finite group, the finite extension $F(H)/F$ is H-Kneser; in other words, the property of a G-radical extension being G-Kneser is of finite character. Note also that for any G-Kneser extension

E/F, the extension $F(H)/F$ is H-Kneser and $F(H) \cap G = H$ for every subgroup H of G with $F^* \leqslant H$.

EXAMPLES.

(1) Let K denote the quadratic field $\mathbb{Q}(\sqrt{-3}) = \mathbb{Q}(\zeta_3)$. Then, the extension $K/\mathbb{Q}$ is clearly $\mathbb{Q}^*\langle\sqrt{-3}\rangle$-Kneser but it is not $\mathbb{Q}^*\langle\zeta_3\rangle$-Kneser since $[K : \mathbb{Q}] = |\mathbb{Q}^*\langle\sqrt{-3}\rangle/\mathbb{Q}^*| = 2 < 3 = |\mathbb{Q}^*\langle\zeta_3\rangle/\mathbb{Q}^*|$.

(2) Consider the extension $\mathbb{Q}(\sqrt[4]{-9})/\mathbb{Q}$, where $\sqrt[4]{-9}$ is one of the complex roots of the irreducible polynomial $X^4 + 9 \in \mathbb{Q}[X]$, say $\sqrt{6}(1+i)/2$. Since $(\sqrt[4]{-9})^2 = 3i$, we have $i = (\sqrt[4]{-9})^2/3 \in \mathbb{Q}(\sqrt[4]{-9})$, and so, $\sqrt{6} = 2\sqrt[4]{-9}/(1+i) \in \mathbb{Q}(\sqrt[4]{-9})$, which implies that $\mathbb{Q}(\sqrt[4]{-9}) = \mathbb{Q}(i, \sqrt{6})$. If $G = \mathbb{Q}^*\langle\sqrt[4]{-9}\rangle$ and $H = \mathbb{Q}^*\langle i, \sqrt{6}\rangle$, then $|G/\mathbb{Q}^*| = |H/\mathbb{Q}^*| = 4 = [\mathbb{Q}(\sqrt[4]{-9}) : \mathbb{Q}]$. Thus, the extension $\mathbb{Q}(\sqrt[4]{-9})/\mathbb{Q}$ is simultaneously G-Kneser and H-Kneser, but $G \neq H$ since $\sqrt{6} \in H \setminus G$. This example shows the non-uniqueness of G for a given G-Kneser extension. However, if E/F is an arbitrary extension which is simultaneously G-Kneser and H-Kneser, and $H \leqslant G$, then $G = E \cap G = F(H) \cap G = H$. The uniqueness of the group G for the so called *G-co-Galois extensions*, which are separable G-Kneser extensions possessing a certain inheritance property, will be discussed in Subsection 4.4.

(3) Let $A = \{\zeta_3, \sqrt{p_1}, \dots, \sqrt{p_n}, \dots\}$, where $p_1, \dots, p_n, \dots$ is the sequence of all positive prime numbers. Set $G = \mathbb{Q}^*\langle A\rangle$ and $E = \mathbb{Q}(A) = \mathbb{Q}(G)$. Then $|G/\mathbb{Q}^*| = [E : \mathbb{Q}] = \aleph_0$, but the G-radical extension $E/\mathbb{Q}$ is not G-Kneser, for otherwise, if $H := \mathbb{Q}^*\langle\zeta_3\rangle$, then it would follow that the quadratic extension $\mathbb{Q}(H)/\mathbb{Q}$ would be H-Kneser, which contradicts example (1) above. This shows that the characterization of finite G-radical extensions E/F being G-Kneser by the equality $|G/F^*| = [E : F]$ fails for infinite G-radical extensions.

(4) A subextension of a Kneser extension is not necessarily Kneser. Indeed, we will see in 6.2.1 that the extension $\mathbb{Q}(\sqrt{2+\sqrt{2}})/\mathbb{Q}$ is not radical, and so, it is not Kneser. On the other hand, it is easily seen that $\mathbb{Q}(\sqrt{2+\sqrt{2}}) \subseteq \mathbb{Q}(\zeta_{16})$, and $\mathbb{Q}(\zeta_{16})/\mathbb{Q}$ is a $\mathbb{Q}^*\langle\zeta_{16}\rangle$-Kneser extension. Observe that the quadratic extensions $\mathbb{Q}(\sqrt{2+\sqrt{2}})/\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ are both Kneser extensions, so the property of an extension being Kneser is, in general, not transitive.

We present now a crucial result which characterizes separable G-Kneser extensions E/F according to whether or not certain roots of unity belonging to G are in F. Originally, it has been established by Kneser in [55] only for finite extensions. The general case has been proved by Albu and Ţena [21] using the fact, mentioned above, that the property of an arbitrary G-radical extension being G-Kneser is of finite character.

THEOREM 3.2 (Kneser criterion [55,21]). *An arbitrary separable G-radical extension E/F is G-Kneser if and only if $\zeta_p \in G \Rightarrow \zeta_p \in F$ for every odd prime p and $1 \pm \zeta_4 \in G \Rightarrow \zeta_4 \in F$.*

The *separability* condition cannot be dropped from the Kneser criterion, as the following example shows. Consider the extension E/F, where $F = \mathbb{F}_2(X^2)$ and $E = \mathbb{F}_2(X)$. If $G =$

$T(E/F)$, then $G/F^* = E^*/F^*$ is a countably infinite group isomorphic to a countably infinite direct sum of copies of $\mathbb{Z}_2$. Observe that E/F is G-radical since $E = F(X)$ and $X^2 \in F$. Further, the conditions from the Kneser criterion are satisfied for $G = T(E/F)$ since $\mu_n(E) = \{1\}$ for every $n \in \mathbb{N}^*$. However, the finite extension E/F is not G-Kneser, since $2 = [E : F] < |G/F^*| = \aleph_0$.

The Kneser criterion is a powerful tool in co-Galois theory. Its immediate applications are in the following directions: (1) in establishing the equality $[\mathbb{Q}(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}) : \mathbb{Q}] = |\mathbb{Q}^*\langle \sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}\rangle/\mathbb{Q}^*|$, where $r, n_1, \ldots, n_r, a_1, \ldots, a_r \in \mathbb{N}^*$ (see Subsection 3.2); (2) in proving the Greither–Harrison criterion (see Theorem 3.3); (3) in classical algebraic number theory (see Subsections 6.3 and 6.4); (4) in investigating G-co-Galois extensions (see Subsection 4.3); (5) in Gröbner bases (see Becker, Grobe, and Niermann [29]).

3.2. *Co-Galois extensions*

Remember that for any extension E/F we use throughout this chapter the following notation:

$$T(E/F) := \{x \in E^* \mid x^n \in F^* \text{ for some } n \in \mathbb{N}^*\}.$$

Clearly $F^* \leqslant T(E/F)$, so it makes sense to consider the quotient group $T(E/F)/F^*$. Note that the quotient group $T(E/F)/F^*$ is precisely the torsion group $t(E^*/F^*)$ of the quotient group E^*/F^*. This group, playing a major role in this chapter, is somewhat dual to the Galois group of E/F, which explains the terminology below.

DEFINITIONS. The *co-Galois group* of an arbitrary field extension E/F, denoted by $\mathrm{Cog}(E/F)$, is the quotient group $T(E/F)/F^*$. The extension E/F is said to be a *co-Galois* if it is $T(E/F)$-Kneser.

Observe that a finite extension E/F is co-Galois if and only if it is radical, i.e., $E = F(T(E/F))$, and $|\mathrm{Cog}(E/F)| = [E : F]$ (just the inequality $|\mathrm{Cog}(E/F)| \leqslant [E : F]$ is sufficient). Note also that an extension E/F is co-Galois if and only if, for every subgroup H of E^* such that $F^* \leqslant H$ and H/F^* is a finite group, the finite extension $F(H)/F$ is co-Galois, in other words, the property of an arbitrary extension being co-Galois is of finite character.

To the best of our knowledge, the term of "co-Galois extension" appeared for the first time in the literature in 1986 in the fundamental paper of Greither and Harrison [45], where co-Galois extensions were introduced and investigated. A finite extension E/F is called *conormal* (respectively *coseparable*) by Greither and Harrison if $|\mathrm{Cog}(E/F)| \leqslant [E : F]$ (respectively if E/F is radical), and it is called *co-Galois* if it is both conormal and coseparable.

As it is well known, one of the characterizations of Galois extensions, finite or not, is that they are exactly those extensions which are both *normal* and *separable*. So, the Greither and Harrison terminology for finite co-Galois extensions has been chosen to agree with the dual of this characterization.

EXAMPLES.

(1) As we have already seen in Subsection 3.1, the co-Galois group of the quadratic extension $\mathbb{F}_2(X)/\mathbb{F}_2(X^2)$ is isomorphic to a countably infinite direct sum of copies of $\mathbb{Z}_2$. More generally, for any prime number $p > 0$, the extension $\mathbb{F}_p(X)/\mathbb{F}_p(X^p)$, of degree p, has co-Galois group isomorphic to a countably infinite direct sum of copies of the cyclic group $\mathbb{Z}_p$. This shows that a finite field extension may have an infinite co-Galois group.

(2) In general, the concrete calculation of the co-Galois group of a given extension is quite hard. However, for any quadratic extension $E = \mathbb{Q}(\sqrt{d})$ of $\mathbb{Q}$, with $d \neq 1$ a square-free integer, we have the following very explicit description of its co-Galois group: (i) $\operatorname{Cog}(E/\mathbb{Q}) = \widehat{\langle\sqrt{d}\rangle} \cong \mathbb{Z}_2$ if $d \neq -1, -3$; (ii) $\operatorname{Cog}(E/\mathbb{Q}) = \widehat{\langle 1+i \rangle} \cong \mathbb{Z}_4$ if $d = -1$; (iii) $\operatorname{Cog}(E/\mathbb{Q}) = \langle \widehat{i\sqrt{3}} \cdot \widehat{(1 + i\sqrt{3})} \rangle \cong \mathbb{Z}_6$ if $d = -3$. This immediately implies that $\mathbb{Q}(\sqrt{d})/\mathbb{Q}$ is a co-Galois extension if and only if $d \neq -1, -3$. In particular, the quadratic extension $\mathbb{Q}(\sqrt{-3})/\mathbb{Q}$ is not co-Galois, but it is of course $\mathbb{Q}^*\langle\sqrt{-3}\rangle$-Kneser.

(3) Using the description of the co-Galois group of a Galois extension by means of crossed homomorphisms, we will see in Corollary 5.6 that any extension E/F of algebraic number fields has a finite co-Galois group.

(4) Calculations of more co-Galois groups are provided in 6.2.8.

A basic concept in the theory of radical extensions is that of *purity*, which is somewhat related to that used in group theory: a subgroup H of an Abelian multiplicative group G is called *pure* if $G^n \cap H = H^n$ for every $n \in \mathbb{N}^*$. Recall that throughout this chapter $\mathbb{P}$ denotes the set of all positive prime numbers and $\mathcal{P} = (\mathbb{P} \setminus \{2\}) \cup \{4\}$.

DEFINITION. An extension E/F is said to be *pure* if $\mu_p(E) \subseteq F$ for every $p \in \mathcal{P}$.

Observe that any extension E/F, where E is any subfield of $\mathbb{R}$, is pure and for any field F and any $m \in \mathbb{N}^*$, the extension $F(X_1, \ldots, X_m)/F$ is pure. Also, a quadratic extension $\mathbb{Q}(\sqrt{d})/\mathbb{Q}$ where d is a square-free integer is pure if and only if $d \neq -1, -3$.

We are now going to present a criterion characterizing co-Galois extensions in terms of purity, due to Greither and Harrison [45] for finite extensions, and generalized by Albu and Ţena [21] to arbitrary extensions. The original proof in [45] involves the machinery of the cohomology of groups. A very short and simple proof, based only on the Kneser criterion, due to Albu and Ţena [21] is sketched below.

THEOREM 3.3 (Greither–Harrison criterion [45,21]). *An arbitrary extension E/F is co-Galois if and only if it is radical, separable, and pure.*

PROOF. We will prove only the implication "$\Leftarrow$", which in the original one-and-a-half-page proof of Greither and Harrison [45], given only for the case of finite extensions, uses among others the Vahlen–Capelli criterion, properties of classical Kummer extensions, an exact sequence of cohomology groups in low dimensions, and Hilbert theorem 90. So, assume that the extension E/F, finite or not, is radical, separable and pure. We are going to prove that the extension E/F is co-Galois. First of all, note that the radical extension E/F is also $T(E/F)$-radical. Let p be an odd prime with $\zeta_p \in T(E/F)$. Then $\zeta_p \in E$, hence $\zeta_p \in F$ by purity. If $1+\zeta_4 \in T(E/F)$, then $1+\zeta_4 \in E$, hence $\zeta_4 \in E$, and so, $\zeta_4 \in F$ again by purity. Now, by Kneser criterion (Theorem 3.2), we deduce that E/F is a $T(E/F)$-Kneser extension, i.e., a co-Galois extension, as desired. □

Observe that an extension E/F is pure $\Leftrightarrow \mu_p(E) = \mu_p(F), \forall p \in \mathcal{P} \Leftrightarrow \zeta_{2p} \notin E \setminus F$, $\forall p \in \mathbb{P}$. This shows that an equivalent form of the Greither–Harrison criterion is the following one.

THEOREM 3.4 (Gay–Vélez criterion [44,21]). *An arbitrary extension E/F is co-Galois if and only if it is radical, separable, and $\zeta_{2p} \notin E \setminus F$ for every $p \in \mathbb{P}$.*

The Gay–Vélez criterion, as formulated above, is an expanded reformulation of Theorem 1.7 in Gay and Vélez [44], where only the implication "$\Leftarrow$" and only for finite extensions has been proved, but the indispensable separability condition in its statement has been omitted. Chronologically, this paper appeared in 1981, so five years earlier than the one of Greither and Harrison [45].

By the Greither–Harrison criterion, any co-Galois extension is separable. However, a Kneser extension is not necessarily separable, as the following example shows: $F = \mathbb{F}_2(X^2)$, $E = \mathbb{F}_2(X)$, $G = F^*\langle X\rangle$. The extension E/F is G-Kneser since $|G/F^*| = \operatorname{ord}(\widehat{X}) = 2 = [E:F]$, but it is not separable.

EXAMPLES OF CO-GALOIS EXTENSIONS.

(1) Any finite G-radical extension E/F with E a subfield of $\mathbb{R}$ is clearly pure, hence it is co-Galois by the Greither–Harrison criterion. Note that for such an extension E/F we have $\operatorname{Cog}(E/F) = G/F^*$. For example, the extension $\mathbb{Q}(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r})/\mathbb{Q}$, with $r, n_1, \ldots, n_r, a_1, \ldots, a_r \in \mathbb{N}^*$, is a G-radical co-Galois extension, where $G = \mathbb{Q}^*\langle \sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}\rangle$, hence its co-Galois group is precisely $\mathbb{Q}^*\langle \sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}\rangle/\mathbb{Q}^*$.

(2) We have already seen that a quadratic extension $\mathbb{Q}(\sqrt{d})/\mathbb{Q}$ where $d \neq 1$ is a square-free integer is co-Galois if and only if $d \neq -1, -3$.

(3) For any odd prime $p > 0$ and any $n \in \mathbb{N}^*$, the radical extension $\mathbb{Q}(\zeta_{p^n})/\mathbb{Q}(\zeta_p)$ is pure, hence it is co-Galois by the Greither–Harrison criterion.

(4) The extension $\mathbb{Q}(\zeta_9, \sqrt[9]{5})/\mathbb{Q}(\zeta_3)$ is a Galois and co-Galois but is not Abelian.

(5) Other examples are provided in Subsection 6.2.

The next result investigates the property of an extension being co-Galois in a tower of fields.

PROPOSITION 3.5. ([45,5].) *If $F \subseteq K \subseteq E$ is a tower of fields then there exists a canonical exact sequence of Abelian groups*

$$\mathbb{1} \to \operatorname{Cog}(K/F) \to \operatorname{Cog}(E/F) \to \operatorname{Cog}(E/K).$$

Moreover, if E/F is a co-Galois extension, then E/K and K/F are both co-Galois extensions, and the groups $\operatorname{Cog}(E/K)$ *and* $\operatorname{Cog}(E/F)/\operatorname{Cog}(K/F)$ *are canonically isomorphic.*

EXAMPLE. Consider the quartic extension $\mathbb{Q}(\sqrt{1+\sqrt{2}})/\mathbb{Q}$. Of course, $\mathbb{Q}(\sqrt{2})$ is an intermediate field of this extension, and the quadratic extensions $\mathbb{Q}(\sqrt{1+\sqrt{2}})/\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ are both co-Galois, but the extension $\mathbb{Q}(\sqrt{1+\sqrt{2}})/\mathbb{Q}$ is neither radical, nor a Kneser extension, and not co-Galois (see 6.2.6). Also, $\mathrm{Cog}(\mathbb{Q}(\sqrt{1+\sqrt{2}})/\mathbb{Q}) = \{\widehat{1}, \widehat{\sqrt{2}}\}$. Observe that the element $\widehat{\sqrt{1+\sqrt{2}}}$ of the group $\mathbb{Q}(\sqrt{1+\sqrt{2}})^*/\mathbb{Q}^*$ has infinite order. This example show that the property of an extension being radical, Kneser, or co-Galois is, in general, not transitive.

We end this section by presenting some results of Barrera-Mora, Rzedowski-Calderón, and Villa-Salvador [26] about the following partial analogue to the *inverse Galois theory problem* for co-Galois extensions: given a finite group Γ and an algebraic number field F does there exist an extension E/F such that E/F is a Galois co-Galois extension with $\mathrm{Gal}(E/F) \cong \Gamma$? Note that when the given finite group Γ is Abelian, then necessarily $\mathrm{Gal}(E/F) \cong \mathrm{Cog}(E/F)$ by Corollary 5.19.

PROPOSITION 3.6. ([45,26].) *Given any finite Abelian group A, and any algebraic number field F, there exists a finite co-Galois extension E/F such that* $\mathrm{Cog}(E/F) \cong A$.

PROPOSITION 3.7. ([26].) *Let Γ be an arbitrary finite group. Then, there exists a Galois co-Galois extension $E/\mathbb{Q}$ with* $\mathrm{Gal}(E/\mathbb{Q}) \cong \Gamma$ *if and only if $\Gamma \cong (\mathbb{Z}_2)^r$ for some $r \in \mathbb{N}$.*

THEOREM 3.8. ([26].) *Let F be an algebraic number field, and let A be a finite Abelian group with $A \cong \prod_{k=1}^{r} \mathbb{Z}_{n_k}$ and $n_1|n_2|\dots|n_r$. Then, there exists a Galois co-Galois extension E/F such that* $\mathrm{Gal}(E/F) \cong A$ *if and only if $\zeta_{n_{r-1}} \in F$ and $F(\zeta_{n_r})/F$ is a pure extension.*

A more general result, dropping the Abelian condition in Theorem 3.8, was obtained again by Barrera-Mora, Rzedowski-Calderón, and Villa-Salvador in a subsequent paper [27]: given an algebraic number field F and an arbitrary allowable group Γ, they found necessary and sufficient conditions on the field F for the existence of a Galois co-Galois extension E/F with $\mathrm{Gal}(E/F) \cong \Gamma$. The concept of an *allowable group* has been introduced by Greither and Harrison [45] in connection with their investigation of the so called *neat presentations*, that we named in this chapter quasi-Kummer extensions (see 5.4.4). These groups turn out to be precisely those finite groups Γ for which the lattice $\mathbb{L}(\Gamma)$ of all subgroups of Γ has an involutive lattice anti-isomorphism $\varphi : \mathbb{L}(\Gamma) \to \mathbb{L}(\Gamma)$ with $\varphi(U) = |\Gamma|/|U|$ for all $U \leqslant \Gamma$. According to [76, Theorem 5], any allowable group Γ is nilpotent and all p-Sylow subgroups of Γ have a modular lattices of subgroups. For more facts about allowable groups and their connections with co-Galois theory, see [45,27,76].

4. Strongly Kneser and G-co-Galois extensions

4.1. *Galois and co-Galois connections*

As in [75], a *Galois connection* between the posets $(X, \leqslant)$ and $(Y, \leqslant)$ is a pair of order-reversing maps $\alpha : X \to Y$ and $\beta : Y \to X$ such that $x \leqslant (\beta \circ \alpha)(x), \forall x \in X$, and $y \leqslant (\alpha \circ \beta)(y), \forall y \in Y$, and in this case we shall use the notation

$$X \underset{\beta}{\overset{\alpha}{\rightleftarrows}} Y.$$

If the maps α and β are both order-preserving instead of order-reversing, we obtain a *co-Galois connection* between X and Y. More precisely, we have the following

DEFINITION. A *co-Galois connection* between two posets $(X, \leqslant)$ and $(Y, \leqslant)$ is a pair of order-preserving maps $\alpha : X \to Y$ and $\beta : Y \to X$ such that $(\beta \circ \alpha)(x) \leqslant x, \forall x \in X$, and $y \leqslant (\alpha \circ \beta)(y), \forall y \in Y$.

If we denote by X^{op} the opposite poset of X, then it is clear that $X \underset{\beta}{\overset{\alpha}{\rightleftarrows}} Y$ is a co-Galois connection if and only if $X^{op} \underset{\beta}{\overset{\alpha}{\rightleftarrows}} Y$ is a Galois connection.

Let $X \underset{\beta}{\overset{\alpha}{\rightleftarrows}} Y$ be a Galois or co-Galois connection. If $x \in X$ and $y \in Y$, then we shall briefly denote $x' = \alpha(x)$, $y' = \beta(y)$, $x'' = (x')'$, and $y'' = (y')'$. An element z of X or Y is said to be a *closed element* of X or Y, if $z = z''$. A closed element is also called *Galois object* (respectively *co-Galois object*) in the case of a Galois (respectively co-Galois) connection. We shall denote by $\overline{X}$ (respectively $\overline{Y}$) the set of all closed elements of X (respectively Y). Then $\overline{X} = \beta(Y)$, $\overline{Y} = \alpha(X)$, and the restrictions $\bar{\alpha} : \overline{X} \to \overline{Y}$ and $\bar{\beta} : \overline{Y} \to \overline{X}$ of α and β to the sets of closed elements of X and Y are bijections inverse to one another.

The most relevant example of a Galois connection, which actually originated the name of this concept, is the one appearing in Galois theory. Let E/F be an arbitrary field extension, and denote by Γ the Galois group $\mathrm{Gal}(E/F)$ of E/F. Then, it is easily seen that the maps

$$\alpha : \mathbb{I}(E/F) \to \mathbb{L}(\Gamma), \quad \alpha(K) = \mathrm{Gal}(E/K),$$

and

$$\beta : \mathbb{L}(\Gamma) \to \mathbb{I}(E/F), \quad \beta(\Delta) = \mathrm{Fix}(\Delta),$$

yield a Galois connection between the lattice $\mathbb{I}(E/F)$ of all intermediate fields of the extension E/F and the lattice $\mathbb{L}(\Gamma)$ of all subgroups of Γ. We will call it the *standard Galois connection associated with the extension E/F*.

One can see that a finite extension E/F with Galois group Γ is a Galois extension $\Leftrightarrow$ every intermediate field of the extension E/F is a closed element in the standard Galois connection associated with E/F $\Leftrightarrow$ F is a closed element in the standard Galois connection associated with E/F $\Leftrightarrow$ the map α is injective $\Leftrightarrow$ the map β is surjective $\Leftrightarrow$ the maps α and β establish anti-isomorphisms of lattices, inverse to one another, between the lattices $\mathbb{I}(E/F)$ and $\mathbb{L}(\Gamma)$.

The prototype of a co-Galois connection is that canonically associated with any radical extension. Let E/F be an arbitrary G-radical extension. Then, the maps

$$\varphi : \mathbb{I}(E/F) \to \mathbb{L}\big(G/F^*\big), \quad \varphi(K) = (K \cap G)/F^*,$$

and

$$\psi : \mathbb{L}(G/F^*) \to \mathbb{I}(E/F), \psi\big(H/F^*\big) = F(H),$$

establish a co-Galois connection between the lattices $\mathbb{I}(E/F)$ and $\mathbb{L}(G/F^*)$, called the *standard co-Galois connection associated with the extension E/F*. Notice that, in contrast with the standard Galois connection which can be associated with any extension, the standard co-Galois connection can be associated only with radical extensions.

The considerations above naturally lead us to define the following dual concepts. An extension E/F with Galois group Γ is said to be an *extension with Γ-Galois correspondence* if the standard Galois connection associated with E/F yields a lattice anti-isomorphism between the lattices $\mathbb{I}(E/F)$ and $\mathbb{L}(\Gamma)$. Dually, a G-radical extension E/F is said to be an *extension with G/F^*-co-Galois correspondence* if the standard co-Galois connection associated with E/F yields a lattice isomorphism between the lattices $\mathbb{I}(E/F)$ and $\mathbb{L}(G/F^*)$.

We have already seen that any finite extension E/F with Γ-Galois correspondence, where $\Gamma = \mathrm{Gal}(E/F)$, is necessarily a Galois extension. Consequently, the equality $[E : F] = |\mathrm{Gal}(E/F)|$ is a consequence of the fact that E/F is an extension with Γ-Galois correspondence. Conversely, if a finite extension E/F is such that $[E : F] = |\mathrm{Gal}(E/F)|$, then E/F is necessarily a Galois extension. This is not the case for finite extensions E/F with G/F^*-co-Galois correspondence; indeed, for such extensions, the equality $[E : F] = |G/F^*|$, saying precisely that E/F is G-Kneser, is, in general, not a consequence of the fact that E/F is an extension with G/F^*-co-Galois correspondence. We will examine more closely this situation in Subsection 4.6.

4.2. *Strongly G-Kneser extensions*

As we already noted in Subsection 3.1, a subextension of a Kneser extension is not necessarily Kneser, just as a subextension of a normal extension is not necessarily normal. So, it makes sense to consider extensions that inherit the property of being Kneser, which will be called *strongly Kneser*. It turns out that such extensions are precisely the G-Kneser extensions with G/F^*-co-Galois correspondence.

PROPOSITION 4.1. ([21].) *Let $F \subseteq K \subseteq E$ be a tower of fields, and let G be a group such that $F^* \leqslant G \leqslant E^*$. Consider the following assertions.*

(1) *K/F is $K^* \cap G$-Kneser.*
(2) *E/K is K^*G-Kneser.*
(3) *E/F is G-Kneser.*

Then, any two of the assertions (1)–(3) *imply the remaining one.*

EXAMPLE. Consider the extension $\mathbb{Q}(\sqrt[4]{-9})/\mathbb{Q}$ discussed in Subsection 3.1. We have seen that the extension $\mathbb{Q}(\sqrt[4]{-9})/\mathbb{Q}$ is $\mathbb{Q}^*\langle\sqrt[4]{-9}\rangle$-Kneser. Observe that $K = \mathbb{Q}(\sqrt{6})$ is an intermediate field of the extension $\mathbb{Q}(\sqrt[4]{-9})/\mathbb{Q}$ and $\mathbb{Q}(\sqrt[4]{-9})/K$ is not a $K^*\mathbb{Q}^*\langle\sqrt[4]{-9}\rangle$-Kneser extension. It follows that for every H with $\mathbb{Q}^* \leqslant H \leqslant \mathbb{Q}^*\langle\sqrt[4]{-9}\rangle$, $\mathbb{Q}(\sqrt{6})/\mathbb{Q}$ is not a H-Kneser extension. However, $\mathbb{Q}(\sqrt{6})/\mathbb{Q}$ is clearly $\mathbb{Q}^*\langle\sqrt{6}\rangle$-Kneser.

DEFINITION. An extension E/F is said to be *strongly G-Kneser* if it is a G-radical extension such that, for every intermediate field K of E/F, the extension E/K is K^*G-Kneser, or equivalently, the extension K/F is $K^*\cap G$-Kneser. The extension E/F is called *strongly Kneser* if it is strongly G-Kneser for some group G.

Observe that a finite extension E/F is strongly Kneser if and only if, for every intermediate field K of E/F, one has $[E : K] = |G/(K^* \cap G)|$, or equivalently, $[K : F] = |(K^* \cap G)/F^*|$. Clearly, any strongly G-Kneser extension is G-Kneser, but not conversely, as the example above shows.

We have seen in Subsection 4.1 that for any G-radical extension E/F the maps

$$\varphi : \mathbb{I}(E/F) \to \mathbb{L}\big(G/F^*\big), \quad \varphi(K) = (K \cap G)/F^*,$$
$$\psi : \mathbb{L}\big(G/F^*\big) \to \mathbb{I}(E/F), \quad \psi\big(H/F^*\big) = F(H),$$

establish a co-Galois connection between the lattices $\mathbb{I}(E/F)$ and $\mathbb{L}(G/F^*)$ which we called the standard co-Galois connection associated with E/F.

The next result gives a characterization of G-Kneser extensions E/F for which the standard associated co-Galois connection yields a bijective correspondence between the lattices $\mathbb{I}(E/F)$ and $\mathbb{L}(G/F^*)$, and is somewhat dual to the corresponding result for Galois extensions discussed in Subsection 4.1.

THEOREM 4.2. ([21].) *The following assertions are equivalent for an arbitrary G-radical extension E/F.*

(1) *E/F is strongly G-Kneser.*
(2) *E/F is G-Kneser, and the map $\psi : \mathbb{L}(G/F^*) \to \mathbb{I}(E/F)$, $H/F^* \mapsto F(H)$, is surjective.*
(3) *E/F is G-Kneser, and every element of $\mathbb{I}(E/F)$ is a closed element in the standard co-Galois connection associated with E/F.*
(4) *E/F is G-Kneser, and the map $\varphi : \mathbb{I}(E/F) \to \mathbb{L}(G/F^*)$, $K \mapsto (K \cap G)/F^*$, is injective.*
(5) *E/F is G-Kneser with G/F^*-co-Galois correspondence, i.e., the maps*

$$\varphi : \mathbb{I}(E/F) \to \mathbb{L}\big(G/F^*\big) \quad \textit{and} \quad \psi : \mathbb{L}\big(G/F^*\big) \to \mathbb{I}(E/F)$$

defined above are isomorphisms of lattices, inverse to one another.

The property of an extension being strongly Kneser behaves nicely with respect to subextensions and quotient extensions, that is, if E/F is a strongly G-Kneser extension, then, for any intermediate field K of E/F, K/F is strongly $K^* \cap G$-Kneser and E/K is strongly K^*G-Kneser.

4.3. *G-co-Galois extensions*

In this subsection we introduce the basic concept of a *G-co-Galois extension*, which plays in co-Galois theory the same role as that of Galois extension in Galois theory. A G-co-Galois extension is nothing else than a separable G-Kneser extension with G/F^*-co-Galois correspondence. These extensions can be nicely characterized within the class of G-radical extensions by means of a certain kind of "purity". Using this characterization, we will see that the class of G-co-Galois extensions is large enough, including important classes of field extensions: co-Galois extensions, classical Kummer extensions, Kummer extensions with few roots of unity, generalized Kummer extensions, and quasi-Kummer extensions.

Recall that $\mathbb{P}$ denotes the set of all strictly positive prime numbers, that $\mathcal{P}$ is the set $(\mathbb{P} \setminus \{2\}) \cup \{4\}$, $\mathbb{D}_n$ is the set of all positive divisors of a given number $n \in \mathbb{N}^*$, and $\mathcal{P}_n$ is the set $\mathcal{P} \cap \mathbb{D}_n$. Also, recall from Subsection 3.2 that an extension E/F is called *pure* when $\mu_p(E) \subseteq F$ for all $p \in \mathcal{P}$. More generally, if $\mathcal{Q}$ is a nonempty subset of $\mathcal{P}$, we say that an extension E/F is *$\mathcal{Q}$-pure* if $\mu_p(E) \subseteq F$, or equivalently, $\mu_p(E) = \mu_p(F)$ for all $p \in \mathcal{Q}$. If $n \in \mathbb{N}^*$, then an extension E/F is called *n-pure* if it is $\mathcal{P}_n$-pure, i.e., if $\mu_p(E) \subseteq F$ for all p, p odd prime or 4, with $p \mid n$. Clearly, an extension E/F is pure if and only if it is n-pure for every $n \in \mathbb{N}^*$. Also, note that an n-pure extension is not necessarily pure; e.g., $\mathbb{Q}(\sqrt{-3})/\mathbb{Q}$ is 2-pure but not pure.

For any torsion multiplicative group T with identity element e we will denote by $\mathcal{O}_T$ the set $\{\mathrm{ord}(x) \mid x \in T\}$ of all orders of elements of T. When the subset $\mathcal{O}_T$ of $\mathbb{N}$ is a bounded set, or equivalently, a finite set, one says that the torsion group T is a *group of bounded order*, and the least number $n \in \mathbb{N}^*$ with the property that $T^n = \{e\}$ is the *exponent* $\exp(T)$ of T. The group T is said to be *n-bounded* if T is a group of bounded order and $\exp(T) = n$.

Observe that for any G-radical extension E/F, which is not necessarily finite, the group G/F^* is a torsion Abelian group, so it makes sense to consider the set $\mathcal{O}_{G/F^*}$ of natural numbers.

DEFINITION. A G-radical extension E/F is said to be a *bounded extension* if G/F^* is a group of bounded order; in this case, if $\exp(G/F^*) = n$, we say that E/F is an *n-bounded extension.*

In the class of field extensions E/F with G/F^*-co-Galois correspondence the most interesting are those extensions which additionally are separable. In view of the equivalence (1) $\Leftrightarrow$ (5) in Theorem 4.2, these are precisely the separable strongly G-Kneser extensions. They deserve a special name.

DEFINITION. An extension E/F is called *G-co-Galois* if it is a separable strongly G-Kneser extension.

For any G-radical extension E/F write $\mathcal{P}_G := \mathcal{P} \cap \mathcal{O}_{G/F^*}$. The next result characterizes the G-co-Galois extensions in terms of $\mathcal{P}_G$-purity.

THEOREM 4.3 (General purity criterion [5]). *A separable G-radical extension is G-co-Galois if and only if it is $\mathcal{P}_G$-pure.*

COROLLARY 4.4. ([5].) *Any co-Galois extension E/F is $T(E/F)$-co-Galois.*

PROOF. In view of the Greither–Harrison criterion (Theorem 3.3), the radical extension E/F is separable and pure, so, a fortiori, $\mathcal{P}_{T(E/F)}$-pure. Thus, E/F is $T(E/F)$-co-Galois by Theorem 4.3. □

Note that when a G-radical extension E/F is n-bounded, then $\mathcal{O}_{G/F^*} = \mathbb{D}_n$, hence $\mathcal{P}_G = \mathcal{P} \cap \mathbb{D}_n = \mathcal{P}_n$, and so, we obtain the following particular case of Theorem 4.3.

COROLLARY 4.5 (n-Purity criterion [16,21]). *A separable n-bounded G-radical extension E/F is G-co-Galois if and only if it is n-pure. In particular, a finite separable G-radical extension E/F with* $\exp(G/F^*) = n$ *is G-co-Galois if and only if it is n-pure.*

REMARKS.

(1) If we combine Theorems 4.2 and 4.3, we deduce that a separable G-radical extension is G-Kneser with G/F^*-co-Galois correspondence if and only if it is $\mathcal{P}_G$-pure. The condition "E/F is G-Kneser" is essential in this statement as the following example shows: the G-radical extension $\mathbb{Q}(\zeta_3)/\mathbb{Q}$, where $G = \mathbb{Q}^*\langle\zeta_3\rangle$, is an extension with $G/\mathbb{Q}^*$-co-Galois correspondence which is not G-co-Galois.

(2) A strongly G-Kneser extension is not necessarily separable. Indeed, if $F = \mathbb{F}_2(X^2)$, $E = \mathbb{F}_2(X)$, $G = F^*\langle X\rangle$, then the extension E/F is G-Kneser since $|G/F^*| = \operatorname{ord}(\widehat{X}) = 2 = [E : F]$, so also strongly G-Kneser, but it is not separable.

The property of an extension being G-co-Galois behaves nicely with respect to subextensions and quotient extensions, that is, if E/F is a G-co-Galois extension, then, for every intermediate field K of E/F, K/F is $K^* \cap G$-co-Galois and E/K is K^*G-co-Galois. Note also that the property of an extension being G-co-Galois, like that of being G-Kneser or co-Galois, is of finite character, i.e., a G-radical extension E/F is G-co-Galois if and only if every finite subextension K/F of E/F is $K^* \cap G$-co-Galois.

We end this subsection by presenting the class, introduced by Vélez [80], of field extensions having the unique subfield property and its connection with the class of G-co-Galois extensions. A finite extension E/F is said to have the *unique subfield property*, abbreviated USP, if for every divisor m of $[E : F]$ there exists a unique intermediate field K of E/F such that $[K : F] = m$. Clearly, a finite extension E/F of degree n has the USP if and only if the canonical map $\mathbb{I}(E/F) \to \mathbb{D}_n$, $K \mapsto [K : F]$, is a lattice isomorphism. The next result characterizes the extensions having the USP in the class of finite G-co-Galois extensions.

PROPOSITION 4.6. ([8].) *A finite G-co-Galois extension E/F of degree n has the* USP *if and only if the group G/F^* is cyclic* (*of order n*). *In particular, a finite co-Galois extension E/F of degree n has the* USP *if and only if the co-Galois group* $\operatorname{Cog}(E/F)$ *of E/F is cyclic* (*of order n*).

The next result, due Acosta de Orozco and Vélez [1] characterizes separable simple radical extensions having the USP. A simple proof of it, via basic facts of co-Galois theory, can be found in [8].

PROPOSITION 4.7. ([1].) *Let F be any field, and let $u \in \Omega$ be a root of an irreducible binomial $X^n - a \in F[X]$ with* $\gcd(n, e(F)) = 1$. *The extension $F(u)/F$ has the* USP *if and only if the following two conditions are satisfied.*

(a) $\zeta_p \notin F(u) \setminus F$ *for every odd prime divisor p of n.*

(b) *If* $4 \mid n$, *then* $\zeta_4 \notin F(u) \setminus F$.

PROPOSITION 4.8. ([8].) *Let F be any field, and let $u \in \Omega$ be a root of an irreducible binomial $X^n - a \in F[X]$ with* $\gcd(n, e(F)) = 1$. *Then, the following assertions are equivalent.*

(1) *The extension $F(u)/F$ has the* USP.

(2) *The extension $F(u)/F$ is n-pure.*

(3) *The extension $F(u)/F$ is $F^*\langle u\rangle$-co-Galois.*

(4) *The extension $F(u)/F$ is G-co-Galois for some group G, and G/F^* is a cyclic group.*

COROLLARY 4.9. ([80].) *Let F be an arbitrary field, let $n \in \mathbb{N}^*$ be such that* $\gcd(n, e(F)) = 1$, *and let $X^n - a, X^n - b$ be irreducible polynomials in $F[X]$ with roots $u, v \in \Omega$, respectively. If the extension $F(u)/F$ has the* USP, *then the following assertions are equivalent.*

(1) *The fields $F(u)$ and $F(v)$ are F-isomorphic.*

(2) *There exist $c \in F$ and $j \in \mathbb{N}$ with* $\gcd(j, n) = 1$ *and $a = b^j c^n$.*

EXAMPLES.

(1) Any extension of degree a prime number has clearly the USP.

(2) A finite G-radical extension which has USP is not necessarily G-co-Galois. Indeed the quadratic extension $\mathbb{Q}(\sqrt{-3})/\mathbb{Q}$ is not $\mathbb{Q}^*\langle\zeta_3\rangle$-Kneser, so it is not $\mathbb{Q}^*\langle\zeta_3\rangle$-co-Galois either, but it has the USP. Note that the quartic extension $\mathbb{Q}(\sqrt{1+\sqrt{2}})/\mathbb{Q}$ has precisely one quadratic intermediate field, so it has the USP, but it is neither a radical nor a co-Galois extension (see 6.2.6). Also, any cyclic Galois extension $E/\mathbb{Q}$ of degree > 2 is not G-co-Galois, but has the USP.

(3) A finite G-co-Galois extension may fail to have the USP. Indeed the quartic co-Galois extension $\mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q}$ does not have the USP.

4.4. *The Kneser group of a G-co-Galois extension*

This subsection essentially shows that a separable G-Kneser extension is G-co-Galois if and only if the group G has a prescribed structure. This implies that the group G/F^* of any G-co-Galois extension E/F is uniquely determined; it is called the *Kneser group* of E/F and denoted by $\mathrm{Kne}(E/F)$.

LEMMA 4.10. ([5].) *Let E/F be a G-co-Galois extension, and let $x \in E^*$ be such that $x^m \in F$ for some $m \in \mathbb{N}^*$. Suppose that one of the following two conditions is satisfied.*

(a) $\mathcal{P}_m \subseteq \mathcal{P}_G$ *(in particular, this holds if E/F is n-bounded and $m \mid n$).*

(b) $\mu_m(E) \subseteq F$ *(in particular, this holds if $\zeta_m \in F$).*

Then, we have $F(x) \subseteq E \Leftrightarrow x \in G$.

If A is an Abelian group, then for every $p \in \mathbb{P}$ we denote by $t_p(A)$ the p-primary component of A. For any extension E/F we denote by $\mathrm{Cog}_2(E/F)$ the subgroup of $\mathrm{Cog}(E/F)$ consisting of all its elements of order $\leqslant 2$. The notation $A = \bigoplus_{i \in I} A_i$ below means that the Abelian group A is the *internal direct sum* of an independent family $(A_i)_{i \in I}$ of its subgroups, that is, any element $x \in A$ can be uniquely expressed as $x = \sum_{i \in I} x_i$, where $(x_i)_{i \in I}$ is a family of finite support, with $x_i \in A_i$ for every $i \in I$.

Using Lemma 4.10, the Kneser criterion, as well as the general purity criterion one obtains the next result characterizing, via the structure of the torsion Abelian group G/F^*, the G-co-Galois extensions among the separable G-Kneser extensions.

THEOREM 4.11. ([5].) *The following statements hold for a separable G-Kneser extension E/F.*

(1) *Assume that* $4 \in \mathcal{P}_G$. *Then E/F is G-co-Galois if and only if*

$$G/F^* = \Big(\bigoplus_{p \in \mathcal{P}_G \setminus \{4\}} t_p\big(\mathrm{Cog}(E/F)\big) \Big) \bigoplus t_2\big(\mathrm{Cog}(E/F)\big).$$

(2) *Assume that* $4 \notin \mathcal{P}_G$. *Then E/F is G-co-Galois if and only if*

$$G/F^* = \Big(\bigoplus_{p \in \mathcal{P}_G} t_p\big(\mathrm{Cog}(E/F)\big) \Big) \bigoplus \mathrm{Cog}_2(E/F).$$

COROLLARY 4.12. ([16,21].) *Let E/F be a separable n-bounded G-Kneser extension, in particular, a finite separable G-Kneser extension with* $\exp(G/F^*) = n$.

(1) *Suppose that* $n \equiv 2 \pmod 4$. *Then E/F is a G-co-Galois extension if and only if*

$$G/F^* = \bigoplus_{p \in \mathbb{P}_n} t_p\big(\mathrm{Cog}(E/F)\big).$$

(2) *Suppose that* $n \equiv 2 \pmod 4$. *Then E/F is a G-co-Galois extension if and only if*

$$G/F^* = \Big(\bigoplus_{p \in \mathbb{P}_n \setminus \{2\}} t_p\big(\mathrm{Cog}(E/F)\big) \Big) \bigoplus \mathrm{Cog}_2(E/F).$$

The condition "E/F is a G-Kneser extension" in Theorem 4.11 and Corollary 4.12 cannot be dropped since, otherwise, E/F may not be G-co-Galois. Indeed, the quadratic G-radical extension $\mathbb{Q}(\zeta_3)/\mathbb{Q}$, with $G = \mathbb{Q}^*\langle \zeta_3 \rangle$ is not G-co-Galois, but $G/\mathbb{Q}^* = t_3(\mathrm{Cog}(\mathbb{Q}(\zeta_3)/\mathbb{Q}))$ and $|G/\mathbb{Q}^*| = \exp(G/\mathbb{Q}^*) = 3$.

Note that if E/F is a finite G-co-Galois extension with $\exp(G/F^*) \equiv 2 \pmod 4$, then we may have $t_2(G/F^*) \neq t_2(\mathrm{Cog}(E/F))$. Indeed, the extension $\mathbb{Q}(i, \sqrt{6})/\mathbb{Q}$ is $\mathbb{Q}^*\langle i, \sqrt{6} \rangle$-co-Galois, but $\sqrt{6}(1+i)/2 \in t_2(\mathrm{Cog}(\mathbb{Q}(i, \sqrt{6})/\mathbb{Q})) \setminus t_2(\mathbb{Q}^*\langle i, \sqrt{6} \rangle/\mathbb{Q}^*)$.

THEOREM 4.13. ([16,21].) *Let E/F be an extension which is simultaneously G-co-Galois and H-co-Galois. Then $G = H$.*

PROOF. Let $x \in G$ be arbitrary, and set $K = F(x)$, $G' = F^*\langle x\rangle$. Since the extension E/F is H-co-Galois it follows by Theorem 4.2 that $K = F(H')$ for some H' with $F^* \leqslant H' \leqslant H$. Then, the finite extension K/F is simultaneously $K^* \cap G$-co-Galois and $K^* \cap H$-co-Galois. But $K^* \cap G = F(G') \cap G = G'$ and $K^* \cap H = F(H') \cap H = H'$. So, the finite extension K/F is simultaneously G'-co-Galois and H'-co-Galois. Therefore, if the result holds for finite extensions it follows that $G' = H'$, and then, since $x \in G'$, we deduce that $x \in H' \subseteq H$. Hence $G \subseteq H$. The proof of the inverse inclusion $H \subseteq G$ is similar.

Thus, without loss of generality, we may assume that the given extension E/F is finite. Set $m = \exp(G/F^*)$, $n = \exp(H/F^*)$, and $k = [E : F]$. Then $|G/F^*| = |H/F^*| = [E : F] = k$. On the other hand, since the order and the exponent of any finite Abelian group have the same prime divisors, we have $\mathbb{P}_m = \mathbb{P}_n = \mathbb{P}_k$, and hence, by Corollary 4.12, it is sufficient to prove only that $4 \mid m \Leftrightarrow 4 \mid n$.

Assume that $4 \mid m$. Then G/F^* contains an element of order 4, say $\widehat{g}$. Set $G_1 = F^*\langle g\rangle$ and $E_1 = F(G_1)$. By Theorem 4.2, there exists H_1 such that $F^* \leqslant H_1 \leqslant H$, $E_1 = F(H_1)$, and $|H_1/F^*| = 4$. Note that E_1/F is an $E_1^* \cap G$-co-Galois extension and $E_1^* \cap G = E_1 \cap G = F(G_1) \cap G = G_1$, hence E_1/F is a G_1-co-Galois extension. Then, using the lattice isomorphism (provided by Theorem 4.2) between the lattice $\mathbb{I}(E_1/F)$ of all intermediate fields of E_1/F and the lattice $\mathbb{L}(G_1/F^*)$ of all subgroups of the cyclic group G_1/F^* of order 4, one deduces that the extension E_1/F has only one proper intermediate field. On the other hand, one shows similarly that the extension E_1/F is H_1-co-Galois. Now, using the lattice isomorphism between the lattices $\mathbb{I}(E_1/F)$ and $\mathbb{L}(H_1/F^*)$, one deduces that the group H_1/F^* of order 4 is necessarily cyclic, and then, $4 \mid n$, as desired. □

In view of Theorem 4.13, the group G of any G-co-Galois extension, finite or not, is uniquely determined. So, it makes sense to introduce the following concept.

DEFINITION. If E/F is a G-co-Galois extension, then the group G/F^* is called the *Kneser group* of the extension E/F and is denoted by $\mathrm{Kne}(E/F)$.

4.5. *Primitive elements for G-co-Galois extensions*

By a well known result, any finite separable extension E/F has a primitive element, i.e., an element $u \in E$ such that $E = F(u)$. However, in general, there is no practical procedure to find such a u effectively for a given extension E/F. In this subsection we show that for a fairly large class of finite separable extensions, namely, for G-co-Galois extensions, it is possible to provide a simple and easily manageable method of finding primitive elements. In particular, it follows that for any set of representatives S of the (finite) Kneser group G/F^* of any finite G-co-Galois extension E/F, the sum $\sum_{s\in S} s$ is a primitive element of E/F.

THEOREM 4.14. ([17].) *Let E/F be a finite G-co-Galois extension, let $n \in \mathbb{N}^*$, and let $(x_i)_{1 \leqslant i \leqslant n}$ be a finite family of elements of G such that $\widehat{x_i} \neq \widehat{x_j}$ for every $i, j \in \{1, \dots, n\}, i \neq j$. Then $x_1 + \dots + x_n$ is a primitive element of E/F if and only if $G = F^*\langle x_1, \dots, x_n \rangle$. In particular, if $\{u_1, \dots, u_r\}$ is any set of representatives of G/F^*, then $u_1 + \dots + u_r$ is a primitive element of the extension E/F.*

LEMMA 4.15. ([17].) *Let E/F be an extension, let $n \in \mathbb{N}^*$, and let $x_1, \dots, x_n \in E^*$. If $F(x_1, \dots, x_n)/F$ is an $F^*\langle x_1, \dots, x_n \rangle$-Kneser extension, then the following assertions are equivalent.*

(1) $[F(x_1, \dots, x_n) : F] = \prod_{i=1}^{n} [F(x_i) : F]$.

(2) $|F^*\langle x_1, \dots, x_n \rangle / F^*| = \prod_{1 \leqslant i \leqslant n} |F^*\langle x_i \rangle / F^*|$.

(3) *The groups* $F^*\langle x_1, \dots, x_n \rangle / F^*$ *and* $\prod_{1 \leqslant i \leqslant n} (F^*\langle x_i \rangle / F^*)$ *are isomorphic.*

(4) *If* $i_1, \dots, i_n \in \mathbb{N}$, *then* $x_1^{i_1} \cdot \dots \cdot x_n^{i_n} \in F^* \Rightarrow x_k^{i_k} \in F^*$ *for every* k, $1 \leqslant k \leqslant n$.

PROPOSITION 4.16. ([17].) *Let $n \in \mathbb{N}^*$, and let $x_1, \dots, x_n \in \Omega^*$ be such that the extension $F(x_1, \dots, x_n)/F$ is $F^*\langle x_1, \dots, x_n \rangle$-co-Galois. If $[F(x_1, \dots, x_n) : F] = \prod_{i=1}^{n} [F(x_i) : F]$, then*

$$F(x_1, \dots, x_n) = F(x_1 + \dots + x_n).$$

4.6. *Almost G-co-Galois extensions*

An analogue of Theorem 5.3 of Greither and Pareigis [46] prompted Albu and Nicolae to raise in [16] the following problem: if E/F is a separable finite G-radical extension which is not G-co-Galois, but is an extension with G/F^*-co-Galois correspondence, then does there exist another group $\widetilde{G}$ such that E/F is $\widetilde{G}$-co-Galois? This problem was solved in *negative* by Lam-Estrada, Barrera-Mora, and Villa-Salvador in [56, Section 4]. They introduced the concept of a *pseudo G-co-Galois extension*, which is precisely our concept of finite *almost G-co-Galois extension* defined below. Recall from Subsection 4.1 that a G-radical extension E/F is said to be an extension with G/F^*-co-Galois correspondence if the standard co-Galois connection associated with E/F yields a lattice isomorphism between the lattices $\mathbb{I}(E/F)$ and $\mathbb{L}(G/F^*)$.

DEFINITION. An extension E/F is said to be *almost G-co-Galois* if it is a separable G-radical extension with G/F^*-co-Galois correspondence. A *strictly almost G-co-Galois extension* is an almost G-co-Galois extension which is not G-co-Galois.

Observe that a G-co-Galois extension is precisely an almost G-co-Galois extension which is also G-Kneser. The property of a radical extension being almost G-co-Galois is inherited by subextensions and quotient extensions, i.e., if E/F is an almost G-co-Galois extension, then for every $K \in \mathbb{I}(E/F)$, K/F is an almost $K \cap G$-co-Galois extension, and E/K is an almost K^*G-co-Galois extension.

A technical result due to Lam-Estrada, Barrera-Mora, and Villa-Salvador [56] characterizes those finite strictly almost G-co-Galois extensions that are $\widetilde{G}$-co-Galois for some group $\widetilde{G}$. If one applies this characterization to algebraic number fields one obtains the following result.

PROPOSITION 4.17. ([56].) *Let E be an algebraic number field such that the extension $E/\mathbb{Q}$ is strictly almost G-co-Galois. Then $E/\mathbb{Q}$ is $\widetilde{G}$-co-Galois for some group $\widetilde{G}$ if and only if there exists an intermediate field K of $E/\mathbb{Q}$ such that $K/\mathbb{Q}$ is co-Galois*, *with* $\gcd(|\operatorname{Cog}(K/\mathbb{Q})|, 6) = 1$, $E = K(\zeta_3)$, *and*

$$G/\mathbb{Q}^* = \operatorname{Cog}(K/\mathbb{Q}) \oplus \big(\mathbb{Q}^*\langle\zeta_3\rangle/\mathbb{Q}^*\big).$$

In this case, *we have*

$$\widetilde{G}/\mathbb{Q}^* = \operatorname{Cog}(K/\mathbb{Q}) \oplus \big(\mathbb{Q}^*\big\langle\sqrt{-3}\big\rangle/\mathbb{Q}^*\big).$$

The next examples are due to Lam-Estrada, Barrera-Mora, and Villa-Salvador [56].

EXAMPLES.

(1) The extension $\mathbb{Q}(\zeta_3, \sqrt[3]{2})/\mathbb{Q}$ is strictly almost G-co-Galois, where $G = \mathbb{Q}^*\langle\zeta_3, \sqrt[3]{2}\rangle$.

(2) Let $a, p \in \mathbb{P}$, $p \geqslant 5$, and let $F = \mathbb{Q}$, $E = \mathbb{Q}(\zeta_3, \sqrt[p]{a})$, and $G = \mathbb{Q}^*\langle\zeta_3, \sqrt[p]{a}\rangle$. Then, the extension $E/\mathbb{Q}$ is strictly almost G-co-Galois and $\widetilde{G}$-co-Galois, where $\widetilde{G}/\mathbb{Q}^* = (\mathbb{Q}^*\langle\sqrt{-3}\rangle/\mathbb{Q}^*) \oplus (\mathbb{Q}^*\langle\sqrt[p]{a}\rangle/\mathbb{Q}^*)$. This shows that there exist infinitely many number fields satisfying the conditions in Proposition 4.17.

(3) Let $p \in \mathbb{P}$, and let $q = p^r \geqslant 3$, $r \in \mathbb{N}^*$. One can choose $t, l \in \mathbb{P}$ such that $l \mid q^t - 1$ and $\gcd(l, q-1) = \gcd(t, q-1) = 1$. Then, the extension $\mathbb{F}_{q^t}/\mathbb{F}_q$ is strictly almost G-co-Galois, where $G = \mathbb{F}_q^*\langle\zeta_l\rangle$, but is not $\widetilde{G}$-co-Galois for any group $\widetilde{G}$.

(4) Let $E = \mathbb{Q}(\zeta_3, \sqrt[3]{2})$, $F = \mathbb{Q}$, and $G = \mathbb{Q}^*\langle\zeta_3, \sqrt[3]{2}\rangle$. Then E/F is a strictly almost G-co-Galois extension, but there exists no group $\widetilde{G}$ such that E/F is $\widetilde{G}$-co-Galois.

5. Galois *G*-co-Galois extensions

5.1. *G-Radical Galois extensions*

In this subsection we characterize G-radical extensions, not necessarily finite, which are separable or Galois.

LEMMA 5.1. ([7].) *A G-radical extension E/F is separable if and only if* $\gcd(m, e(F)) = 1$ *for every $m \in \mathcal{O}_{G/F^*}$.*

PROPOSITION 5.2. ([7].) *A G-radical extension E/F is a Galois extension if and only if* $\gcd(m, e(F)) = 1$ *and $\zeta_m \in E, \forall m \in \mathcal{O}_{G/F^*}$.*

Observe that if E/F is an n-bounded G-radical extension, then $\mathcal{O}_{G/F^*} = \mathbb{D}_n$, and so, we obtain the next two results.

COROLLARY 5.3. ([21].) *Let E/F be an n-bounded G-radical extension. Then E/F is a Galois extension if and only if* $\gcd(n, e(F)) = 1$ *and $\zeta_n \in E$.*

COROLLARY 5.4. ([16].) *Let E/F be a finite G-radical extension with G/F^* a finite group of exponent n. Then E/F is a Galois extension if and only if* $\gcd(n, e(F)) = 1$ *and* $\zeta_n \in E$.

5.2. *Galois extensions and crossed homomorphisms*

In this subsection we investigate Galois extensions E/F by means of continuous crossed homomorphisms of $\mathrm{Gal}(E/F)$ with coefficients in the group $\mu(E)$ of all roots of unity of E. As an application of this result, one deduces very easily that the co-Galois group of any extension of algebraic number fields is a finite group. We also describe the Kneser group $\mathrm{Kne}(E/F)$ of any Galois G-co-Galois extension E/F via continuous crossed homomorphisms.

Recall first only those basic facts on *Galois cohomology* which will be used in the sequel. Let E/F be an arbitrary extension with Galois group Γ, and let $M \leqslant E^*$ be such that $\sigma(M) \subseteq M$ for every $\sigma \in \Gamma$. A *crossed homomorphism* (or a 1-*cocycle*) of Γ with coefficients in M is a map $f : \Gamma \to M$ satisfying the condition $f(\sigma\tau) = f(\sigma) \cdot \sigma(f(\tau))$ for every $\sigma, \tau \in \Gamma$. The set of all crossed homomorphisms of Γ with coefficients in M is an Abelian group, which will be denoted by $Z^1(\Gamma, M)$. For every $\alpha \in M$ we shall denote by f_α the 1-*coboundary* $f_\alpha : \Gamma \to M$, defined as $f_\alpha(\sigma) = \sigma(\alpha) \cdot \alpha^{-1}, \sigma \in \Gamma$. The set $B^1(\Gamma, M) = \{f_\alpha \mid \alpha \in M\}$ is a subgroup of $Z^1(\Gamma, M)$. The quotient group $Z^1(\Gamma, M)/B^1(\Gamma, M)$ is called the first *cohomology group* of Γ with coefficients in M, and is denoted by $H^1(\Gamma, M)$. Note that for any group G, for any G-module A, and for any $n \in \mathbb{N}$ one can define the more general concept of n-th *cohomology group* $H^n(G, A)$ of G with coefficients in A (see, e.g., Cassels and Fröhlich [34, Chapter IV] or Karpilovsky [54, p. 369]). The famous *Hilbert theorem 90* asserts that $H^1(\mathrm{Gal}(E/F), E^*) = \mathbb{1}$, for any finite Galois extension E/F (see, e.g., Cassels and Fröhlich [34, Chapter V, Proposition 2.2] or Karpilovsky [54, Chapter 6, Theorem 9.2]).

In case the Galois extension E/F is infinite, then an *infinite* Hilbert theorem 90 still holds. As it is well known, the Galois group Γ of E/F is a profinite group, or equivalently, a Hausdorff, compact, and totally disconnected topological group with respect to its *Krull topology*. A fundamental system of open neighborhoods of the identity element 1_E of Γ consists of normal subgroups of Γ of finite index, that is, of subgroups of the form $\mathrm{Gal}(E/N)$, with N/F is a finite normal subextension of E/F. As above, let $M \leqslant E^*$ be such that $\sigma(M) \subseteq M$ for every $\sigma \in \Gamma$. A *continuous crossed homomorphism* or a *continuous* 1-*cocycle* of Γ with coefficients in M is a continuous function $f \in Z^1(\Gamma, M)$, where M is endowed with the discrete topology. The set of all continuous crossed homomorphisms of Γ with coefficients in M is a subgroup of $Z^1(\Gamma, M)$, and will be denoted in the sequel by $Z^1_c(\Gamma, M)$, where the subscript "c" stands for "continuous". Observe that for every $\alpha \in M$, the 1-coboundary $f_\alpha : \Gamma \to M$, $f_\alpha(\sigma) = \sigma(\alpha) \cdot \alpha^{-1}, \sigma \in \Gamma$, is a continuous map. Consequently, the set $B^1(\Gamma, M) = \{f_\alpha \mid \alpha \in M\}$ of all 1-coboundaries of Γ with coefficients in M coincides with the set $B^1_c(\Gamma, M)$ of all *continuous* 1-coboundaries of Γ with coefficients in M. The quotient group $Z^1_c(\Gamma, M)/B^1_c(\Gamma, M)$ is called the first *continuous cohomology group* of Γ with coefficients in M, and is denoted by $H^1_c(\Gamma, M)$. Note that when E/F is a finite Galois extension, then $Z^1_c(\Gamma, M) = Z^1(\Gamma, M)$, hence

$H^1(\Gamma, M) = H^1_c(\Gamma, M)$. The *continuous* (or *infinite*) *Hilbert theorem 90* asserts that if E/F is an arbitrary Galois extension, finite or infinite, then $H^1_c(\Gamma, E^*) = \mathbb{1}$ (see, e.g., Cassels and Fröhlich [34, Chapter V, Proposition 2.2] or Serre [72, Chapitre 2, Proposition 1]).

Recall that for any extension E/F used in this chapter there is the following notation:

$$\begin{aligned}
&\mu(E) = \{x \in E^* \mid x^n = 1 \text{ for some } n \in \mathbb{N}^*\},\\
&T(E/F) = \{x \in E^* \mid x^n \in F^* \text{ for some } n \in \mathbb{N}^*\},\\
&\mathrm{Cog}(E/F) = T(E/F)/F^*,\\
&\widehat{x} = \text{the coset } xF^* \in E^*/F^* \text{ of any } x \in E.
\end{aligned}$$

For an arbitrary extension E/F we consider the following map

$$\begin{aligned}
&f : \mathrm{Gal}(E/F) \times \mathrm{Cog}(E/F) \to \mu(E),\\
&\quad f(\sigma, \widehat{\alpha}) = f_\alpha(\sigma) = \sigma(\alpha) \cdot \alpha^{-1}, \ \sigma \in \mathrm{Gal}(E/F), \alpha \in T(E/F).
\end{aligned}$$

For every fixed $\sigma \in \mathrm{Gal}(E/F)$, the partial map $f(\sigma, -)$ is multiplicative on $\mathrm{Cog}(E/F)$, and for every fixed $\widehat{\alpha} \in \mathrm{Cog}(E/F)$, the partial map $f(-, \widehat{\alpha})$ is precisely the 1-coboundary $f_\alpha \in Z^1(\mathrm{Gal}(E/F), \mu(E))$, so f induces the group morphism

$$\psi : \mathrm{Cog}(E/F) \to Z^1(\mathrm{Gal}(E/F), \mu(E)), \quad \psi(\widehat{\alpha}) = f_\alpha.$$

The Galois group $\mathrm{Gal}(E/F) = \Gamma$ is a profinite group, in particular it is a topological group with respect to its Krull topology. The morphism ψ has its image in the corresponding group of continuous crossed homomorphisms since $\psi(\widehat{\alpha}) = f_\alpha \in B^1(\Gamma, \mu(E)) \subseteq Z^1_c(\Gamma, \mu(E))$. Consequently, we can assume that the canonical map ψ is defined as follows:

$$\psi : \mathrm{Cog}(E/F) \to Z^1_c(\mathrm{Gal}(E/F), \mu(E)), \quad \psi(\widehat{\alpha}) = f_\alpha.$$

THEOREM 5.5. ([39,59,26].) *For any Galois extension E/F, the map $\widehat{\alpha} \mapsto f_\alpha$ establishes a group isomorphism*

$$\mathrm{Cog}(E/F) \cong Z^1_c(\mathrm{Gal}(E/F), \mu(E)).$$

In particular, if E/F is a finite Galois extension, then the groups $\mathrm{Cog}(E/F)$ *and* $Z^1(\mathrm{Gal}(E/F), \mu(E))$ *are canonically isomorphic.*

COROLLARY 5.6. ([45,26].) *Let E/F be a finite Galois extension with $\mu(E)$ finite. Then* $\mathrm{Cog}(E/F)$ *is a finite group. In particular, for any extension K/L of algebraic number fields, which is not necessarily Galois, the group* $\mathrm{Cog}(K/L)$ *is finite.*

PROOF. Since $\mathrm{Gal}(E/F)$ and $\mu(E)$ are finite groups, it is obvious that the group $Z^1(\mathrm{Gal}(E/F), \mu(E))$ is finite, hence $\mathrm{Cog}(E/F)$ is also finite by Theorem 5.5. Since $K/\mathbb{Q}$ and $L/\mathbb{Q}$ are both finite extensions, it follows that the extension K/L is a finite separable extension. Consider the normal closure $\widetilde{K}/L$ of the extension K/L, which is a finite Galois extension. Then, $\mathrm{Cog}(\widetilde{K}/L)$ is a finite group since $\mu(N)$ is a finite group for any algebraic number field N. Now, observe that $\mathrm{Cog}(K/L)$ is a subgroup of the finite group $\mathrm{Cog}(\widetilde{K}/L)$, hence it is also finite. □

COROLLARY 5.7. ([6].) *If E/F is a Galois extension with Galois group Γ, then the map*

$$\varphi : \{H \mid F^* \leqslant H \leqslant T(E/F)\} \to \{U \mid U \leqslant Z_c^1(\Gamma, \mu(E))\},$$
$$H \mapsto \{f_\alpha \in Z_c^1(\Gamma, \mu(E)) \mid \alpha \in H\},$$

is a lattice isomorphism, which induces a canonical lattice isomorphism

$$\mathbb{L}(\mathrm{Cog}(E/F)) \cong \mathbb{L}(Z_c^1(\Gamma, \mu(E))).$$

For every cyclic subgroup C of $Z_c^1(\Gamma, \mu(E))$ there exists $\alpha \in T(E/F)$ such that $\varphi(F^*\langle\alpha\rangle) = \langle f_\alpha\rangle = C$. Moreover, $H/F^* \cong \varphi(H)$ for every H with $F^* \leqslant H \leqslant T(E/F)$.

For an arbitrary G-radical extension E/F we shall use the following notation:

$$\mu_G(E) := \bigcup_{m \in \mathcal{O}_{G/F^*}} \mu_m(E).$$

Observe that $\mu_G(E)$ is a subgroup of the group $\mu(E)$, and $\mathrm{Im}(f_\alpha) \in \mu_G(E)$ for all $\alpha \in G$. We deduce that the group isomorphism $\psi : \mathrm{Cog}(E/F) \to Z_c^1(\mathrm{Gal}(E/F), \mu(E))$ induces by restriction to G/F^* a monomorphism

$$\psi_G : G/F^* \to Z_c^1(\mathrm{Gal}(E/F), \mu_G(E)).$$

The next result shows that if additionally the extension E/F is G-co-Galois, then the monomorphism ψ_G is also surjective, in other words, it is a group isomorphism.

THEOREM 5.8. ([6].) *For any Galois G-co-Galois extension E/F, the map $\widehat{\alpha} \mapsto f_\alpha$ yields a group isomorphism*

$$\mathrm{Kne}(E/F) \cong Z_c^1(\mathrm{Gal}(E/F), \mu_G(E)).$$

PROOF. Denote by Γ the Galois group of E/F. Since Γ is a profinite group, every $f \in Z_c^1(\Gamma, \mu_G(E))$ is locally constant, and so, taking into account that $\mathrm{lcm}(s, t) \in \mathcal{O}_{G/F^*}$ for all $s, t \in \mathcal{O}_{G/F^*}$, we deduce that $Z_c^1(\Gamma, \mu_G(E)) = \bigcup_{m \in \mathcal{O}_{G/F^*}} Z_c^1(\Gamma, \mu_m(E))$. Let $h \in Z_c^1(\Gamma, \mu_G(E))$. Then there exists $m \in \mathcal{O}_{G/F^*}$ such that $h \in Z_c^1(\Gamma, \mu_m(E))$. But, the map $\psi_m : \mathrm{Cog}_m(E/F) \to Z_c^1(\Gamma, \mu_m(E))$ is an isomorphism, hence $h = \psi_m(\widehat{\alpha})$ for some $\alpha \in T_m(E/F)$, where $T_m(E/F) = \{x \in E^* \mid x^m \in F^*\}$ and $\mathrm{Cog}_m(E/F) = T_m(E/F)/F^*$. Since clearly $\mathcal{P}_m \subseteq \mathcal{P}_G$, Lemma 4.10 implies that $\alpha \in G$. Thus $\widehat{\alpha} \in G/F^* = \mathrm{Kne}(E/F)$, and so $h = \psi_m(\widehat{\alpha}) = \psi_G(\widehat{\alpha})$, which proves that ψ_G is a surjective map, as desired. □

COROLLARY 5.9. ([19,6].) *Let E/F be a Galois n-bounded G-co-Galois extension, in particular a finite Galois G-co-Galois extension with $n = \exp(G/F^*)$. Then*

$$\mathrm{Kne}(E/F) \cong Z_c^1(\mathrm{Gal}(E/F), \mu_n(E)).$$

For any Galois G-co-Galois extension E/F, the map

$$f : \mathrm{Gal}(E/F) \times \mathrm{Cog}(E/F) \to \mu(E),$$

considered at the beginning of this subsection yields by restriction the map

$$g : \mathrm{Gal}(E/F) \times \mathrm{Kne}(E/F) \to \mu_G(E), \quad (\sigma, \widehat{\alpha}) \mapsto \sigma(\alpha) \cdot \alpha^{-1}.$$

For every $\varDelta \leqslant \mathrm{Gal}(E/F)$ and $W \leqslant \mathrm{Kne}(E/F)$ let

$$\begin{aligned} \varDelta^{\top} &= \bigl\{c \in \mathrm{Kne}(E/F) \mid g(\sigma, c) = 1, \forall \sigma \in \varDelta\bigr\}, \\ W^{\top} &= \bigl\{\sigma \in \mathrm{Gal}(E/F) \mid g(\sigma, c) = 1, \forall c \in W\bigr\}. \end{aligned}$$

Proposition 5.10. ([6].) *For any Galois G-co-Galois extension E/F, the assignments* (-)$^{\top}$ *define mutually inverse anti-isomorphisms between the lattices* $\overline{\mathbb{L}}(\mathrm{Gal}(E/F))$ *and* $\mathbb{L}(\mathrm{Kne}(E/F))$.

Let E/F be a Galois extension with Galois group $\varGamma$. Then, by Theorem 5.5, there exists a canonical isomorphism $\mathrm{Cog}(E/F) \cong Z_c^1(\varGamma, \mu(E))$, hence the canonical map $f : \mathrm{Gal}(E/F) \times \mathrm{Cog}(E/F) \to \mu(E)$, considered above produces, by replacing $\mathrm{Cog}(E/F)$ with its isomorphic copy $Z_c^1(\varGamma, \mu(E))$, precisely the *evaluation map*

$$\langle \text{-}, \text{-} \rangle : \varGamma \times Z_c^1(\varGamma, \mu(E)) \to \mu(E), \quad \langle \sigma, h \rangle = h(\sigma).$$

For any $\varDelta \leqslant \varGamma$, $U \leqslant Z_c^1(\varGamma, \mu(E))$, and $\chi \in Z_c^1(\varGamma, \mu(E))$ we shall write

$$\begin{aligned} \varDelta^{\perp} &= \bigl\{h \in Z_c^1\bigl(\varGamma, \mu(E)\bigr) \mid \langle \sigma, h \rangle = 1, \forall \sigma \in \varDelta\bigr\}, \\ U^{\perp} &= \bigl\{\sigma \in \varGamma \mid \langle \sigma, h \rangle = 1, \forall h \in U\bigr\}, \\ \chi^{\perp} &= \bigl\{\sigma \in \varGamma \mid \langle \sigma, \chi \rangle = 1\bigr\}. \end{aligned}$$

One verifies easily that $\varDelta^{\perp} \leqslant Z_c^1(\varGamma, \mu(E))$, $U^{\perp} \leqslant \varGamma$, and $\chi^{\perp} = \langle \chi \rangle^{\perp}$.

The next two results characterize the radical, G-Kneser, and G-co-Galois subextensions of a given Galois extension E/F with Galois group $\varGamma$ via subgroups of $Z_c^1(\varGamma, \mu(E))$.

Proposition 5.11. ([6].) *Let E/F be a Galois extension with Galois group $\varGamma$, and let $K \in \mathbb{I}(E/F)$. Then K/F is a radical extension* (*respectively a simple radical extension*) *if and only if there exists a $U \leqslant Z_c^1(\varGamma, \mu(E))$* (*respectively a $\chi \in Z_c^1(\varGamma, \mu(E))$*) *such that* $\mathrm{Gal}(E/K) = U^{\perp}$ (*respectively* $\mathrm{Gal}(E/K) = \chi^{\perp}$).

Corollary 5.12. ([6].) *Let E/F be a Galois extension with Galois group $\varGamma$, and let K/F be a G-radical subextension of E/F. For every H with $F^* \leqslant H \leqslant G$ set $U_H = \{f_\alpha \mid \alpha \in H\} \leqslant Z_c^1(\varGamma, \mu(E))$. Then, the following assertions hold.*

1. *The extension K/F is G-Kneser if and only if $(\varGamma : U_H^{\perp}) = |U_H|$ for every H with $F^* \leqslant H \leqslant G$ and H/F^* finite.*
2. *The extension K/F is G-co-Galois if and only if it is G-Kneser and the "perpendicular" map $V \mapsto V^{\perp}$ yields a bijection, or equivalently, an anti-isomorphism of lattices*

$$\{V \mid V \leqslant U_G\} \to \bigl\{\varDelta \mid U_G^{\perp} \leqslant \varDelta \leqslant \varGamma, \varDelta \text{ closed subgroup of } \varGamma\bigr\}.$$

Now, we analyze the transfer under change of base fields of the property of a Galois extension being radical. For any extension E/F we denote by $\mathcal{R}(E/F)$ the set of all subextensions K/F of E/F which are radical. If E/F is an arbitrary Galois extension and L/F is any extension with $L \cap E = F$, and such that E and L are subfields of some other field, then it is well known that the canonical map $\mathrm{Gal}(EL/L) \overset{\sim}{\to} \mathrm{Gal}(E/F), \sigma \mapsto \sigma_{|E}$, is an isomorphism of topological groups. Using this map, it follows that there exists an injective map

$$\begin{aligned}&\rho : \mathcal{R}(E/F) \to \mathcal{R}(EL/L),\\ &\quad F(G)/F \mapsto F(G)L/L = L(GL^*)/L, \quad F^* \leqslant G \leqslant T(E/F).\end{aligned}$$

Proposition 5.13. ([6].) *Let E/F be a Galois extension with Galois group Γ, and let L/F be an arbitrary field extension such that $E \cap L = F$. If E and L are subfields of some other field, and $\mu(EL) = \mu(E)$, then the map*

$$\rho : \mathcal{R}(E/F) \to \mathcal{R}(EL/L), \quad F(G)/F \mapsto L(GL^*)/L, \quad F^* \leqslant G \leqslant T(E/F),$$

is bijective, with inverse

$$\begin{aligned}&\mathcal{R}(EL/L) \to \mathcal{R}(E/F),\\ &L(G_1)/L \mapsto F(G_1 \cap E^*)/F, \quad L^* \leqslant G_1 \leqslant T(EL/L).\end{aligned}$$

5.3. *Abelian G-co-Galois extensions*

The aim of this subsection is to show that for any Abelian G-co-Galois extension E/F, the Kneser group G/F^* of E/F is isomorphic to the group $\mathrm{Ch}(\Gamma)$ of characters of the profinite group $\Gamma = \mathrm{Gal}(E/F)$. To do that, observe that the lattice anti-isomorphism from the lattice $\mathbb{I}(E/F)$ of all intermediate fields of the extension E/F onto the lattice $\overline{\mathbb{L}}(\Gamma)$ of all closed subgroups of Γ given by the fundamental theorem of infinite Galois theory, produces, by taking the characters, a lattice isomorphism from the lattice $\mathbb{I}(E/F)$ onto the lattice $\mathbb{L}(\mathrm{Ch}(\Gamma))$ of all subgroups of the character group $\mathrm{Ch}(\Gamma)$. On the other hand, since E/F is a G-co-Galois extension, the map $H/F^* \mapsto F(H)$ is an isomorphism from the lattice $\mathbb{L}(\mathrm{Kne}(E/F))$ onto the lattice $\mathbb{I}(E/F)$. If we compose these two lattice isomorphisms, we obtain a lattice isomorphism from $\mathbb{L}(\mathrm{Kne}(E/F))$ onto $\mathbb{L}(\mathrm{Ch}(\Gamma))$. It is natural to ask whether or not such a lattice isomorphism yields a group isomorphism between $\mathrm{Kne}(E/F)$ and $\mathrm{Ch}(\mathrm{Gal}(E/F))$.

Thus, the following natural question arises: given two groups A and B, when does a lattice isomorphism $\varphi : \mathbb{L}(A) \to \mathbb{L}(B)$ produce a group isomorphism $f : A \to B$? The groups A and B are called *lattice-isomorphic* if there exists a lattice isomorphism between their subgroup lattices $\mathbb{L}(A)$ and $\mathbb{L}(B)$. With this terminology, the problem we just stated can be briefly reformulated as follows: *when are two lattice-isomorphic groups isomorphic*? In general, lattice-isomorphic groups are not isomorphic, as a classical example discovered in 1928 by A. Rottlaender shows (see Baer [23]). However, if some restrictive conditions on one or both groups A and B are imposed, then any lattice-isomorphism between A and B produces a group isomorphism between them. By chance, such conditions are satisfied in our case, and so, the canonical lattice isomorphism between $\mathbb{L}(\mathrm{Kne}(E/F))$ and $\mathbb{L}(\mathrm{Ch}(\mathrm{Gal}(E/F)))$ yields a (non-canonical) group isomorphism between the groups $\mathrm{Kne}(E/F)$ and $\mathrm{Ch}(\mathrm{Gal}(E/F))$. In case E/F is a finite Abelian G-co-Galois, this implies that the two Abelian groups $\mathrm{Kne}(E/F)$ and $\mathrm{Gal}(E/F)$ are isomorphic, but not in a canonical way.

We shall denote by **Gr** the category of all groups, by **Pos** the category of all posets, and by **Lat** the category of all lattices. The morphisms in the category **Pos** are the order-preserving (i.e., increasing) maps. Recall that if L and L' are lattices, then a morphism in **Lat** from L to L' is a map $\alpha : L \to L'$ commuting with the joins and meets. Clearly, any group morphism $f : A \to B$ yields a morphism $\mathbb{L}(f) : \mathbb{L}(A) \to \mathbb{L}(B)$, $H \mapsto f(H)$, in **Pos**, which is not necessarily a morphism in **Lat**. A map $\alpha : L \to L'$ is an isomorphism in **Lat** if and only if α is an order-preserving bijection such that its inverse α^{-1} is also an order-preserving map, in other words, α is an isomorphism in the category **Pos**. This implies that for any isomorphism $f : A \to B$ in **Gr**, the map $\mathbb{L}(f) : \mathbb{L}(A) \to \mathbb{L}(B)$, $H \mapsto f(H)$, is an isomorphism in **Lat**. This fact can be described briefly by saying that we have a canonical map

$$\mathbb{L}_{A,B} : \mathrm{Isom}_{\mathbf{Gr}}(A, B) \to \mathrm{Isom}_{\mathbf{Lat}}\big(\mathbb{L}(A), \mathbb{L}(B)\big), \quad f \mapsto \mathbb{L}(f),$$

where, if $\mathcal{C}$ is any category, then $\mathrm{Isom}_{\mathcal{C}}(X, Y)$ denotes the set, possibly empty, of all isomorphisms from the object X of $\mathcal{C}$ to the object Y of $\mathcal{C}$.

DEFINITION. A *lattice-isomorphism* from a group A to a group B is any isomorphism of lattices $\varphi \in \mathrm{Isom}_{\mathbf{Lat}}(\mathbb{L}(A), \mathbb{L}(B))$. The lattice-isomorphism φ is said to be *induced by a group isomorphism* if there exists a group isomorphism $f : A \to B$ such that $\varphi = \mathbb{L}(f)$, and in that case, φ is said to be *induced by* f. The groups A and B are called *lattice-isomorphic* if $\mathrm{Isom}_{\mathbf{Lat}}(\mathbb{L}(A), \mathbb{L}(B)) \neq \varnothing$, and we denote this situation by $A \cong_{\mathbb{L}} B$.

The term of *lattice-isomorphism* of groups also has various other names in the literature: *subgroup-isomorphism* in Baer [23], *projectivity* in Schmidt [71] and Suzuki [76], *L-isomorphism*, etc. If A and B are isomorphic groups, then, as usually, we shall denote this situation by $A \cong B$. Clearly, if $A, B \in \mathrm{Gr}$ and $A \cong B$, then $A \cong_{\mathbb{L}} B$, but not conversely. Therefore, the following natural question arises: *Given a class $\mathcal{X}$ of groups, what kind of conditions* (C) *on lattice-isomorphisms of groups should be imposed such that for every $A \in \mathcal{X}$ and for every $B \in$* **Gr**, *every lattice-isomorphism $A \cong_{\mathbb{L}} B$ satisfying the conditions* (C) *implies that $A \cong B$?*

For $\mathcal{X}$ = the class of all Abelian torsion groups, two conditions are sufficient to completely answer the question above, namely, "index-preserving" and "normal": we say that a lattice-isomorphism $\varphi \in \mathrm{Isom}_{\mathbf{Lat}}(\mathbb{L}(A), \mathbb{L}(B))$ between the groups A and B is *index-preserving* if $(C : D) = (\varphi(C) : \varphi(D))$ for every cyclic subgroup C of A and for every $D \leqslant C$. The lattice-isomorphism φ is said to be *normal* if $\varphi(N) \lhd B$ for every $N \lhd A$.

THEOREM 5.14. ([23,14].) *Let A be a torsion Abelian group, and let B be a group which is lattice-isomorphic to A via $\varphi : \mathbb{L}(A) \xrightarrow{\sim} \mathbb{L}(B)$. If φ is index-preserving and normal, then $A \cong B$.*

Recall that for any topological group T we denote by $\overline{\mathbb{L}}(T)$ the lattice of all closed subgroups of T, and by $\mathrm{Ch}(T)$ or by $\widehat{T}$ the *character group* of T, that is, the group $\mathrm{Hom}_c(T, \mathbb{U})$ of all continuous group morphisms of T into the unit circle $\mathbb{U}$. Note that if T is a profinite group, then $\widehat{T}$ can be identified with the torsion Abelian group $\mathrm{Hom}_c(T, \mathbb{Q}/\mathbb{Z})$. For any locally compact Abelian group A and subsets $\varnothing \neq X \subseteq A$ and $\varnothing \neq Y \subseteq \widehat{A}$ we shall use the following notation:

$$X^{\perp} = \{\chi \in \widehat{A} \mid \chi(a) = 1, \forall a \in X\}, \qquad Y^{\perp} = \{a \in A \mid \chi(a) = 1, \forall \chi \in Y\}.$$

From the *Pontryagin duality* one deduces easily that the map $\overline{\mathbb{L}}(A) \to \overline{\mathbb{L}}(\widehat{A})$, $X \mapsto \widehat{X}$, is an anti-isomorphism of lattices, with inverse $\overline{\mathbb{L}}(\widehat{A}) \to \overline{\mathbb{L}}(A)$, $Y \mapsto \widehat{Y}$. Note that the Abelian group A is compact if and only if $\widehat{A}$ is a discrete group, and then $\overline{\mathbb{L}}(\widehat{A}) = \mathbb{L}(\widehat{A})$.

PROPOSITION 5.15. ([14].) *For any Abelian G-co-Galois extension E/F, the discrete torsion Abelian groups* $\mathrm{Kne}(E/F)$ *and* $\widehat{\mathrm{Gal}(E/F)}$ *are lattice-isomorphic via the canonical lattice isomorphism*

$$\mathbb{L}(\mathrm{Kne}(E/F)) \xrightarrow{\sim} \mathbb{L}(\widehat{\mathrm{Gal}(E/F)}), \quad H/F^* \mapsto \mathrm{Gal}(E/F(H))^{\perp}.$$

PROOF. Consider the following canonical maps:

$$\begin{aligned}
&\gamma_1 : \mathbb{L}(\mathrm{Kne}(E/F)) \to \mathbb{I}(E/F), \quad H/F^* \mapsto F(H)/F,\\
&\gamma_2 : \mathbb{I}(E/F) \to \overline{\mathbb{L}}(\mathrm{Gal}(E/F)), \quad K/F \mapsto \mathrm{Gal}(E/K),\\
&\gamma_3 : \overline{\mathbb{L}}(\mathrm{Gal}(E/F)) \to \mathbb{L}(\widehat{\mathrm{Gal}(E/F)}), \quad B \mapsto B^{\perp}.
\end{aligned}$$

Since γ_1 is an isomorphism of lattices by Theorem 4.2, γ_2 is an anti-isomorphism of lattices by the fundamental theorem of infinite Galois theory, and γ_3 is an anti-isomorphism of lattices as observed above, we deduce that their composition $\gamma = \gamma_3 \circ \gamma_2 \circ \gamma_1$ yields an isomorphism of lattices

$$\gamma : \mathbb{L}(\mathrm{Kne}(E/F)) \to \mathbb{L}(\widehat{\mathrm{Gal}(E/F)}), \quad H/F^* \mapsto \mathrm{Gal}(E/F(H))^{\perp}.$$ □

THEOREM 5.16. ([7,14].) *If E/F is an Abelian G-co-Galois extension, then the discrete torsion Abelian groups* $\mathrm{Kne}(E/F)$ *and* $\widehat{\mathrm{Gal}(E/F)}$ *are isomorphic.*

PROOF. By Proposition 5.15, the Abelian groups $\mathrm{Kne}(E/F)$ and $\widehat{\mathrm{Gal}(E/F)}$ are lattice-isomorphic via the isomorphism of lattices

$$\gamma : \mathbb{L}(\mathrm{Kne}(E/F)) \to \mathbb{L}(\widehat{\mathrm{Gal}(E/F)}), \quad H/F^* \mapsto \mathrm{Gal}(E/F(H))^{\perp},$$

for every $H/F^* \leqslant \mathrm{Kne}(E/F) = G/F^*$, where $F^* \leqslant H \leqslant G$. Since $\mathrm{Kne}(E/F)$ is clearly a torsion Abelian group, the lattice-isomorphism γ is index-preserving as can be easily checked, and γ is obviously normal because $\widehat{\mathrm{Gal}(E/F)}$ is an Abelian group, we can apply Theorem 5.14 to obtain the desired result. □

By Pontryagin duality, any locally compact Abelian group A is topologically isomorphic to its second character group $\mathrm{Ch}(\mathrm{Ch}(A)) = \widehat{\widehat{A}}$, so Theorem 5.16 immediately implies the next result.

COROLLARY 5.17. ([14].) *For any Abelian G-co-Galois extension E/F, the totally disconnected compact Abelian groups* $\mathrm{Gal}(E/F)$ *and* $\widehat{\mathrm{Kne}(E/F)}$ *are topologically isomorphic.*

COROLLARY 5.18. ([14].) *If E/F is an Abelian co-Galois extension, then the discrete torsion Abelian groups* $\mathrm{Cog}(E/F)$ *and* $\widehat{\mathrm{Gal}(E/F)}$ *are isomorphic.*

PROOF. Any co-Galois extension E/F is $T(E/F)$-co-Galois by Corollary 4.4. But $\mathrm{Kne}(E/F) = T(E/F)/F^* = \mathrm{Cog}(E/F)$, so the result follows at once from Theorem 5.16. □

Since any finite Abelian group A is isomorphic with its character group $\widehat{A}$, there is the following particular case of Theorem 5.16.

COROLLARY 5.19. ([16,26].) *For any finite Abelian G-co-Galois extension E/F, the finite Abelian groups* $\mathrm{Kne}(E/F)$ *and* $\mathrm{Gal}(E/F)$ *are isomorphic. In particular, the groups* $\mathrm{Cog}(E/F)$ *and* $\mathrm{Gal}(E/F)$ *are isomorphic for any finite Abelian co-Galois extension E/F.*

A proof of Corollary 5.19 by induction on $[E : F]$ is given in Albu and Nicolae [16], and an alternate proof of Theorem 5.16, that avoids the lattice-isomorphism technique and is based on reducing the problem to the case when the Kneser group $\mathrm{Kne}(E/F)$ is a p-group, is provided in Albu [11, Remark 15.3.12].

5.4. *Kummer theory via co-Galois theory*

The concept of a *Kummer extension* is a classical one in field theory. To distinguish such extensions from various generalizations, we refer to them as *classical Kummer extensions*. A good account of the (classical) *Kummer theory* is provided by Artin [22], Bourbaki [31], Karpilovsky [54], and Lang [57]. The prototype of a G-co-Galois extension is, by Kummer theory, any classical Kummer extension. In this subsection we show that all of Kummer theory can be immediately deduced from our co-Galois theory using the n-purity criterion (Corollary 4.5). Moreover, this criterion allows us to give large classes of G-co-Galois extensions which generalize or are closely related to classical Kummer extensions: *generalized Kummer extensions*, *Kummer extensions with few roots of unity*, and *quasi-Kummer extensions*. The prototype of an (infinite) Kummer extension with few roots of unity is any subextension of $\mathbb{R}/\mathbb{Q}$ of the form $\mathbb{Q}(\{\sqrt[n]{a_i} \mid i \in I\})/\mathbb{Q}$, where $\{a_i \mid i \in I\}$ is an arbitrary nonempty set of strictly positive rational numbers. Note that, in general, these extensions are not Galois if $n \geqslant 3$. Placing the classical Kummer extensions in the framework of G-co-Galois extensions allows us to derive, from the co-Galois theory, results on generalized Kummer extensions and quasi-Kummer extensions. In particular, we obtain results on Kummer extensions with few roots of unity, which are very similar to the known ones for classical Kummer extensions.

5.4.1. *Classical Kummer extensions* First, we present various characterizations of classical Kummer extensions. Then, we show that the classical Kummer extensions are G-co-Galois. As an immediate consequence of this fact we deduce the whole of classical *Kummer theory*.

Recall that Ω is a fixed algebraically closed field containing the fixed base field F as a subfield; any overfield of F considered is supposed to be a subfield of Ω. For any nonempty subset A of F^* and any $n \in \mathbb{N}^*$ we will denote by $\sqrt[n]{A}$ the subset of $T(\Omega/F)$ defined by $\sqrt[n]{A} := \{x \in \Omega \mid x^n \in A\}$. In particular, if A is a singleton $\{a\}$, then $\sqrt[n]{\{a\}}$ is precisely the set of all roots (in Ω) of the polynomial $X^n - a \in F[X]$. We shall use throughout this chapter the notation $\sqrt[n]{a}$ to designate a root, which in general is not specified, of this polynomial. Thus, $\sqrt[n]{a} \in \sqrt[n]{\{a\}}$. More precisely, for any choice of the root $\sqrt[n]{a}$, we have $\sqrt[n]{\{a\}} = \{\zeta_n^k \sqrt[n]{a} \mid 0 \leqslant k \leqslant n-1\}$. In particular, if $\zeta_n \in F$, then $F(\sqrt[n]{\{a\}}) = F(\sqrt[n]{a})$. However, in certain cases, for instance when F is a subfield of the field $\mathbb{R}$ of all real numbers and $a > 0$, then $\sqrt[n]{a}$ will always mean the unique positive root in $\mathbb{R}$ of the polynomial $X^n - a$.

Definition. A *classical n-Kummer extension*, where $n \in \mathbb{N}^*$, is an Abelian extension E/F such that $\gcd(n, e(F)) = 1$, $\mu_n(\Omega) \subseteq F$ and $\mathrm{Gal}(E/F)$ is a group of exponent a divisor of n. A *classical Kummer extension* is an extension which is a classical n-Kummer extension for some $n \in \mathbb{N}^*$. If E/F is a classical Kummer extension, we also say that E is a classical Kummer extension of F.

Theorem 5.20. *The following assertions are equivalent for an extension E/F and a number $n \in \mathbb{N}^*$.*

(1) *E/F is a classical n-Kummer extension.*
(2) $\gcd(n, e(F)) = 1$, $\mu_n(\Omega) \subseteq F$, *and* $E = F(\sqrt[n]{A})$ *for some* $\varnothing \neq A \subseteq F^*$.
(3) $\gcd(n, e(F)) = 1$, $\mu_n(\Omega) \subseteq F$, *and* $E = F(B)$ *for some* $\varnothing \neq B \subseteq E^*$ *with* $B^n \subseteq F$.

Note that a classical n-Kummer extension E/F is finite if and only if any of the sets A and B in Theorem 5.20 can be chosen to be finite.

We are now going to show how all of Kummer theory can be very easily derived from co-Galois theory.

Lemma 5.21. ([21].) *Let E/F be a bounded separable G-radical extension, and let $n \in \mathbb{N}^*$ be such that $G^n \subseteq F^*$. If the extension E/F is n-pure, then E/F is G-co-Galois.*

Proof. If $m = \exp(G/F^*)$, then clearly $m \mid n$, hence for any $p \in \mathcal{P}_m$, we have $p \mid n$. By hypothesis, $\mu_p(E) \subseteq F$, and so, the extension E/F is also m-pure, hence it is G-co-Galois by Corollary 4.5. □

Lemma 5.22. ([11].) *The following statements hold for any extension E/F such that $\mu_n(E) \subseteq F$ for a given $n \in \mathbb{N}^*$, and any group G such that $F^* \leqslant G \leqslant E^*$.*

(1) *The map* $G/F^* \to G^n/F^{*n}$, $xF^* \mapsto x^nF^{*n}$, *is a group isomorphism.*

(2) *The maps* $H \mapsto H^n$ *and* $M \mapsto \sqrt[n]{M} \cap G$ *establish lattice isomorphisms, inverse to one another, between the lattices* $\{H \mid F^* \leqslant H \leqslant G\}$ *and* $\{M \mid F^{*n} \leqslant M \leqslant G^n\}$.

Excepting point (1) which is due to Albu and Ţena [21], the next result is the core of the classical Kummer theory. Our approach, based on (1), is completely different from the standard one, as presented, e.g., in the books of Artin [22], Bourbaki [31], Karpilovsky [54], and/or Lang [57] that offer a good account of (classical) Kummer theory.

THEOREM 5.23. *Let* E/F *be a classical n-Kummer extension, with* $\varnothing \neq A \subseteq F^*$, $E = F(\sqrt[n]{A})$, $\gcd(n, e(F)) = 1$, *and* $\mu_n(\Omega) \subseteq F$. *Then, the following assertions hold.*

(1) E/F *is an* $F^*\langle\sqrt[n]{A}\rangle$*-co-Galois extension.*

(2) *The maps* $H \mapsto F(\sqrt[n]{H})$ *and* $K \mapsto K^n \cap (F^{*n}\langle A\rangle)$ *establish isomorphisms of lattices, inverse to one another, between the lattice of all subgroups* H *of* $F^{*n}\langle A\rangle$ *containing* F^{*n}, *and the lattice* $\mathbb{I}(E/F)$ *of all intermediate fields* K *of* E/F. *Moreover, any subextension* K/F *of* E/F *is a classical n-Kummer extension.*

(3) *If* H *is any subgroup of* $F^{*n}\langle A\rangle$ *containing* F^{*n}, *then any set of representatives of the group* $\sqrt[n]{H}/F^*$ *is a vector space basis of* $F(\sqrt[n]{H})$ *over* F, *and* $[F(\sqrt[n]{H}) : F] = |H/F^{*n}|$.

(4) *There exists a canonical group isomorphism*

$$F^*\big\langle\sqrt[n]{A}\big\rangle/F^* \cong \mathrm{Hom}_c\big(\mathrm{Gal}(E/F), \mu_n(F)\big).$$

PROOF. Set $B = \sqrt[n]{A}$ and $G = F^*\langle B\rangle$. Since $\mu_n(\Omega) \subseteq F$ it follows that the extension E/F is n-pure. Clearly $G^n \subseteq F$, hence the extension E/F is G-co-Galois by Lemma 5.21. By (1) and Theorem 4.2, the maps $H \mapsto F(H)$ and $K \mapsto K \cap (F^*\langle\sqrt[n]{A}\rangle)$ are lattice isomorphisms, inverse to one another, between the lattice of all subgroups H of $F^*\langle\sqrt[n]{A}\rangle$ containing F^* and the lattice $\mathbb{I}(E/F)$ of all intermediate fields K of E/F. Now, apply Lemma 5.22 for $G = F^*\langle\sqrt[n]{A}\rangle$ and observe that $G^n = F^{*n}\langle A\rangle$. The extension E/F is $F^*\langle\sqrt[n]{A}\rangle$-co-Galois by (1), hence the extension $F(\sqrt[n]{H})/F$ is $\sqrt[n]{H}$-Kneser. Use again Lemma 5.22 to deduce (3).

To prove (4), observe that $\mathrm{Kne}(E/F) = G/F^* = F^*\langle\sqrt[n]{A}\rangle/F^*$ by (1). On the other hand, by Corollary 5.9, there exists a canonical group isomorphism $\mathrm{Kne}(E/F) \cong Z^1_c(\mathrm{Gal}(E/F), \mu_m(E))$, where $m = \exp(G/F^*)$. Clearly $m \mid n$, so $\mu_m(E) \subseteq \mu_n(E) \subseteq F$. It follows that every 1-cocycle f of $\mathrm{Gal}(E/F)$ with coefficients in $\mu_m(E)$ is actually a morphism of groups, since $f(\sigma\tau) = f(\sigma) \cdot \sigma(f(\tau)) = f(\sigma) \cdot f(\tau)$, for every $\sigma, \tau \in \mathrm{Gal}(E/F)$. Thus,

$$\begin{aligned}\mathrm{Kne}(E/F) &\cong Z^1_c\big(\mathrm{Gal}(E/F), \mu_m(E)\big) = \mathrm{Hom}_c\big(\mathrm{Gal}(E/F), \mu_m(E)\big)\\ &= \mathrm{Hom}_c\big(\mathrm{Gal}(E/F), \mu_n(F)\big).\end{aligned}$$ □

Let F be any field, and let $n \in \mathbb{N}^*$ be such that $\gcd(n, e(F)) = 1$ and $\mu_n(\Omega) \subseteq F$. In view of Theorem 5.20, any classical n-Kummer extension E of F, with E a subfield of Ω, has the form $F(\sqrt[n]{A})$ for some $\varnothing \neq A \subseteq F^*$. Since obviously $F(\sqrt[n]{F^*})/F$ is a classical n-Kummer extension, it follows that the set of all classical n-Kummer extensions of F contained in Ω, ordered by inclusion, has a greatest element, namely $F(\sqrt[n]{F^*})$, called the *maximal classical n-Kummer extension of* F contained in Ω; clearly, it is the greatest

Abelian extension of F contained in Ω for which its Galois group is a group of exponent a divisor of n. Taking $A = F^*$ in Theorem 5.23, we obtain the next result, which is Kummer theory for the maximal classical n-Kummer extension of F.

COROLLARY 5.24. *Let F be any field and let $n \in \mathbb{N}^*$ be such that* $\gcd(n, e(F)) = 1$ *and* $\mu_n(\Omega) \subseteq F$. *Then, the maps $H \mapsto F(\sqrt[n]{H})$ and $E \mapsto E^n \cap F^*$ establish isomorphisms of lattices, inverse to one another, between the lattice of all subgroups H of F^* containing F^{*n} and the lattice of all classical n-Kummer extensions E of F (contained in Ω).*

5.4.2. *Generalized Kummer extensions* We present another class of G-co-Galois extensions, which is larger than the class of classical Kummer extensions, namely the class of *generalized Kummer extensions*, introduced and investigated by Albu and Nicolae [16] for finite extensions, and by Albu and Ţena [21] for infinite extensions. This new class also includes the class of *Kummer extension with few roots of unity* which will be discussed in 5.4.3. We show that a theory of generalized Kummer extensions, which is very similar to that of classical Kummer extensions, can be developed using the properties of G-co-Galois extensions; since, in general, they are not Galois extensions, no other approach (e.g., via Galois theory, as in the case of classical Kummer extensions) seems to be applicable.

DEFINITION. We say that an extension E/F is a *generalized n-Kummer extension*, where $n \in \mathbb{N}^*$, if $E = F(B)$ for some $\varnothing \neq B \subseteq E^*$, with $\gcd(n, e(F)) = 1$, $B^n \subseteq F$, and $\mu_n(E) \subseteq F$. A *generalized Kummer extension* is an extension which is a generalized n-Kummer extension for some $n \in \mathbb{N}^*$.

Observe that any classical n-Kummer extension $F(\sqrt[n]{A})/F$ is a generalized n-Kummer extension, since if we write $B = \sqrt[n]{A}$, then $B^n = A \subseteq F^*$. If we proceed as in the proof of Theorem 5.23, the following result is obtained.

THEOREM 5.25. ([21,11].) *Let E/F be a generalized n-Kummer extension, with $\varnothing \neq B \subseteq E^*$, $E = F(B)$, $B^n \subseteq F$,* $\gcd(n, e(F)) = 1$, *and $\mu_n(E) \subseteq F$. If we write $G = F^*\langle B\rangle$, then the following statements hold.*

(1) *The extension E/F is G-co-Galois.*

(2) *The maps $H \mapsto F(\sqrt[n]{H} \cap G)$ and $K \mapsto K^n \cap (F^{*n}\langle B^n\rangle)$ establish isomorphisms of lattices, inverse to one another, between the lattice of all subgroups H of $F^{*n}\langle B^n\rangle$ containing F^{*n} and the lattice $\mathbb{I}(E/F)$ of all intermediate fields K of E/F. Moreover, any subextension K/F of E/F is a generalized n-Kummer extension.*

(3) *If H is any subgroup of $F^{*n}\langle B^n\rangle$ containing F^{*n}, then any set of representatives of the group $(\sqrt[n]{H} \cap G)/F^*$ is a vector space basis of $F(\sqrt[n]{H} \cap G)$ over F, and $[F(\sqrt[n]{H} \cap G) : F] = |H/F^{*n}|$.*

5.4.3. *Kummer extensions with few roots of unity* The concept of *Kummer extension with few roots of unity* was introduced and investigated by Albu [3] for finite extensions, and by Albu and Ţena [21] for infinite extensions.

DEFINITION. We say that an extension E/F is an *n-Kummer extensions with few roots of unity* if $E = F(B)$ for some $\varnothing \neq B \subseteq E^*$, with $\gcd(n, e(F)) = 1$, $B^n \subseteq F$ and

$\mu_n(E) \subseteq \{-1, 1\}$. A *Kummer extensions with few roots of unity* is an extension which is an n-Kummer extensions with few roots of unity for some $n \in \mathbb{N}^*$.

Kummer extensions with few roots of unity are very particular cases of generalized Kummer extensions, so Theorem 5.25 is applicable for them.

5.4.4. *Quasi-Kummer extensions* We present now still another class of G-co-Galois extensions which is close to the class of classical Kummer extensions. The notion of a (finite) *neat presentation* has been introduced by Greither and Harrison [45] as follows: a neat presentation is a Galois extension E/F with $E = F(x_1, \ldots, x_r)$, $\mathrm{Char}(F) = 0$, and $x_1^n, \ldots, x_r^n \in F$ for some $n \in \mathbb{N}^*$, such that $\mu_p(\Omega) \subseteq F$ for every $p \in \mathcal{P}_n$. Using heavy cohomological machinery, which includes the Lyndon–Hochschild spectral sequence, Greither and Harrison proved that any such extension is an extension with $F^*\langle x_1, \ldots, x_r\rangle/F^*$-co-Galois correspondence, i.e., in our terminology, is an $F^*\langle x_1, \ldots, x_r\rangle$-co-Galois extension. Albu and Nicolae [16] introduced the more general concept of (finite) *generalized neat presentation* by dropping from Greither and Harrison's definition the condition "E/F is Galois" and by weakening the condition "$\mathrm{Char}(F) = 0$" to "$\gcd(e(F), n) = 1$", and proved in a very simple manner that any such extension is still $F^*\langle x_1, \ldots, x_r\rangle$-co-Galois. Infinite generalized neat presentations were introduced and investigated by Albu and Ţena [21], and then renamed *quasi-Kummer extensions* in Albu [11].

DEFINITION. An extension E/F is said to be a *quasi-Kummer extension* if $E = F(B)$ for some $\varnothing \neq B \subseteq E^*$, and there exists an $n \in \mathbb{N}^*$ with $B^n \subseteq F$, $\gcd(n, e(F)) = 1$, such that $\zeta_p \in F$ for every $p \in \mathcal{P}_n$.

THEOREM 5.26. ([21].) *Any quasi-Kummer extension $F(B)/F$ with B as in definition above is an $F^*\langle B\rangle$-co-Galois extension.*

PROOF. Set $G = F^*\langle B\rangle$. Then $G^n \subseteq F^*$ and $E = F(G)$, so, E/F is a bounded G-radical extension. If $m = \exp(G/F^*)$, then $m|n$, hence $\gcd(m, e(F)) = 1$, and consequently, E/F is a separable extension by Lemma 5.1. For every $p \in \mathcal{P}_n$ we have $\mu_p(E) \subseteq \mu_p(\Omega) \subseteq F$, so E/F is n-pure. By Lemma 5.21, it follows that $F(B)/F$ is an $F^*\langle B\rangle$-co-Galois extension. □

PROPOSITION 5.27. ([21].) *Any Galois n-bounded G-co-Galois extension is a quasi-Kummer extension.*

PROOF. Let E/F be a Galois n-bounded G-co-Galois extension. Then $E = F(G)$, with $F^* \leqslant G \leqslant E^*$ and $\exp(G/F^*) = n$. Since the extension E/F is Galois, it follows that $\gcd(n, e(F)) = 1$ and $\zeta_n \in E$ by Corollary 5.3, so $\mu_n(\Omega) \subseteq \mu_n(E)$. By Corollary 4.5, the G-co-Galois extension E/F is n-pure, hence $\mu_p(E) \subseteq F$ for every $p \in \mathcal{P}_n$. Then $\mu_p(\Omega) \subseteq \mu_p(E) \subseteq F$ for any such p. Consequently, the extension E/F is quasi-Kummer. □

COROLLARY 5.28. ([21].) *Let E be a subfield of $\mathbb{C}$. Then $E/\mathbb{Q}$ is a Galois n-bounded G-co-Galois extension if and only if $E/\mathbb{Q}$ is a classical 2-Kummer extension.*

PROPOSITION 5.29. ([21].) *Let E/F be a Galois generalized m-Kummer extension, with $E = F(B)$, $\varnothing \neq B \subseteq E^*$, $\gcd(m, e(F)) = 1$, $B^m \subseteq F$, and $\mu_m(E) \subseteq F$. Then E/F is $F^*\langle B\rangle$-co-Galois and $\mu_n(\Omega) \subseteq F$, where $n = \exp(F^*\langle B\rangle/F^*)$, i.e., E/F is a classical n-Kummer extension.*

PROOF. Let $p \in \mathcal{P}_n$. Since $n \mid m$, we have $\mu_p(E) \subseteq \mu_m(E) \subseteq F$, hence E/F is n-pure. Observe that E/F is G-radical, where $G = F^*\langle B\rangle$. By the n-purity criterion (Corollary 4.5), E/F is a G-co-Galois extension. Now, in view of Corollary 5.3, we have $\gcd(n, e(F)) = 1$ and $\zeta_n \in E$. But $n \mid m$, hence $\zeta_n \in E \cap \mu_n(\Omega) = \mu_n(E) \subseteq \mu_m(E) \subseteq F$. Thus E/F is a classical n-Kummer extension. □

6. Applications to elementary field arithmetic and algebraic number theory

6.1. *When is a biquadratic extension Galois, radical, or co-Galois?*

For any $r, d \in \mathbb{Q}$, let $\mathbb{Q}_{r,d} = \mathbb{Q}(\sqrt{r + \sqrt{d}})$. Recall that if $n \in \mathbb{N}^*$ and $a \in \mathbb{R}_+^*$, then $\sqrt[n]{a}$ designates the unique positive root of the equation $x^n - a = 0$, and if $a \in \mathbb{C} \setminus \mathbb{R}_+^*$, then $\sqrt[n]{a}$ designates one of the not specified roots in $\mathbb{C}$ of the polynomial $X^n - a \in \mathbb{C}[X]$. We are interested in when the biquadratic extension $\mathbb{Q}_{r,d}/\mathbb{Q}$ is Galois, co-Galois, or radical.

PROPOSITION 6.1. ([4].) *The following statements are equivalent for $r \in \mathbb{Q}$ and $d \in \mathbb{Q} \setminus \mathbb{Q}^2$.*

(1) *The polynomial $X^4 - 2rX^2 + r^2 - d$ is reducible in $\mathbb{Q}[X]$.*
(2) *There exist $c, k \in \mathbb{Q}_+^*$ such that $r^2 - d = c^2$ and $r \pm c = k^2/2$.*
(3) *There exists a $c \in \mathbb{Q}_+^*$ such that $r^2 - d = c^2$ and $\mathbb{Q}_{r,d} = \mathbb{Q}(\sqrt{(r \pm c)/2})$.*
(4) $[\mathbb{Q}_{r,d} : \mathbb{Q}] = 2$.

COROLLARY 6.2. ([4].) *For any square-free integer $d \in \mathbb{Z} \setminus \{1\}$ and any $n \in \mathbb{Z}$, the polynomial $X^4 - 2nX^2 + n^2 - d$ is irreducible in $\mathbb{Q}[X]$, and so, $[\mathbb{Q}_{n,d} : \mathbb{Q}] = 4$.*

The next result completely answers the question when $\mathbb{Q}_{r,d}/\mathbb{Q}$, $r \in \mathbb{Q}$, $d \in \mathbb{Q} \setminus \mathbb{Q}^2$, is a Galois extension.

PROPOSITION 6.3. ([4].) *The following statements hold for $r \in \mathbb{Q}$ and $d \in \mathbb{Q} \setminus \mathbb{Q}^2$.*

(1) *If the polynomial $X^4 - 2rX^2 + r^2 - d$ is reducible in $\mathbb{Q}[X]$, then $\mathbb{Q}_{r,d}/\mathbb{Q}$ is a Galois extension, and $\mathrm{Gal}(\mathbb{Q}_{r,d}/\mathbb{Q}) \cong \mathbb{Z}_2$.*
(2) *If the polynomial $X^4 - 2rX^2 + r^2 - d$ is irreducible in $\mathbb{Q}[X]$, then $\mathbb{Q}_{r,d}/\mathbb{Q}$ is a Galois extension if and only if $\sqrt{r^2 - d} \in \mathbb{Q}(\sqrt{d})$. In this case, we have*
    (a) $\mathrm{Gal}(\mathbb{Q}_{r,d}/\mathbb{Q}) \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \Leftrightarrow \sqrt{r^2 - d} \in \mathbb{Q}^*$,
    (b) $\mathrm{Gal}(\mathbb{Q}_{r,d}/\mathbb{Q}) \cong \mathbb{Z}_4 \Leftrightarrow \sqrt{r^2 - d} = s\sqrt{d}$ *for some $s \in \mathbb{Q}^*$.*

COROLLARY 6.4. ([4].) *The following assertions are equivalent for an $n \in \mathbb{Z}$ and a square-free integer $d \in \mathbb{Z} \setminus \{1\}$.*

(1) *$\mathbb{Q}_{n,d}/\mathbb{Q}$ is a Galois extension.*
(2) *There exists $k \in \mathbb{Z}^*$ such that either $\sqrt{n^2-d} = k$ or $\sqrt{n^2-d} = k\sqrt{d}$.*
(3) *$n^2 - d \in \mathbb{N}^{*2}$, or there exist $x, y \in \mathbb{Z}^*$ such that $x^2 - dy^2 = -1$ and $n = dy$.*

We are now going to find conditions on the rational numbers r, d for which $\mathbb{Q}_{r,d}/\mathbb{Q}$ is a co-Galois extension.

THEOREM 6.5. ([4].) *The following statements hold for $r \in \mathbb{Q}$ and $d \in \mathbb{Q} \setminus \mathbb{Q}^2$.*
(1) *$\mathbb{Q}_{r,d}/\mathbb{Q}$ is a quadratic co-Galois extension if and only if $\sqrt{r^2-d} \in \mathbb{Q}_+^*$ and, either $2(r - \sqrt{r^2-d}) \in \mathbb{Q}^2$, or $2(r + \sqrt{r^2-d}) \in \mathbb{Q}^2$ and $2(\sqrt{r^2-d} - r)$, $6(\sqrt{r^2-d} - r) \notin \mathbb{Q}^2$.*
(2) *If $[\mathbb{Q}_{r,d} : \mathbb{Q}] = 4$ and $\sqrt{r^2-d} \in \mathbb{Q}$, then $\mathbb{Q}_{r,d}/\mathbb{Q}$ is a co-Galois extension if and only if $-d, -3d, 2(-r \pm \sqrt{r^2-d}), 6(-r \pm \sqrt{r^2-d}) \notin \mathbb{Q}^2$.*
(3) *If $[\mathbb{Q}_{r,d} : \mathbb{Q}] = 4$, $\sqrt{r^2-d} \in \mathbb{Q}(\sqrt{d}) \setminus \mathbb{Q}$, and $-d \notin \mathbb{Q}^2$, then $\mathbb{Q}_{r,d}/\mathbb{Q}$ is not a co-Galois extension.*
(4) *If $[\mathbb{Q}_{r,d} : \mathbb{Q}] = 4$, $\sqrt{r^2-d} \notin \mathbb{Q}(\sqrt{d})$, and either $\sqrt{d-r^2} \notin \mathbb{Q}(\sqrt{d})$ or $d^2 - dr^2 \notin \mathbb{Q}^2$, then $\mathbb{Q}_{r,d}/\mathbb{Q}$ is not a co-Galois extension.*

COROLLARY 6.6. ([4].) *Let $d \in \mathbb{N}, d \geqslant 2$, be a square-free integer, and let $n \in \mathbb{Z}^*$ be such that $\sqrt{n^2-d} \notin \mathbb{Q}(\sqrt{d})$. Then $\mathbb{Q}_{n,d}/\mathbb{Q}$ is not a co-Galois extension.*

Next, we discuss when is $\mathbb{Q}_{r,d}/\mathbb{Q}$ a radical extension. We will mainly consider those $\mathbb{Q}_{r,d}$ which are subfields of $\mathbb{R}$. Since any extension E/F with E a subfield of $\mathbb{R}$ is clearly pure and separable, by the Greither–Harrison criterion it follows that E/F is radical if and only if it is co-Galois. Thus, the radical extensions of type $\mathbb{Q}_{r,d}/\mathbb{Q}$, with $r + \sqrt{d} > 0$, are precisely the co-Galois ones.

PROPOSITION 6.7. ([4].) *The following statements hold for $r \in \mathbb{Q}$ and $d \in \mathbb{Q} \setminus \mathbb{Q}^2$.*
(1) *If $\sqrt{r^2-d} \in \mathbb{Q}$ then $\mathbb{Q}_{r,d}/\mathbb{Q}$ is a radical Galois extension.*
(2) *If $\sqrt{r^2-d} \in \mathbb{Q}(\sqrt{d}) \setminus \mathbb{Q}$, $d > 0$, and $r + \sqrt{d} > 0$, then $\mathbb{Q}_{r,d}/\mathbb{Q}$ is a nonradical Galois extension.*
(3) *If $d \in \mathbb{N}, d \geqslant 2$ is square-free, $r \in \mathbb{Z}^*, r + \sqrt{d} > 0$, and $\sqrt{r^2-d} \notin \mathbb{Q}(\sqrt{d})$, then the extension $\mathbb{Q}_{r,d}/\mathbb{Q}$ is neither Galois nor radical.*

6.2. *Some examples in co-Galois theory with applications to elementary field arithmetic*

In this subsection we present first some very concrete examples, dues to Albu [4], of field extensions, finite or not, involving the concepts of radical, Kneser, Galois, co-Galois, and G-co-Galois extension, and effectively describe the co-Galois groups of some biquadratic and infinite field extensions. Then we are interested to see when a positive algebraic number can or cannot be written as a sum of monomials of form $c \cdot \sqrt[n_1]{a_1}^{\,j_1} \cdot \dots \cdot \sqrt[n_r]{a_r}^{\,j_r}$, with $r, n_1, \dots, n_r, a_1, \dots, a_r \in \mathbb{N}^*$, $j_1, \dots, j_r \in \mathbb{N}$ and $c \in \mathbb{Q}^*$. Finally, we present applications of co-Galois theory to elementary field arithmetic essentially based on the results concerning the primitive elements of G-co-Galois extensions and generalized Kummer extensions discussed in Subsections 4.5 and 5.4.

6.2.1. *A nonradical quartic Galois extension* Consider the quartic Galois extension $\mathbb{Q}_{2,2}/\mathbb{Q}$, where $\mathbb{Q}_{2,2} = \mathbb{Q}(\sqrt{2+\sqrt{2}})$. Then $K = \mathbb{Q}(\sqrt{2})$ is a subfield of $\mathbb{Q}_{2,2}$, and $K/\mathbb{Q}$, $\mathbb{Q}_{2,2}/K$ are both co-Galois extensions, hence also radical, as well as Kneser extensions, but the extension $\mathbb{Q}_{2,2}/\mathbb{Q}$ is neither radical, nor Kneser, and also not co-Galois by Proposition 6.7(2). This example also shows that the property of an extension being radical, Kneser, or co-Galois is, in general, not transitive.

6.2.2. *Cyclotomic Kneser, co-Galois, and G-co-Galois extensions* A direct calculation shows that the cyclotomic extension $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ is $\mathbb{Q}^*\langle\zeta_n\rangle$-Kneser if and only if n is a power of 2. Using Corollaries 5.19 and 5.28, it follows that the extension $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ is G-co-Galois for some group G if and only if $n \in \{1, 2, 3, 4, 6, 8, 12, 24\}$. However, by the Greither–Harrison criterion (Theorem 3.3), the extension $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ is co-Galois if and only if $n \in \{1, 2\}$.

6.2.3. *A nonradical Galois extension of degree* 2^r More generally, consider the extension $E_r/\mathbb{Q}$, where $E_r = \mathbb{Q}(\theta_r)$, and

$$\theta_r = \underbrace{\sqrt{2+\sqrt{2+\sqrt{2+\cdots+\sqrt{2}}}}}_{r \text{ radicals}}.$$

Observe that $E_2 = \mathbb{Q}_{2,2}$, $\theta_r = 2\cos(\pi/2^{r+1}) = \zeta_{2^{r+2}} + \zeta_{2^{r+2}}^{-1}$, and $E_r/\mathbb{Q}$ is a Galois extension having as Galois group the cyclic group of order 2^r generated by the automorphism sending $\cos(\pi/2^{r+1})$ to $\cos(5\pi/2^{r+1})$. However, for every $r \in \mathbb{N}, r \geqslant 2$, the extension $E_r/\mathbb{Q}$ is not a radical extension, for otherwise it would be co-Galois by the Greither–Harrison criterion, so $\operatorname{Cog}(E_r/F) \simeq \operatorname{Gal}(E_r/F) \simeq \mathbb{Z}_{2^r}$ by Corollary 5.19. Consequently, the extension $E_r/\mathbb{Q}$ is neither Kneser nor co-Galois. Notice that for every $r \geqslant 2$, the extension $E_r/\mathbb{Q}$ is a subextension of the cyclotomic extension $\mathbb{Q}(\zeta_{2^{r+2}})/\mathbb{Q}$, which is a Kneser extension by 6.2.2. This shows that a subextension of a radical (respectively Kneser) extension is not necessarily radical (respectively Kneser), in contrast with the fact that any subextension of a co-Galois extension is still co-Galois (see Proposition 3.5).

6.2.4. *An infinite nonradical Abelian extension* Now consider the directed union $E_\infty = \bigcup_{r\geqslant 1} E_r$ of the subfields E_r of the $\mathbb{R}$ defined in 6.2.3. The field E_∞ is a subfield of $\mathbb{Q}^{2,ab} := \bigcup_{r\geqslant 1} \mathbb{Q}(\zeta_{2^r})$, the maximal 2-primary Abelian extension of $\mathbb{Q}$ contained in $\mathbb{C}$. Since the extension $\mathbb{Q}(\zeta_{2^r})/\mathbb{Q}$ is $\mathbb{Q}^*\langle\zeta_{2^r}\rangle$-Kneser for every $r \geqslant 1$, we deduce that the infinite Abelian extension $\mathbb{Q}^{2,ab}/\mathbb{Q}$ is $\mathbb{Q}^*\langle\{\zeta_{2^r} \mid r \geqslant 1\}\rangle$-Kneser. However, its subextension $E_\infty/\mathbb{Q}$ is not Kneser. Thus $E_\infty/\mathbb{Q}$ is a Galois extension of infinite degree which is neither radical, nor Kneser, and nor co-Galois.

6.2.5. *Other nonradical non-Galois quartic extensions* For a given $r \in \mathbb{N}^*$ there are infinitely many d such that the extension $\mathbb{Q}_{r,d}/\mathbb{Q}$ is neither radical nor Galois. Indeed, take as d any prime number with $d > r^2$. Since $r^2 - d < 0$ and $d > 0$ we cannot have $\sqrt{r^2 - d} \in \mathbb{Q}(\sqrt{d})$, hence, $\mathbb{Q}_{r,d}/\mathbb{Q}$ is a nonradical non-Galois quartic extension by Proposition 6.7(3).

6.2.6. *A nonradical non-Galois extension of degree* 2^r Let $r \in \mathbb{N}$, $r \geqslant 2$. We claim that $F_r/\mathbb{Q}$ is such an example, where $F_r = \mathbb{Q}(\nu_r)$ and

$$\nu_r = \underbrace{\sqrt{1 + \sqrt{1 + \sqrt{1 + \cdots + \sqrt{2}}}}}_{r \text{ radicals}}.$$

Indeed, if $F_r/\mathbb{Q}$ would be radical, it would be co-Galois and then, so would be also its subextension $F_2/\mathbb{Q}$, which contradicts Corollary 6.6. An induction on r shows that the extension $F_r/\mathbb{Q}$ has degree 2^r and is not Galois. In particular, for $r = 2$, we deduce that the non-Galois extension $\mathbb{Q}(\sqrt{1 + \sqrt{2}})/\mathbb{Q}$ is neither radical, nor Kneser, and nor co-Galois.

6.2.7. *An infinite nonradical non-Galois extension* Consider the directed union $F_\infty = \bigcup_{r \geqslant 1} F_r$ of subfields F_r of $\mathbb{R}$ defined in 6.2.6. Then $F_\infty/\mathbb{Q}$ is a non-Galois extension of infinite degree which is neither radical, nor Kneser, and nor co-Galois. Indeed, if it would be radical, then necessarily it would be co-Galois, hence any of its subextensions, in particular $F_2/\mathbb{Q}$ would be so, which would contradict Corollary 6.6.

6.2.8. *Calculation of some co-Galois groups* We determine explicitly the co-Galois groups of the extensions $\mathbb{Q}_{r,d}/\mathbb{Q}$ ($r \in \mathbb{Q}, d \in \mathbb{Q} \setminus \mathbb{Q}^2$) and $E_n/\mathbb{Q}$ ($n \in \mathbb{N}^* \cup \{\infty\}$) considered above.

(1) $\operatorname{Cog}(E_n/\mathbb{Q}) = \{\widehat{1}, \widehat{\sqrt{2}}\}$ for every $n \in \mathbb{N}^* \cup \{\infty\}$.

(2) If $c := \sqrt{r^2 - d} \in \mathbb{Q}_+^*$ and $2(r - c) \in \mathbb{Q}^2$, then $\operatorname{Cog}(\mathbb{Q}_{r,d}/\mathbb{Q}) = \{\widehat{1}, \widehat{\sqrt{2(r + c)}}\} \cong \mathbb{Z}_2$.

(3) If $c := \sqrt{r^2 - d} \in \mathbb{Q}_+^*$ and $2(r + c) \in \mathbb{Q}^2$, then

$$\operatorname{Cog}(\mathbb{Q}_{r,d}/\mathbb{Q}) = \begin{cases} \langle \widehat{1 + i} \rangle \cong \mathbb{Z}_4 & \text{if } 2(c - r) \in \mathbb{Q}^2, \\ \langle \widehat{i\sqrt{3}} \cdot (\widehat{1 + i\sqrt{3}}) \rangle \cong \mathbb{Z}_6 & \text{if } 6(c - r) \in \mathbb{Q}^2, \\ \{\widehat{1}, \widehat{\sqrt{2(r - c)}}\} \cong \mathbb{Z}_2 & \text{otherwise.} \end{cases}$$

(4) If $c := \sqrt{r^2 - d} \in \mathbb{Q}_+^*$, $2(r \pm c) \notin \mathbb{Q}^2$, and $-d, -3d, 2(-r \pm c), 6(-r \pm c) \notin \mathbb{Q}^2$, then

$$\operatorname{Cog}(\mathbb{Q}_{r,d}/\mathbb{Q}) = \left\{\widehat{1}, \widehat{\sqrt{2(r + c)}}, \widehat{\sqrt{2(r - c)}}, \widehat{\sqrt{d}}\right\} \cong \mathbb{Z}_2 \times \mathbb{Z}_2.$$

(5) If $\sqrt{r^2 - d} \in \mathbb{Q}(\sqrt{d}) \setminus \mathbb{Q}$, $[\mathbb{Q}_{r,d} : \mathbb{Q}] = 4$, and the extension $\mathbb{Q}_{r,d}/\mathbb{Q}$ is co-Galois, then

$$\operatorname{Cog}(\mathbb{Q}_{r,d}/\mathbb{Q}) = \langle \widehat{1 + i} \rangle \cong \mathbb{Z}_4.$$

(6) If $[\mathbb{Q}_{r,d} : \mathbb{Q}] = 4$ and the extension $\mathbb{Q}_{r,d}/\mathbb{Q}$ is not radical, then $\operatorname{Cog}(\mathbb{Q}_{r,d}/\mathbb{Q}) = \{\widehat{1}, \widehat{\sqrt{d}}\} \cong \mathbb{Z}_2$.

PROOF. We will prove only (1). Let $n \in \mathbb{N}^* \cup \{\infty\}$. The inclusion $\{\widehat{1}, \widehat{\sqrt{2}}\} \subseteq \mathrm{Cog}(E_n/\mathbb{Q})$ is clear. Now let $\widehat{x} \in \mathrm{Cog}(E_n/\mathbb{Q})$. We may assume that $x \in T(E_\infty/\mathbb{Q}) \setminus \mathbb{Q}, x > 0$. Then $x^k = a \in \mathbb{Q}^*$ for some $k \in \mathbb{N}^*$, and let $m \in \mathbb{N}^*$ be minimum with $x^m = a \in \mathbb{Q}^*$. Thus $x = \sqrt[m]{a}$, and $m = \mathrm{ord}(\widehat{x})$. Then, the extension $\mathbb{Q}(\sqrt[m]{a})/\mathbb{Q}$ is a finite radical subextension of $E_\infty/\mathbb{Q}$, so there exists an $r \in \mathbb{N}^*$ such that $\mathbb{Q}(\sqrt[m]{a}) \subseteq E_r$. By 6.2.3, $E_r/\mathbb{Q}$ is an Abelian extension, and so is $\mathbb{Q}(\sqrt[m]{a})/\mathbb{Q}$ as well. Observe that $\mathbb{Q}(\sqrt[m]{a})/\mathbb{Q}$ is a $\mathbb{Q}\langle\sqrt[m]{a}\rangle$-radical extension, and $m = \exp(\mathbb{Q}\langle\sqrt[m]{a}\rangle/\mathbb{Q}^*)$, hence $\zeta_m \in \mathbb{Q}(\sqrt[m]{a})$ by Corollary 5.4. But $\mathbb{Q}(\sqrt[m]{a}) \subseteq \mathbb{R}$, hence necessarily $m = 2$. By Galois theory, the cyclic extension $E_r/\mathbb{Q}$ has a unique quadratic subextension. Since $\mathbb{Q}(\sqrt{a})/\mathbb{Q}$ and $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ are such subextensions, we deduce that $\mathbb{Q}(\sqrt{a}) = \mathbb{Q}(\sqrt{2})$, i.e., $x = \sqrt{a} = b\sqrt{2}$ for some $b \in \mathbb{Q}^*_+$. Thus $\widehat{x} = \widehat{\sqrt{2}}$. This proves that $\mathrm{Cog}(E_n/\mathbb{Q}) = \{\widehat{1}, \widehat{\sqrt{2}}\}$. □

Next, we show that the property of the extension $\mathbb{Q}(\sqrt{2+\sqrt{2}})/\mathbb{Q}$ not being co-Galois (see 6.2.1) can be equivalently expressed more attractively and elementarily as the impossibility to write $\sqrt{2+\sqrt{2}}$ as a finite sum of real numbers of type $\pm\sqrt[n_i]{a_i}$, where $r, n_1, \ldots, n_r, a_1, \ldots, a_r \in \mathbb{N}^*$. The same problem will be discussed for any algebraic number $\alpha \in \mathbb{R}^*_+$.

PROPOSITION 6.8. ([4].) *The following statements are equivalent for a real algebraic number field K.*

(1) *There exist $r \in \mathbb{N}^*$ and $n_1, \ldots, n_r, a_1, \ldots, a_r \in \mathbb{N}^*$ such that*

$$K = \mathbb{Q}\big(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}\big).$$

(2) *The extension $K/\mathbb{Q}$ is radical.*

(3) *The extension $K/\mathbb{Q}$ is Kneser.*

(4) *The extension $K/\mathbb{Q}$ is co-Galois.*

PROOF. (1) ⇒ (4) and (2) ⇒ (4): The extension $K/\mathbb{Q}$ is clearly separable, radical by hypothesis, and pure since $K \subseteq \mathbb{R}$, hence it is co-Galois by the Greither–Harrison criterion.

(4) ⇒ (1): Let $\{x_1, \ldots, x_r\}$ be a set of representatives of the finite group $\mathrm{Cog}(K/\mathbb{Q}) = T(K/\mathbb{Q})/\mathbb{Q}^*$. Since $x_i \equiv -x_i \pmod{\mathbb{Q}^*}$, we may assume that $x_i > 0$ for all i, $1 \leqslant i \leqslant r$. Then $K = \mathbb{Q}(x_1, \ldots, x_r)$, and for every i, $1 \leqslant i \leqslant r$, there exists $n_i \in \mathbb{N}^*$ such that $x_i^{n_i} = a_i \in \mathbb{Q}$. Clearly, $a_i > 0$ for all i. Then $K = \mathbb{Q}(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r})$. Of course, we may also assume that all $a_i \in \mathbb{N}^*$.

The other implications are obvious. □

COROLLARY 6.9. ([4].) *The following assertions are equivalent for an algebraic number $\alpha \in \mathbb{R}^*_+$.*

(1) *α can be written as a finite sum of real numbers of type $\pm\sqrt[n_i]{a_i}$, $1 \leqslant i \leqslant r$, r, n_i, $a_i \in \mathbb{N}^*$.*

(2) *The extension $\mathbb{Q}(\alpha)/\mathbb{Q}$ is radical.*

(3) *The extension $\mathbb{Q}(\alpha)/\mathbb{Q}$ is Kneser.*

(4) *The extension $\mathbb{Q}(\alpha)/\mathbb{Q}$ is co-Galois.*

We deduce from Corollary 6.9 that each of the numbers

$$\sqrt{1+\sqrt{2}}, \qquad \sqrt{2+\sqrt{2+\sqrt{2+\cdots+\sqrt{2}}}}$$

cannot be written as a finite sum of real numbers of type $\pm\sqrt[n_i]{a_i}$, $1 \leqslant i \leqslant r$, where $r, n_i, a_i \in \mathbb{N}^*$; also if a square-free integer $d \geqslant 2$ and $r \in \mathbb{Z}^*$ are such that $r > -\sqrt{d}$ and $\sqrt{r^2-d} \notin \mathbb{Q}(\sqrt{d})$, then $\sqrt{r+\sqrt{d}}$ cannot be written as a finite sum of real numbers of type $\pm\sqrt[n_i]{a_i}$, $1 \leqslant i \leqslant r$, with $r, n_i, a_i \in \mathbb{N}^*$, by Proposition 6.7(3).

We are now going to present further applications of co-Galois theory to elementary field arithmetic, essentially based on Theorem 4.14, Proposition 4.16, and Theorem 5.25.

PROPOSITION 6.10. ([17].) *Let F be an arbitrary field, let $r, n_1, \ldots, n_r \in \mathbb{N}^*$, let $a_1, \ldots, a_r \in F^*$, and let $n = \operatorname{lcm}(n_1, \ldots, n_r)$. Assume that $\gcd(n, e(F)) = 1$, $\mu_n(F(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r})) \subseteq F$, and $\widehat{\sqrt[n_i]{a_i}} \neq \widehat{\sqrt[n_j]{a_j}}$ for all $i \neq j$ in $\{1, \ldots, r\}$. Then $F(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}) = F(\sqrt[n_1]{a_1} + \cdots + \sqrt[n_r]{a_r})$.*

PROOF. The first two conditions show that $F(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r})/F$ is a generalized n-Kummer extension, hence according to Theorem 5.25, it is $F^*\langle\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}\rangle$-co-Galois. This fact and the third condition imply the desired equality in view of Theorem 4.14. □

COROLLARY 6.11. ([17].) *Let F be a subfield of $\mathbb{R}$, let $r, n_1, \ldots, n_r \in \mathbb{N}^*$, and let $a_1, \ldots, a_r \in F^*$ be positive numbers. Then $F(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}) = F(\sqrt[n_1]{a_1} + \cdots + \sqrt[n_r]{a_r})$. In particular, if $a_1, \ldots, a_r \in \mathbb{Q}_+^*$, then $\mathbb{Q}(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}) = \mathbb{Q}(\sqrt[n_1]{a_1} + \cdots + \sqrt[n_r]{a_r})$.*

COROLLARY 6.12. ([17].) *Let F be a subfield of $\mathbb{R}$, let $r, n_1, \ldots, n_r \in \mathbb{N}^*$, and let $a_1, \ldots, a_r \in F^*$ be positive numbers. Then $\sqrt[n_1]{a_1} + \cdots + \sqrt[n_r]{a_r} \in F \Leftrightarrow \sqrt[n_i]{a_i} \in F$ for all i, $1 \leqslant i \leqslant r$. In particular, if $a_1, \ldots, a_r \in \mathbb{Q}_+^*$, then $\sqrt[n_1]{a_1} + \cdots + \sqrt[n_r]{a_r} \in \mathbb{Q} \Leftrightarrow \sqrt[n_i]{a_i} \in \mathbb{Q}$ for all i, $1 \leqslant i \leqslant r$.*

PROOF. If $\sqrt[n_1]{a_1} + \cdots + \sqrt[n_r]{a_r} \in F$, then $F(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}) = F(\sqrt[n_1]{a_1} + \cdots + \sqrt[n_r]{a_r}) = F$ by Corollary 6.11, and consequently $\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r} \in F$. □

COROLLARY 6.13. ([17].) *Let $n, r \in \mathbb{N}^*$, let F be a field with $\gcd(n, e(F)) = 1$, and let $a_1, \ldots, a_r \in F^*$. Suppose that $\mu_n(F(\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_r})) \subseteq F$ and*

$$\left[F\left(\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_r}\right) : F\right] = \prod_{i=1}^{r}\left[F\left(\sqrt[n]{a_i}\right) : F\right].$$

Then

$$F\left(\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_r}\right) = F\left(\sqrt[n]{a_1} + \cdots + \sqrt[n]{a_r}\right).$$

PROOF. By Theorem 5.25, the extension $F(\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_r})/F$ is $F^*\langle\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}\rangle$-co-Galois. Now apply Proposition 4.16. □

PROPOSITION 6.14. ([4].) *Let $r, n_0, n_1, \ldots, n_r \in \mathbb{N}^*$ and $a_0, a_1, \ldots, a_r \in F^*$. Let $n = \operatorname{lcm}(n_0, n_1, \ldots, n_r)$, and suppose that $\gcd(n, e(F)) = 1$ and*

$$\mu_n\big(F\big(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}\big)\big) \subseteq F.$$

Then $\sqrt[n_0]{a_0} \in F(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r})$ if and only if there exist $j_1, \ldots, j_r \in \mathbb{N}$ and $c \in F^*$ such that $\sqrt[n_0]{a_0} = c \cdot \sqrt[n_1]{a_1}^{\,j_1} \cdot \ldots \cdot \sqrt[n_r]{a_r}^{\,j_r}$.

PROOF. Set $E = F(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r})$ and $G = F^*\langle \sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r}\rangle$. By Theorem 5.25, E/F is a G-co-Galois extension. Now, observe that since $n_0 \mid n$, we have $\mu_{n_0}(E) \subseteq \mu_n(E) \subseteq F$. So, we can apply Lemma 4.10 to deduce that $\sqrt[n_0]{a_0} \in F(\sqrt[n_1]{a_1}, \ldots, \sqrt[n_r]{a_r})$ if and only if $\sqrt[n_0]{a_0} \in G$, as desired. □

COROLLARY 6.15. ([4].) *Let $r, n \in \mathbb{N}^*$, let $a_0, a_1, \ldots, a_r \in F^*$, and suppose that $\gcd(n, e(F)) = 1$, and $\mu_n(F(\sqrt[n]{a_0}, \sqrt[n]{a_1}, \ldots, \sqrt[n]{a_r})) \subseteq \{-1, 1\}$ or $\mu_n(\Omega) \subseteq F$, i.e., the extension $F(\sqrt[n]{a_0}, \sqrt[n]{a_1}, \ldots, \sqrt[n]{a_r})/F$ is either an n-Kummer extension with few roots of unity or a classical n-Kummer extension. Then $\sqrt[n]{a_0} \in F(\sqrt[n]{a_1}, \ldots, \sqrt[n]{a_r})$ if and only if there exist $j_1, \ldots, j_r \in \mathbb{N}$ and $c \in F^*$ such that $a_0 = c^n \cdot a_1^{j_1} \cdot \ldots \cdot a_r^{j_r}$.*

COROLLARY 6.16. ([4].) *Let $r, n_0, n_1, \ldots, n_r \in \mathbb{N}^*$, and let $a_0, a_1, \ldots, a_r \in \mathbb{Q}_+^*$. Then $\sqrt[n_0]{a_0}$ can be written as a finite sum of monomials of form $c \cdot \sqrt[n_1]{a_1}^{\,j_1} \cdot \ldots \cdot \sqrt[n_r]{a_r}^{\,j_r}$, with $j_1, \ldots, j_r \in \mathbb{N}$ and $c \in \mathbb{Q}^*$, if and only if $\sqrt[n_0]{a_0}$ is itself such a monomial.*

We will now examine the elementary arithmetic of finite G-radical extensions E/F with $\exp(G/F^*)$ a prime number p, in other words, the elementary arithmetic of finite p-bounded radical extensions.

LEMMA 6.17. ([9,10].) *Let F be an arbitrary field, let $p > 0$ be a prime number, other than the characteristic of F, let $r \in \mathbb{N}^*$, let $a_1, \ldots, a_r \in F^*$, and let $\sqrt[p]{a_1}, \ldots, \sqrt[p]{a_r} \in \Omega$ denote any fixed p-th roots. Assume that $[F(\sqrt[p]{a_1}, \ldots, \sqrt[p]{a_r}) : F] = p^r$. Then, we have either $\zeta_p \in F$ or $\mu_p(F(\sqrt[p]{a_1}, \ldots, \sqrt[p]{a_r})) = \{1\}$, in other words, $F(\sqrt[p]{a_1}, \ldots, \sqrt[p]{a_r})/F$ is either a classical p-Kummer extension or a p-Kummer extension with few roots of unity.*

PROPOSITION 6.18. ([9,10].) *Let F be an arbitrary field, let $p > 0$ be a prime number, other than the characteristic of F, let $r \in \mathbb{N}^*$, let $a_1, \ldots, a_r \in F^*$, and let $\sqrt[p]{a_1}, \ldots, \sqrt[p]{a_r} \in \Omega$ denote any fixed p-th roots. If $[F(\sqrt[p]{a_1}, \ldots, \sqrt[p]{a_r}) : F] = p^r$, then the following assertions hold.*

(1) *The extension $F(\sqrt[p]{a_1}, \ldots, \sqrt[p]{a_r})/F$ is $F^*\langle \sqrt[p]{a_1}, \ldots, \sqrt[p]{a_r}\rangle$-co-Galois.*

(2) *The Kneser group $F^*\langle \sqrt[p]{a_1}, \ldots, \sqrt[p]{a_r}\rangle/F^*$ of the extension E/F is isomorphic to the direct product $\mathbb{Z}_p^r$ of r copies of the group $\mathbb{Z}_p$.*

(3) *$|\langle \widehat{\sqrt[p]{a_1}}, \ldots, \widehat{\sqrt[p]{a_r}}\rangle| = |\langle \widehat{\widehat{a_1}}, \ldots, \widehat{\widehat{a_r}}\rangle| = p^r$, where $\widehat{\widehat{a}}$ denotes for any $a \in F^*$ its coset in the group F^*/F^{*p}.*

(4) *If $i_1, \ldots, i_n \in \mathbb{N}$ and $a_1^{i_1} \cdot \ldots \cdot a_r^{i_r} \in F^{*p}$, then $p \mid i_1, \ldots, p \mid i_r$.*

(5) *$F(\sqrt[p]{a_1}, \ldots, \sqrt[p]{a_r}) = F(\sqrt[p]{a_1} + \cdots + \sqrt[p]{a_r})$.*

COROLLARY 6.19. ([53].) *Let p be a prime other than the characteristic of a field F. Let $u, v \in \Omega$ be any roots of the irreducible polynomials $X^p - a$ and $X^p - b \in F[X]$, respectively. Then $[F(u, v) : F] = p^2$ unless $b = c^p a^n$ for some $c \in F$ and $n \in \mathbb{N}$, and if $[F(u, v) : F] = p^2$ then $F(u, v) = F(u + v)$.*

COROLLARY 6.20. ([24].) *Let F be an arbitrary field, let $p > 0$ be a prime number, other than the characteristic of F, let $r \in \mathbb{N}, r \geqslant 2$, let $a_1, \ldots, a_r \in F^*$, and let $\sqrt[p]{a_1}, \ldots, \sqrt[p]{a_r} \in \Omega$ denote any fixed p-th roots. Further, let denote $E = F(\sqrt[p]{a_1}, \ldots, \sqrt[p]{a_{r-1}})$. Then either $[E(\sqrt[p]{a_r}) : E] = p$, or $a_r = c^p \cdot a_1^{j_1} \cdot \cdots \cdot a_{r-1}^{j_{r-1}}$ for some $j_1, \ldots, j_r \in \mathbb{N}$ and $c \in F^*$.*

6.3. *Some classical results on algebraic number fields via co-Galois theory*

We show in this subsection how some classical results on algebraic number fields, chronologically due to Hasse, Besicovitch, Mordell, Siegel, etc., can be immediately deduced from basic facts of co-Galois theory, especially from the Kneser criterion. Throughout this subsection K denotes a fixed algebraic number field, and for any $a \in K^*$ and $n \in \mathbb{N}^*$, $\sqrt[n]{a}$ designates a root (which in general is not specified) in $\mathbb{C}$ of the polynomial $X^n - a \in K[X]$. However, if K is a subfield of $\mathbb{R}$ and $a > 0$, then $\sqrt[n]{a}$ will always mean the unique positive root in $\mathbb{R}$ of the polynomial $X^n - a$.

THEOREM 6.21 (Hasse [50], 1930). *Let K be an algebraic number field containing ζ_n for some $n \in \mathbb{N}^*$, let $r \in \mathbb{N}^*$, and let $x_1, \ldots, x_r \in \mathbb{C}^*$ be such that $x_k^n \in K$ for all k, $1 \leqslant k \leqslant r$. Assume that the following condition is satisfied:*

$$m_1, \ldots, m_r \in \mathbb{N} \text{ and } x_1^{m_1} \cdot \cdots \cdot x_r^{m_r} \in K \quad \Rightarrow \quad x_k^{m_k} \in K, k = 1, \ldots, r. \qquad (\dagger)$$

Then $[K(x_1, \ldots, x_r) : K] = \prod_{1 \leqslant k \leqslant r} |K^*\langle x_k \rangle / K^*|$.

PROOF. Clearly, $K(x_1, \ldots, x_r)/K$ is a classical n-Kummer extension, so it is a G-co-Galois extension by Theorem 5.23, where $G = K\langle x_1, \ldots, x_r \rangle$. In particular, the extension $K(x_1, \ldots, x_r)/K$ is G-Kneser, hence

$$\big[K(x_1, \ldots, x_r) : K\big] = \big|G/K^*\big| = \big|K^*\langle x_1, \ldots, x_r \rangle / K^*\big|.$$

Now apply Lemma 4.15 to deduce that

$$\big|K^*\langle x_1, \ldots, x_r \rangle / K^*\big| = \prod_{1 \leqslant k \leqslant r} \big|K^*\langle x_k \rangle / K^*\big|. \qquad \square$$

THEOREM 6.22 (Siegel [73], 1972). *Let K be an algebraic number field, let $r \in \mathbb{N}^*$, let $n_1, \ldots, n_r \in \mathbb{N}^*$, and let $x_1, \ldots, x_r \in \mathbb{C}$ be such that $x_k^{n_k} \in K$ for all k, $1 \leqslant k \leqslant r$. Assume that either $\zeta_{n_k} \in K$ for all k, $1 \leqslant k \leqslant r$, or $K \subseteq \mathbb{R}$ and $x_k \in \mathbb{R}_+^*$ for all k, $1 \leqslant k \leqslant r$. Then $[K(x_1, \ldots, x_r) : K] = |K^*\langle x_1, \ldots, x_r \rangle / K^*|$.*

PROOF. Set $G = K\langle x_1, \ldots, x_r \rangle$ and $L = K(x_1, \ldots, x_r)$. Observe that L/K is a separable G-radical extension, and so, the desired equality means precisely that L/K is a G-Kneser extension. If $K \subseteq \mathbb{R}$ and $x_k \in \mathbb{R}_+^*$ for all k, $1 \leqslant k \leqslant r$, then $L \subseteq \mathbb{R}$, hence L/K is clearly a pure extension, and thus, it is G-Kneser. Now, assume that $\zeta_{n_k} \in K$ for all k, $1 \leqslant k \leqslant r$, and let $n = \operatorname{lcm}(n_1, \dots, n_r)$. Then $n = n_k b_k$ for some $b_k \in \mathbb{Z}$, $1 \leqslant k \leqslant r$. Since $\gcd(b_1, \dots, b_r) = 1$, $1 = \sum_{1 \leqslant k \leqslant r} b_k c_k$ for some $c_1, \dots, c_r \in \mathbb{Z}$. It follows that $\zeta_n = \zeta_n^{b_1c_1+\cdots+b_rc_r} = (\zeta_n^{b_1})^{c_1} \cdot \dots \cdot (\zeta_n^{b_r})^{c_r} \in K$, since $\zeta_{n_k} \in K$ and $\zeta_n^{b_k}$ is a primitive n_k-th root of unity for all $k = 1, \dots, r$. Using the Kneser criterion, this easily implies that L/K is a G-Kneser extension. □

THEOREM 6.23 (Mordell [62], 1953). *Let K be an algebraic number field, let $r \in \mathbb{N}^*$, let $n_1, \dots, n_r \in \mathbb{N}^*$, and let $x_1, \dots, x_r \in \mathbb{C}$ be such that $x_k^{n_k} \in K$ for all k, $1 \leqslant k \leqslant r$, and*

$$m_1, \dots, m_r \in \mathbb{N} \text{ and } x_1^{m_1} \cdot \dots \cdot x_r^{m_r} \in K \quad \Rightarrow \quad n_k \mid m_k, k = 1, \dots, r. \qquad (\dagger\dagger)$$

Assume that either $\zeta_{n_k} \in K$ for all k, $1 \leqslant k \leqslant r$, or $K \subseteq \mathbb{R}$ and $x_k \in \mathbb{R}_+^*$ for all k, $1 \leqslant k \leqslant r$. Then $[K(x_1, \dots, x_r) : K] = n_1 \cdot \dots \cdot n_r$.

PROOF. The extension $K(x_1, \dots, x_r)/K$ is $K\langle x_1, \dots, x_r\rangle$-Kneser by Theorem 6.22. To conclude, observe that condition $(\dagger\dagger)$ is the same as condition $(\dagger)$ in Theorem 6.21. Now, apply Lemma 4.15 to obtain the desired equality. □

COROLLARY 6.24 (Ursell [79], 1974). *Let $r \in \mathbb{N}$, $r \geqslant 2$, and let $a_1, \dots, a_r, n_1, \dots, n_r \in \mathbb{N}^*$ be such that for every k, $1 \leqslant k \leqslant r$, and every s_k, $1 \leqslant s_k < n_k$ one has $(\sqrt[n_k]{a_k})^{s_k} \notin \mathbb{N}$. If $a_1, \dots, a_k$ are relatively prime in pairs, then $[\mathbb{Q}(\sqrt[n_1]{a_1}, \dots, \sqrt[n_r]{a_r}) : \mathbb{Q}] = n_1 \cdot \dots \cdot n_r$.*

PROOF. Observe that the condition "$a_1, \dots, a_k$ are relatively prime in pairs" implies the condition $(\dagger\dagger)$ in Theorem 6.23. □

THEOREM 6.25 (Besicovitch [30], 1940). *Let $r \in \mathbb{N}^*$, let $p_1, \dots, p_r$ be different positive prime integers, let $b_1, \dots, b_r \in \mathbb{N}^*$ be not divisible by any of these primes, and let $a_1 = b_1 p_1, \dots, a_r = b_r p_r$. Then, for any $n_1, \dots, n_r \in \mathbb{N}^*$ one has $[\mathbb{Q}(\sqrt[n_1]{a_1}, \dots, \sqrt[n_r]{a_r}) : \mathbb{Q}] = n_1 \cdot \dots \cdot n_r$.*

PROOF. One easily shows that the condition $(\dagger\dagger)$ in Theorem 6.23 is satisfied. □

COROLLARY 6.26. *Let $r \in \mathbb{N}^*$, let $p_1, \dots, p_r$ be different positive prime integers, and let $n_1, \dots, n_r \in \mathbb{N}^*$ be arbitrary. Then $[\mathbb{Q}(\sqrt[n_1]{p_1}, \dots, \sqrt[n_r]{p_r}) : \mathbb{Q}] = n_1 \cdot \dots \cdot n_r$.*

PROOF. Apply Theorem 6.25 for $b_1 = \dots = b_r = 1$. □

REMARKS.

(1) In view of Theorem 5.23 and Lemma 4.15, Theorem 6.21 holds not only for algebraic number fields K, but also for any field K and any $n \in \mathbb{N}^*$ such that $\gcd(n, e(F)) = 1$. Of course, the field $\mathbb{C}$ in the statement of Theorem 6.21 should be replaced by an algebraically closed field Ω containing K as a subfield. Also, according to Theorem 5.25 and Lemma 4.15, Theorems 6.22 and 6.23 are valid for

any field K, any $n_1, \dots, n_r \in \mathbb{N}^*$, and any $x_1, \dots, x_r \in \Omega$ with $x_k^{n_k} \in K$ for all k, $1 \leqslant k \leqslant r$, such that $\gcd(n, e(K)) = 1$ and $\mu_n(K(x_1, \dots, x_r)) \subseteq K$, where $n = \operatorname{lcm}(n_1, \dots, n_r)$.

(2) We have seen that the conditions in Theorem 6.25 as well as those in Corollary 6.24 imply for $K = \mathbb{Q}$ the condition (††) in Theorem 6.23. Observe that the extension $\mathbb{Q}(\sqrt{2}, \sqrt{6})/\mathbb{Q}$ satisfies the condition (††) in Theorem 6.23 with $r = 2, n_1 = n_2 = 2, x_1 = \sqrt{2}, x_2 = \sqrt{6}$, but satisfies neither the conditions in Corollary 6.24 nor the conditions in Theorem 6.25. On the other hand, the extension $\mathbb{Q}(\sqrt{8}, \sqrt{3})/\mathbb{Q}$ satisfies the conditions in Corollary 6.24, but does not satisfy the conditions in Theorem 6.25. Also, the extension $\mathbb{Q}(\sqrt{6}, \sqrt{10})/\mathbb{Q}$ satisfies the conditions in Theorem 6.25, but does not satisfy the ones in Corollary 6.24. This shows that we cannot deduce Corollary 6.24 from Theorem 6.25, and vice versa.

(3) Corollary 6.24 and Theorem 6.25 can be extended from $\mathbb{Q}$ to the field of quotients of any UFD.

Necessary and sufficient conditions for an arbitrary radical extension $F(x_1, \dots, x_r)/F$ to have degree $n_1 \cdot \cdots \cdot n_r$, where $x_1, \dots, x_r \in \Omega^*$ are such that $x_i^{n_i} \in F^*$ for every i, $1 \leqslant i \leqslant r$, are provided in the next result.

THEOREM 6.27 (Schinzel [70]). *Let F be any field, let $r \in \mathbb{N}^*$, let $n_1, \dots, n_r \in \mathbb{N}^*$ be positive integers such that at most one of them is divisible by the characteristic of F, let $a_1, \dots, a_r \in F^*$, and let $x_1, \dots, x_r \in \Omega$ be such that $x_i^{n_i} = a_i$ for every $i = 1, \dots, r$. Then $[F(x_1, \dots, x_r) : F] = n_1 \cdot \cdots \cdot n_r$ if and only if the following two conditions are satisfied.*

(a) *Whenever $p \in \mathbb{P}$ and $k_1, \dots, k_r \in \mathbb{N}$ are such that $p \mid n_i k_i$ for every $i = 1, \dots, r$ and $a_1^{k_1} \cdot \cdots \cdot a_r^{k_r} \in F^p$, then $p \mid k_i$ for every $i = 1, \dots, r$.*

(b) *Whenever $k_1, \dots, k_r \in \mathbb{N}$ are such that $4 \mid n_i k_i$ for every $i = 1, \dots, r$ and $a_1^{k_1} \cdot \cdots \cdot a_r^{k_r} \in -4F^4$, then $p \mid k_i$ for every $i = 1, \dots, r$.*

EXAMPLE. We are going to calculate the degree $[\mathbb{Q}(\sqrt[4]{12} + \sqrt[6]{108}) : \mathbb{Q}]$ and to exhibit a vector space basis of the extension $\mathbb{Q}(\sqrt[4]{12} + \sqrt[6]{108})/\mathbb{Q}$. To do this, we first apply Corollary 6.11 to deduce that $\mathbb{Q}(\sqrt[4]{12} + \sqrt[6]{108}) = \mathbb{Q}(\sqrt[4]{12}, \sqrt[6]{108})$. Since $\sqrt[4]{12} = \sqrt[12]{2^6 \cdot 3^3}$ and $\sqrt[6]{108} = \sqrt[12]{2^4 \cdot 3^6}$, we have $\mathbb{Q}(\sqrt[4]{12}, \sqrt[6]{108}) = \mathbb{Q}(\sqrt[12]{2^6 \cdot 3^3}, \sqrt[12]{2^4 \cdot 3^6})$. Set $E = \mathbb{Q}(\sqrt[12]{2^6 \cdot 3^3}, \sqrt[12]{2^4 \cdot 3^6})$, and observe that $E/\mathbb{Q}$ is a 12-Kummer extension with few roots of unity; so $[E : \mathbb{Q}] = |\mathbb{Q}^*\langle \sqrt[12]{a}, \sqrt[12]{b}\rangle/\mathbb{Q}^*| = |\langle \widehat{\sqrt[12]{a}, \sqrt[12]{b}}\rangle| = |\langle \widehat{\widehat{a}}, \widehat{\widehat{b}}\rangle|$ by Theorem 5.25, where $a = 2^4 \cdot 3^6$, $b = 2^6 \cdot 3^3$, and $\widehat{\widehat{x}}$ denotes for any $x \in \mathbb{Q}^*$ its coset $x\mathbb{Q}^{*12}$ in the group $\mathbb{Q}^*/\mathbb{Q}^{*12}$. We describe now explicitly the group $\langle \widehat{a}, \widehat{b}\rangle$. Since $\operatorname{ord}(\widehat{a}) = 6$, $\operatorname{ord}(\widehat{\widehat{b}}) = 4$, and $\widehat{\widehat{b^2}} = \widehat{\widehat{a^3}} = \widehat{\widehat{3^6}}$, we have $\langle \widehat{\widehat{a}}, \widehat{\widehat{b}}\rangle = \{\widehat{\widehat{a^i}} \cdot \widehat{\widehat{b^j}} \mid 0 \leqslant i \leqslant 5, 0 \leqslant j \leqslant 1\} = \{\widehat{\widehat{1}}, \widehat{\widehat{a}}, \widehat{\widehat{a^2}}, \widehat{\widehat{a^3}}, \widehat{\widehat{a^4}}, \widehat{\widehat{a^5}}, \widehat{\widehat{b}}, \widehat{\widehat{a}} \cdot \widehat{\widehat{b}}, \widehat{\widehat{a^2}} \cdot \widehat{\widehat{b}}, \widehat{\widehat{a^3}} \cdot \widehat{\widehat{b}}, \widehat{\widehat{a^4}} \cdot \widehat{\widehat{b}}, \widehat{\widehat{a^5}} \cdot \widehat{\widehat{b}}\}$. Note that $\widehat{\widehat{b}} \notin \langle \widehat{\widehat{a}}\rangle$, and consequently $|\langle \widehat{\widehat{a}}, \widehat{\widehat{b}}\rangle| = 12$. Thus $[E : \mathbb{Q}] = 12$, and, by Proposition 3.1, a basis of the extension $\mathbb{Q}(\sqrt[4]{12} + \sqrt[6]{108})/\mathbb{Q}$ is the set $\{\sqrt[4]{12}^i \cdot \sqrt[6]{108}^j \mid 0 \leqslant i \leqslant 5, 0 \leqslant j \leqslant 1\}$.

6.4. *Applications to Hecke systems of ideal numbers*

The Kneser criterion (Theorem 3.2) is not only a powerful as well as indispensable tool in investigating radical field extensions, but, as we have already seen in the previous subsection, it has nice applications in proving some classical results of algebraic number theory. In this subsection we present other such applications.

A classical construction from 1920 in algebraic number theory is the following one: to every algebraic number field K one can associate a so-called *system of ideal numbers* S, which is a certain subgroup of the multiplicative group $\mathbb{C}^*$ of complex numbers such that $K^* \leqslant S$ and the quotient group S/K^* is canonically isomorphic to the ideal class group $\mathcal{C}\ell_K$ of K. This construction, originating with Hecke [51], has the following important property, that the Hilbert class field also possesses: every ideal of K becomes a principal ideal in the algebraic number field $K(S)$. The equality $[K(S) : K] = |\mathcal{C}\ell_K|$ was claimed by Hecke on page 122 of his monograph [52] published in 1948, but never proved by him. To the best of our knowledge, no proof of this assertion, excepting the very short one due to Albu and Nicolae [18] and reproduced below, is available in the literature. Note that Ribenboim gives on page 124 of his monograph [65] only the inequality $[K(S) : K] \leqslant h$.

The main aim of this subsection is to provide a short proof of this equality by using the Kneser criterion, and to discuss some other related questions.

First let us fix the notation and terminology needed to explain the equality mentioned above. Throughout this subsection K will denote a fixed algebraic number field. We will denote by $\mathfrak{O}_K$ the ring of algebraic integers of K, by $\mathcal{F}_K$ the group of nonzero fractional ideals of K, by $\mathcal{P}_K$ the group of nonzero principal fractional ideals of K, by $\mathcal{C}\ell_K = \mathcal{F}_K/\mathcal{P}_K$ the ideal class group of K, and by $h = |\mathcal{C}\ell_K|$ the class number of K. For any $a \in K^*$, (a) will denote the principal fractional ideal $a\mathfrak{O}_K$ of K.

As any finite Abelian group, the group $\mathcal{C}\ell_K$ is an internal direct sum of finitely many cyclic subgroups. This means that there exist ideal classes $\mathcal{C}_1, \dots, \mathcal{C}_s$ in $\mathcal{C}\ell_K$, $s \geqslant 1$, such that every ideal class $\mathcal{C} \in \mathcal{C}\ell_K$ has a unique decomposition $\mathcal{C} = \mathcal{C}_1^{r_1} \cdot \dots \cdot \mathcal{C}_s^{r_s}$, where $0 \leqslant r_k < h_k$, $h_k > 1$ is the order of the ideal class $\mathcal{C}_k$ in $\mathcal{C}\ell_K$, $k = 1, \dots, s$, and $h = h_1 \cdot \dots \cdot h_s$. For any $k = 1, \dots, s$, let I_k be an integral ideal from the ideal class $\mathcal{C}_k$. Then, every fractional ideal $I \in \mathcal{F}_K$ has a unique decomposition $I = (a)I_1^{r_1} \cdot \dots \cdot I_s^{r_s}$, where $a \in K^*$, $0 \leqslant r_k < h_k$, $k = 1, \dots, s$, and the exponents r_k are uniquely determined. Since $\mathcal{C}_k^{h_k} = 1$, one deduces that $I_k^{h_k} = (c_k) \in \mathcal{P}_K$ for suitable numbers $c_k \in K^*$, $k = 1, \dots, s$, which are uniquely determined up to units of K. Assume that we have fixed the numbers c_k and consider the number field $K(\gamma_1, \dots, \gamma_s)$, where γ_k is a complex root of the polynomial $X^{h_k} - c_k \in K[X]$, i.e., $\gamma_k^{h_k} = c_k$.

Definition. With the notation above, the group $K^*\langle\gamma_1, \dots, \gamma_s\rangle$ is called a *Hecke system of ideal numbers* of K, and the field $K(\gamma_1, \dots, \gamma_s)$, denoted by H_K, is called the *Hecke field* of K associated with the Hecke system of ideal numbers $K^*\langle\gamma_1, \dots, \gamma_s\rangle$ of K.

The Hecke field of any algebraic number field K, which clearly depends on the chosen Hecke system of ideal numbers of K, is uniquely determined up to a K-isomorphism. Therefore, we will just call it the *Hecke field* of K. Observe that $H_K = K \Leftrightarrow h = 1$, so we may assume that $h > 1$. The group morphism

$$\mathcal{F}_K \to K^*\langle\gamma_1, \dots, \gamma_s\rangle/K^*, \quad (a)I_1^{r_1} \cdot \cdots \cdot I_s^{r_s} \longmapsto \widehat{\gamma_1^{r_1} \cdot \cdots \cdot \gamma_s^{r_s}},$$

with $a \in K^*$, $0 \leqslant r_k < h_k$, $k = 1, \dots, s$, clearly induces a surjective group morphism

$$\psi_K : \mathcal{C}\ell_K \to K^*\langle\gamma_1, \dots, \gamma_s\rangle/K^*.$$

LEMMA 6.28. ([63,18].) *With the notation above, the map*

$$\psi_K : \mathcal{C}\ell_K \to K^*\langle\gamma_1, \dots, \gamma_s\rangle/K^*$$

is a group isomorphism. In particular, one has $|K^*\langle\gamma_1, \dots, \gamma_s\rangle/K^*| = |\mathcal{C}\ell_K| = h$.

LEMMA 6.29. *If* $\varepsilon \in K^*\langle\gamma_1, \dots, \gamma_s\rangle$ *is a unit of* H_K, *then* $\varepsilon \in K$.

THEOREM 6.30. ([51,18].) *Let* K *be an algebraic number field, and let* H_K *be its Hecke field. Then* $[H_K : K] = h$.

PROOF. By Lemma 6.28, we have to prove that

$$\bigl[K(\gamma_1, \dots, \gamma_s) : K\bigr] = \bigl|K^*\langle\gamma_1, \dots, \gamma_s\rangle/K^*\bigr| = h.$$

Since $K(\gamma_1, \dots, \gamma_s) = K(K^*\langle\gamma_1, \dots, \gamma_s\rangle)$ and $\gamma_i^{h_i} = c_i \in K^*$ for all $i = 1, \dots, s$, it follows that the extension $K(\gamma_1, \dots, \gamma_s)/K$ is $K^*\langle\gamma_1, \dots, \gamma_s\rangle$-radical. Consequently, the desired equality means precisely that the extension $K(\gamma_1, \dots, \gamma_s)/K$ is $K^*\langle\gamma_1, \dots, \gamma_s\rangle$-Kneser. To prove that, we will check that the conditions from the Kneser criterion are satisfied. Let p be an odd prime such that $\zeta_p \in K^*\langle\gamma_1, \dots, \gamma_s\rangle$. Then $\zeta_p \in K$ by Lemma 6.29. Now, assume that $1 + \zeta_4 \in K^*\langle\gamma_1, \dots, \gamma_s\rangle$. Then $(1 + \zeta_4)^2 = 2\zeta_4 \in K^*\langle\gamma_1, \dots, \gamma_s\rangle$, so $\zeta_4 \in K^*\langle\gamma_1, \dots, \gamma_s\rangle$. Thus, $\zeta_4 \in K$ again by Lemma 6.29. This proves that H_K/K is a $K^*\langle\gamma_1, \dots, \gamma_s\rangle$-Kneser extension, as desired. □

COROLLARY 6.31. ([18].) *Let* $\mathcal{C} \in \mathcal{C}\ell_K$ *and let* $\gamma \in K^*\langle\gamma_1, \dots, \gamma_s\rangle$ *be such that* $\widehat{\gamma} = \psi_K(\mathcal{C})$, *where* ψ_K *is the isomorphism in Lemma 6.28. If* $m = \operatorname{ord}(\mathcal{C})$ *in* $\mathcal{C}\ell_K$, *then* $[K(\gamma) : K] = m$, *and every ideal of* $\mathcal{C}$ *becomes a principal ideal in* $K(\gamma)$.

PROOF. Since $K^*\langle\gamma\rangle \leqslant K^*\langle\gamma_1, \dots, \gamma_s\rangle$ and $K(\gamma_1, \dots, \gamma_s)/K$ is a $K^*\langle\gamma_1, \dots, \gamma_s\rangle$-Kneser extension, it follows that $K(K^*\langle\gamma\rangle)/K$ is a $K^*\langle\gamma\rangle$-Kneser extension by Proposition 3.1. But $K(K^*\langle\gamma\rangle) = K(\gamma)$, hence $[K(\gamma) : K] = |K^*\langle\gamma\rangle/K^*| = \operatorname{ord}(\widehat{\gamma}) = \operatorname{ord}(\mathcal{C}) = m$.

Let I be any integral ideal of the ideal class $\mathcal{C}$. Then $I = I_1^{r_1} \cdot \cdots \cdot I_s^{r_s}$ with $0 \leqslant r_k < h_k$, $k = 1, \dots, s$. We may assume that $\gamma = \gamma_1^{r_1} \cdot \cdots \cdot \gamma_s^{r_s}$. Since $I_k^{h_k} = (\gamma_k)^{h_k} = (c_k)$, it follows that $(I_k\mathfrak{O}_{H_K})^{h_k} = (\gamma_k\mathfrak{O}_{H_K})^{h_k}$, hence $I_k\mathfrak{O}_{H_K} = \gamma_k\mathfrak{O}_{H_K}$ for all $k = 1, \dots, s$. Thus, $I\mathfrak{O}_{H_K} = (\gamma_1^{r_1} \cdot \cdots \cdot \gamma_s^{r_s})\mathfrak{O}_{H_K} = \gamma\mathfrak{O}_{H_K}$, so $I = I\mathfrak{O}_{H_K} \cap \mathfrak{O}_K = \gamma\mathfrak{O}_{H_K} \cap \mathfrak{O}_K = \gamma\mathfrak{O}_K$. This implies that $I\mathfrak{O}_{K(\gamma)} = \gamma\mathfrak{O}_{K(\gamma)}$, and we are done. □

The Hecke field H_K of a number field K has two of the basic properties of the *Hilbert class field* HCF_K of K, namely: (i) $[H_K : K] = h$; and (ii) every ideal of K becomes a principal ideal in H_K. However, the fields H_K and HCF_K are different since the extension HCF_K/K is always Galois, while, in general, the extension H_K/K is not necessarily Galois. Corollary 6.31 shows that every ideal class $\mathcal{C} \in \mathcal{C}\ell_K$ of order m becomes a principal ideal class in a suitable intermediate field of H_K/K of degree m over K. As is known, the question of whether HCF_K has the same property, was answered in the negative by Artin and Furtwängler (see Hasse [50, pp. 173–174]).

PROPOSITION 6.32. ([18].) *Let $I \in \mathcal{F}_K$, let $\mathcal{C}$ the class of I in $\mathcal{C}\ell_K$, let $m = \operatorname{ord}(\mathcal{C})$ in $\mathcal{C}\ell_K$, and let $c \in K^*$ with $I^m = (c)$. Then, the polynomial $X^m - c$ is irreducible in $K[X]$.*

PROOF. Let γ denote a root in $\mathbb{C}$ of the polynomial $X^m - c$. Then $(I\mathfrak{O}_{K(\gamma)})^m = c\mathfrak{O}_{K(\gamma)} = (\gamma\mathfrak{O}_{K(\gamma)})^m$, hence $I\mathfrak{O}_{K(\gamma)} = \gamma\mathfrak{O}_{K(\gamma)}$. Let ε be a unit of $\mathfrak{O}_{K(\gamma)}$ with $\varepsilon \in K^*\langle\gamma\rangle$. Then $\varepsilon = a\gamma^r$ for some $a \in K^*$ and $0 \leqslant r < m$. So $I^r = (I\mathfrak{O}_{K(\gamma)})^r \cap K = (\gamma)^r\mathfrak{O}_{K(\gamma)} \cap K = (\varepsilon a^{-1})\mathfrak{O}_{K(\gamma)} \cap K = (a^{-1})\mathfrak{O}_{K(\gamma)} \cap K = a^{-1}\mathfrak{O}_K = (a^{-1})$. Thus, I^r is a principal fractional ideal, hence its ideal class $\mathcal{C}^r$ is the identity class. Since $\operatorname{ord}(\mathcal{C}) = m$, it follows that $m \mid r$, hence necessarily $r = 0$, and so, $\varepsilon = a \in K^*$. Now, proceed as in the proof of Theorem 6.30 to deduce that $K(\gamma)/K$ is a $K^*\langle\gamma\rangle$-Kneser extension. Therefore

$$\big[K(\gamma) : K\big] = \big|K^*\langle\gamma\rangle/K^*\big| = \operatorname{ord}(\widehat{\gamma}).$$

If $n = \operatorname{ord}(\widehat{\gamma})$, then $n \mid m$, hence $m = nt$ for some $t \in \mathbb{N}^*$. On the other hand, we have $(c) = (\gamma)^m = (\gamma^n)^t = I^m = (I^n)^t$, which implies that $I^n = (\gamma^n)$, hence I^n is a principal fractional ideal. As above, we deduce that $m \mid n$. Then $m = n$, and so $[K(\gamma) : K] = m$. We conclude that the polynomial $X^m - c$ is irreducible in $K[X]$. □

Related to Hecke systems of ideal numbers a natural question arose: are the polynomials $X^{h_k} - c_k$ irreducible in $K[X]$, where $c_k = \gamma_k^{h_k} \in K$? This problem was only mentioned (but not settled) by Hasse [49, p. 544] as follows: “Auf die Frage nach der Irreduzibilität der Polynome $X^{h_k} - c_k$ über K wollen wir hier nicht eingehen”. The positive answer to this question, due to Albu and Nicolae [18] and presented below, immediately follows from Proposition 6.32.

COROLLARY 6.33. ([18].) *With the notation above, the polynomials $X^{h_k} - c_k$ are irreducible in $K[X]$ for every k, $1 \leqslant k \leqslant s$, the fields $K(\gamma_1), \dots, K(\gamma_s)$ are linearly disjoint over K, and there exists a canonical group isomorphism*

$$K^*\langle\gamma_1, \dots, \gamma_s\rangle/K^* \cong \prod_{1 \leqslant i \leqslant s} \big(K^*\langle\gamma_i\rangle/K^*\big).$$

PROPOSITION 6.34. ([18].) *Let K be an algebraic number field, let $S_K = K^*\langle\gamma_1, \dots, \gamma_s\rangle$ be any Hecke system of ideal numbers of K, and let $H_K = K^*(\gamma_1, \dots, \gamma_s)$ be its associated Hecke field. Then H_K/K is an S_K-Kneser extension.*

PROOF. The result follows at once from the proof of Theorem 6.30. □

EXAMPLE. In general, the S_K-Kneser extension H_K/K is not necessarily S_K-co-Galois. To see that, consider the quadratic field $K = \mathbb{Q}(\sqrt{-87})$. Then $\mathfrak{O}_K = \mathbb{Z}[(1+\sqrt{-87})/2]$ and the class number h of K is 6. The decomposition of 3 as a product of prime ideals in the ring $\mathfrak{O}_K$ is $3\mathfrak{O}_K = (3, \sqrt{-87})^2$, and the prime ideal $I = (3, \sqrt{-87})$ is not a principal ideal of $\mathfrak{O}_K$. Thus, the ideal class $\mathcal{C}_1$ of I in $\mathcal{C}\ell_K$ has order 2, and $I^2 = (3) = (-3)$. We choose $\gamma_1 = \sqrt{-3}$ as an ideal number of I. Since the group $\mathcal{C}\ell_K$ is cyclic of order 6, it contains an ideal class $\mathcal{C}_2$ of order 3. Then, we have necessarily $\mathcal{C}\ell_K = \langle \mathcal{C}_1 \rangle \oplus \langle \mathcal{C}_2 \rangle$. Let γ_2 be an ideal number of an integral ideal I_2 from the ideal class $\mathcal{C}_2$, that is, γ_2 is one of the complex roots of the polynomial $X^3 - c_2 \in K[X]$, where $c_2 \in K^*$ is such that $I_2^3 = (c_2)$.

Thus, the group $K^*\langle \sqrt{-3}, \gamma_2 \rangle$ is a Hecke system of ideal numbers of K, and $H_K = K(\sqrt{-3}, \gamma_2)$ is the associated Hecke field of K. By Proposition 6.34, the extension H_K/K is a $K^*\langle \sqrt{-3}, \gamma_2 \rangle$-Kneser extension. Observe that $\exp(K^*\langle \sqrt{-3}, \gamma_2 \rangle / K^*) = 6$, $3 \mid 6$ and $\zeta_3 = (-1+\sqrt{-3})/2 \in H_K \setminus K$, so the extension H_K/K is not 6-pure. By the n-purity criterion (Corollary 4.5), it follows that H_K/K is not a $K^*\langle \sqrt{-3}, \gamma_2 \rangle$-co-Galois extension.

The next result provides two cases when the extension H_K/K is S_K-co-Galois.

PROPOSITION 6.35. ([18].) *Let K be an algebraic number field with class number h.*

(1) *If $\zeta_h \in K$, then the extension $K^*(\gamma_1, \dots, \gamma_s)/K$ is $K^*\langle \gamma_1, \dots, \gamma_s \rangle$-co-Galois for any choice of the Hecke system $K^*\langle \gamma_1, \dots, \gamma_s \rangle$ of ideal numbers of K.*

(2) *If K can be embedded into the field $\mathbb{R}$, then there exists a Hecke system*

$$K^*\langle \gamma_1, \dots, \gamma_s \rangle$$

of ideal numbers of K such that the extension $K^*(\gamma_1, \dots, \gamma_s)/K$ is $K^*\langle \gamma_1, \dots, \gamma_s \rangle$-co-Galois.

PROOF. (1) If $\zeta_h \in K$, then clearly $\mu_h(K^*(\gamma_1, \dots, \gamma_s)) \subseteq K$. Since $h = h_1 \cdot \dots \cdot h_s$ and $\gamma_k^{h_k} = c_k \in K$ for all k, $1 \leqslant k \leqslant s$, we deduce that $K^*(\gamma_1, \dots, \gamma_s)/K$ is a generalized h-Kummer extension. Now apply Theorem 5.25 to obtain the desired result.

(2) Without loss of generality, we may assume that K is a subfield of $\mathbb{R}$. In the construction of ideal numbers we may choose $c_1 > 0, \dots, c_s > 0$. For γ_k we choose the positive real root of the polynomial $X^{h_k} - c_k$, $1 \leqslant k \leqslant s$. Then, $H_K = K(\gamma_1, \dots, \gamma_s)$ is a subfield of $\mathbb{R}$. Thus, H_K/K is a pure extension, and, a fortiori, an n-pure extension, where $n = \exp(K^*\langle \gamma_1, \dots, \gamma_s \rangle / K^*)$. By the n-purity criterion, we deduce that H_K/K is a $K^*\langle \gamma_1, \dots, \gamma_s \rangle$-co-Galois extension. □

Let A be a Dedekind ring with a finite ideal class group $\mathcal{C}\ell_A$ of order h, and denote by L its quotient field. Assume that the characteristic of L is not 2 and is relatively prime with h. Then, we can perform mutatis-mutandis the construction presented at the beginning of this subsection to define a *Hecke system* $L^*\langle \gamma_1, \dots, \gamma_s \rangle$ *of ideal elements* of L and a *Hecke field* H_L of L. One can show that the main part of the results of this subsection can be extended from algebraic number fields K to such more general fields L.

7. Connections with graded algebras and Hopf algebras

7.1. *Kneser and G-co-Galois extensions via strongly graded algebras*

In this subsection we describe the concepts of G-radical, G-Kneser, and G-co-Galois extensions in terms of graded ring theory.

Throughout this subsection all algebras are assumed to be associative with unit, and K will denote a fixed commutative ring with nonzero identity element. If A is a K-algebra and X, Y are subsets of A, then XY will denote the K-submodule of the underlying K-module of the algebra A which is generated by the set $\{xy \mid x \in X, y \in Y\}$. For a K-module M and a family $(M_i)_{i\in I}$ of submodules of M, the notation $M = \bigoplus_{i\in I} M_i$ will mean throughout this chapter that M is the *internal direct sum* of the independent family $(M_i)_{i\in I}$ of its submodules, that is, any element $x \in M$ can be uniquely expressed as $x = \sum_{i\in I} x_i$, where $(x_i)_{i\in I}$ is a family of finite support, with $x_i \in M_i$ for every $i \in I$.

Let Γ be a multiplicative group with identity element e. Recall that a K-algebra A is said to be a *Γ-graded algebra* if $A = \bigoplus_{\gamma\in\Gamma} A_\gamma$ is a direct sum of K-submodules A_γ of A, with $A_\gamma A_\delta \subseteq A_{\gamma\delta}$ for every $\gamma, \delta \in \Gamma$. A Γ-graded algebra $A = \bigoplus_{\gamma\in\Gamma} A_\gamma$ is said to be *strongly graded* if $A_\gamma A_\delta = A_{\gamma\delta}$ for every $\gamma, \delta \in \Gamma$. A (*strongly*) *$\Gamma$-graded ring* is a (strongly) Γ-graded algebra over the ring $\mathbb{Z}$ of rational integers. A left module M over the Γ-graded algebra $A = \bigoplus_{\gamma\in\Gamma} A_\gamma$ is said to be a *graded module* (respectively a *strongly graded module*) if $M = \bigoplus_{\gamma\in\Gamma} M_\gamma$ is a direct sum of K-submodules M_γ of M, with $A_\gamma M_\delta \subseteq M_{\gamma\delta}$ (respectively $A_\gamma M_\delta = M_{\gamma\delta}$) for every $\gamma, \delta \in \Gamma$. The elements of $h(M) = \bigcup_{\gamma\in\Gamma} M_\gamma$ are called *homogeneous elements* of M. Any element $x \in M$ has a unique decomposition $x = \sum_{\gamma\in\Gamma} x_\gamma$, with $x_\gamma \in M_\gamma$, $\gamma \in \Gamma$, where all but a finite number of the x_γ are zero; the elements x_γ are called the *homogeneous components* of x.

DEFINITION. Let E/F be a field extension, and let Γ be a multiplicative group with identity element e. One says that E/F is a *Γ-Clifford extension* (respectively a *strongly Γ-graded extension*) if there exists a family $(E_\gamma)_{\gamma\in\Gamma}$ of F-subspaces of the vector space ${}_FE$ indexed by the group Γ, satisfying the following conditions.

(1) $E = \sum_{\gamma\in\Gamma} E_\gamma$ (respectively $E = \bigoplus_{\gamma\in\Gamma} E_\gamma$).
(2) $E_\gamma E_\delta = E_{\gamma\delta}$ for every $\gamma, \delta \in \Gamma$.
(3) $E_\gamma = F \Leftrightarrow \gamma = e$.

An element $x \in E$ is said to be *homogeneous* of degree γ if $x \in E_\gamma$. The set of all nonzero homogeneous elements of E will be denoted by $U^h(E)$.

LEMMA 7.1. ([74,12].) *The following statements hold for a Γ-Clifford extension E/F.*

(1) *If $\gamma, \delta \in \Gamma$, then $E_\gamma = E_\delta \Leftrightarrow \gamma = \delta$.*
(2) $\dim_F(E_\gamma) = 1$ *for all $\gamma \in \Gamma$. In particular, $E_\gamma = Fx$ for every $x \in E^*_\gamma$.*
(3) *For every $\gamma \in \Gamma$ let $x_\gamma \in E^*_\gamma$ be arbitrary. Then $X = \{x_\gamma \mid \gamma \in \Gamma\}$ is a set of generators of the vector space ${}_FE$, and $[E : F] \leqslant |\Gamma|$. If E/F is a strongly Γ-graded extension, then X is a basis of the vector space ${}_FE$, and $[E : F] = |\Gamma|$. Conversely, if there exists a family $(x_\gamma)_{\gamma\in\Gamma} \in \prod_{\gamma\in\Gamma} E^*_\gamma$ which is a basis of the vector space ${}_FE$, then E/F is a strongly Γ-graded extension.*

(4) *If E/F is a finite extension, then E/F is a strongly Γ-graded extension if and only if $[E : F] = |\Gamma|$.*
(5) *The group Γ is Abelian.*
(6) *If Γ is a torsion group, then E/F is an algebraic extension.*

EXAMPLE. The next example shows that, in general, the converse implication in Lemma 7.1(6) does not hold. Let $\theta = \sqrt{1+\sqrt{2}}$, $F = \mathbb{Q}$ and $E = \mathbb{Q}(\theta)$. Then E/F is a quartic extension, hence an algebraic extension, which is also a $\mathbb{Z}$-Clifford extension with $E = \sum_{n\in\mathbb{Z}} E_n$, $E_n = F\theta^n, n \in \mathbb{Z}$. Indeed, the only nontrivial fact is that $E_n = F \Rightarrow n = 0$. But $E_n = F \Leftrightarrow \theta^n \in \mathbb{Q}$, which can happen only if $n = 0$ by the example following Proposition 3.5 in Subsection 3.2. Thus, E/F is a $\mathbb{Z}$-Clifford extension, but $\mathbb{Z}$ is a torsion-free group.

For any Γ-Clifford extension E/F and any $\Delta \leqslant \Gamma$ we shall write $E_\Delta := \sum_{\gamma\in\Delta} E_\gamma$. Obviously, E_Δ/F is a Δ-Clifford extension, which is strongly Δ-graded whenever the extension E/F is strongly Γ-graded.

PROPOSITION 7.2. ([12].) *The following assertions hold for a field extension E/F.*
(1) *Let G be a group with $F^* \leqslant G \leqslant E^*$. If E/F is a G-radical extension, then E/F is a G/F^*-Clifford extension.*
(2) *Conversely, if E/F is a Γ-Clifford extension for some torsion group Γ, then there exists a group G such that $F^* \leqslant G \leqslant E^*$, $\Gamma \cong G/F^*$, E/F is G-radical, and $U^h(E) = G$.*
(3) *If E/F is a G/F^*-Clifford extension, where $F^* \leqslant G \leqslant T(E/F)$ and $U^h(E) \subseteq G$, then E/F is G-radical.*
(4) *Let G be a group with $F^* \leqslant G \leqslant E^*$. If E/F is a G-Kneser extension, then E/F is a strongly G/F^*-graded extension.*
(5) *Conversely, if E/F is a strongly Γ-graded extension for some torsion group Γ, then there exists a group G such that $F^* \leqslant G \leqslant E^*$, $\Gamma \cong G/F^*$, E/F is G-Kneser, and $U^h(E) = G$.*
(6) *If E/F is a strongly G/F^*-graded extension, with $F^* \leqslant G \leqslant T(E/F)$ and $U^h(E) \subseteq G$, then E/F is G-Kneser.*

The graded version of the concept of G-co-Galois extension is that of Γ-Clifford-co-Galois extension we are going to introduce below. For any Γ-Clifford extension E/F with $E = \sum_{\gamma\in\Gamma} E_\gamma$, and any intermediate field K of E/F we shall use the following notation: $\Gamma_K = \{\gamma \in \Gamma \mid E_\gamma \subseteq K\}$. Then clearly Γ_K is a subgroup of Γ, hence it makes sense to consider the order-preserving map

$$\Phi : \mathbb{I}(E/F) \to \mathbb{L}(\Gamma), \quad \Phi(K) = \Gamma_K.$$

We have also another order-preserving map

$$\Psi : \mathbb{L}(\Gamma) \to \mathbb{I}(E/F), \quad \Psi(\Delta) = E_\Delta = \sum_{\gamma\in\Delta} E_\gamma,$$

and the maps Φ and Φ define a co-Galois connection between the lattices $\mathbb{I}(E/F)$ and $\mathbb{L}(\Gamma)$.

PROPOSITION 7.3. ([12].) *The following assertions are equivalent for a Γ-Clifford extension E/F with Γ a torsion group.*

(1) *Every subextension K/F of E/F is a strongly Γ_K-graded extension.*
(2) *E/F is a strongly Γ-graded extension, and the map $\Phi : \mathbb{I}(E/F) \to \mathbb{L}(\Gamma)$, $\Phi(K) = \Gamma_K$, is a lattice isomorphism.*
(3) *E/F is a strongly Γ-graded extension, and the maps*

$$\Phi : \mathbb{I}(E/F) \to \mathbb{L}(\Gamma), \Phi(K) = \Gamma_K,$$
$$\Psi : \mathbb{L}(\Gamma) \to \mathbb{I}(E/F), \Psi(\Delta) = E_\Delta,$$

are isomorphisms of lattices, inverse to one another.
(4) *E/F is a strongly Γ-graded extension, and every intermediate field K of E/F has a vector space basis over F consisting of homogeneous elements.*
(5) *E/F is a strongly Γ-graded extension, and every intermediate field K of E/F is obtained by adjoining to F a set of homogeneous elements of E.*
(6) *E/F is a strongly Γ-graded extension, and every nonzero element $x \in E$ has its homogeneous components in $F(x)$.*
(7) *E/K is a strongly Γ/Γ_K-graded extension for every intermediate field K of E/F.*

DEFINITION. A field extension E/F is said to be *Γ-Clifford-co-Galois* if it a separable Γ-Clifford extension which satisfies one of the equivalent conditions of Proposition 7.3.

PROPOSITION 7.4. ([12].) *The following statements hold for a field extension E/F.*

(1) *If the extension E/F is G-co-Galois for some G with $F^* \leqslant G \leqslant E^*$, then E/F is a G/F^*-Clifford-co-Galois extension.*
(2) *Conversely, if E/F is a Γ-Clifford-co-Galois extension for some group Γ, then there exists a uniquely determined group G such that $F^* \leqslant G \leqslant E^*$, $\Gamma \cong G/F^*$, E/F is G-co-Galois, and $U^h(E) = G$.*

The first of the next two results is the graded version of Theorem 4.13 on the uniqueness of the Kneser group of a G-co-Galois field extension, while the second one is the graded version of Theorem 4.14 concerning primitive elements of such field extensions.

COROLLARY 7.5. ([12].) *Let E/F be a field extension which is simultaneously Γ-Clifford-co-Galois and Δ-Clifford-co-Galois. Then the groups Γ and Δ are isomorphic.*

We present below a series of results due to Ştefan [74] about Γ-Clifford extensions with Γ a finite group, other than the ones that were extended by Albu [12] to arbitrary Γ-Clifford extensions and presented above.

PROPOSITION 7.6. ([74].) *Let E/F be a finite Γ-Clifford-co-Galois extension, let Λ be a finite nonempty subset of Γ, and let $\{x_\lambda \mid \lambda \in \Lambda\} \subseteq U^h(E)$. Then $\sum_{\lambda\in\Lambda} x_\lambda$ is a primitive element of the extension E/F if and only if $\langle\Lambda\rangle = \Gamma$.*

PROPOSITION 7.7 (Graded version of the Kneser criterion [74]). *The following assertions are equivalent for a separable Γ-Clifford extension E/F with Γ a finite group of exponent n.*

(a) *E/F is a strongly Γ-graded extension.*

(b) *For every $p \in \mathcal{P}_n$ and for every cyclic subgroup Δ of order p of Γ one has $[E_\Delta : F] = p$.*

(c) *For every odd prime p, $\zeta_p \in U^h(E) \Rightarrow \zeta_p \in F$, and $1 + \zeta_4 \in U^h(E) \Rightarrow \zeta_4 \in F$.*

(d) *$\mu_p(E) \cap U^h(E) \subseteq F$ and $(1 + \mu_4(E)) \cap U^h(E) \subseteq F$.*

PROPOSITION 7.8 (Graded version of the finite n-purity criterion [74]). *A separable Γ-Clifford extension E/F with Γ a finite group of exponent n is a Γ-Clifford-co-Galois extension if and only if $\mu_p(E) = \mu_p(F)$ for all $p \in \mathcal{P}_n$.*

PROPOSITION 7.9. ([74].) *Let E/F be a Galois Γ-Clifford-co-Galois extension with Γ a finite group of exponent n. If $n \in \mathbb{P}$ or $\mu_n(E) = \mu_n(F)$, then* $\mathrm{Gal}(E/F) \cong \Gamma$.

We strongly believe that infinite graded versions of the Kneser criterion and of the general purity criterion for field extensions also hold.

We end this subsection by presenting a similar group-graded/coring approach in investigating co-Galois extensions E/F, finite or not, due to Masuoka [59]. Masuoka's terminology is slightly different from ours; for instance, if E/F is a field extension, finite or not, and Γ is a torsion subgroup of the quotient group E^*/F^*, then his notion of *strongly Γ-graded extension* is our notion of a Γ-Clifford-co-Galois extension. Then, the concept of coring is used to interpret the quotient group E^*/F^* for an arbitrary field extension E/F as a group of group-like elements of the coring $E \otimes_F E$ and to define the concepts of (possibly infinite) coseparable and co-Galois field extension as follows.

Recall that if R is a ring with identity element 1, an *R-coring* is an (R, R)-bimodule C together with (R, R)-bimodule maps $\Delta_C : C \to C \otimes_R C$ called a coproduct and $\varepsilon_C : C \to R$ called a counit, satisfying the usual coassociativity and left and right counit conditions for Hopf algebras (see Brzeziński and Wisbauer [32]). A *group-like* element in C is any element $g \in C$ satisfying $\Delta_C(g) = g \otimes g$ and $\varepsilon_C(g) = 1$. We denote by $\mathrm{Gr}(C)$ the set of all group-like elements of C.

If E/F is an arbitrary field extension, then $E \otimes_F E$ has a natural structure of (E, E)-bimodule and a E-coring structure defined by the following maps:

$$\begin{aligned} &\Delta : E \otimes_F E \to (E \otimes_F E) \otimes_E (E \otimes_F E) = E \otimes_F E \otimes_F E, \\ &\quad \Delta(x \otimes y) = x \otimes 1 \otimes y, \\ &\varepsilon : E \otimes_F E \to E, \quad \varepsilon(x \otimes y) = xy. \end{aligned}$$

The F-algebra structure on $E \otimes_F E$ makes $\mathrm{Gr}(E \otimes_F E)$ a group. A nice interpretation of this group, due to Masuoka [59], is the following: the map

$$E^*/F^* \to \mathrm{Gr}(E \otimes_F E), \quad xF^* \mapsto x^{-1} \otimes x,$$

is a group isomorphism.

For any subgroup Γ of $\mathrm{Gr}(E \otimes_F E)$, a canonical E-algebra morphism

$$\Theta_\Gamma : E[\Gamma] \to E \otimes_F E$$

is induced by the inclusion map of Γ into the group of units of $E \otimes_F E$, where $E \otimes_F E$ is viewed as an E-algebra via the map $E \to E \otimes_F E, x \mapsto x \otimes 1$. Masuoka [59] calls the field extension E/F *coseparable* (respectively *co-Galois*) if the map $\Theta_{\mathrm{Cog}(E/F)} : E[\mathrm{Cog}(E/F)] \to E \otimes_F E$ is surjective (respectively bijective). It turns out that the co-Galois (respectively coseparable) extensions in Masuoka's sense are precisely the co-Galois (respectively radical) extensions in the usual sense. Proofs in this coring setting of Theorems 3.2, 3.3, Proposition 3.5, and Corollary 4.4 are also given in [59].

Now, if E/F is a Galois extension with $\Delta = \mathrm{Gal}(E/F)$ then the canonical E-algebra isomorphism $E \otimes_F E \overset{\sim}{\to} \mathrm{Map}_c(\Delta, E), x \otimes y \mapsto (\sigma \mapsto x\sigma(y))$, where $\mathrm{Map}_c(\Delta, E)$ is the E-algebra of all continuous maps from the profinite group Δ into the discrete space E, induces a group isomorphism $E^*/F^* \cong \mathrm{Gr}(E \otimes_F E) \cong Z^1_c(\Delta, E^*)$, and hence, by taking the torsion part we retrieve the group isomorphism $\mathrm{Cog}(E/F) \cong Z^1_c(\Delta, \mu(E))$ from Theorem 5.5.

7.2. *Kneser and co-Galois extensions via Hopf algebras*

In this subsection we show that a G-Kneser extension E/F is nothing else than a field extension E/F such that E is a Galois H-object, where H is the group algebra $F[G/F^*]$. Since an extension E/F is co-Galois if and only if it is $T(E/F)$-Kneser, one obtains a similar description for co-Galois extensions.

Throughout this subsection K will denote a fixed field, and H will denote a K-Hopf algebra. Tensor products are assumed to be over K, unless stated otherwise. We shall denote by K-Mod the category of all K-vector spaces. For the standard concepts and facts of Hopf algebras presented in this subsection the reader is referred to Sweedler [77], Montgomery [61], Caenepeel [33], and/or Dăscălescu, Năstăsescu, and Raianu [38]. In particular, if (C, Δ, ε) is a K-coalgebra, then for any right C-comodule M with structure map $\rho : M \to M \otimes C$ and any $m \in M$, we shall also use the *sigma notation*: $\rho(m) = \sum m_0 \otimes m_1$. If A is a right H-comodule algebra then by $\mathcal{M}^H_A$ we will denote the category of all right (H, A)-Hopf modules.

DEFINITION. A *Galois H-object* is a right H-comodule algebra A such that the map

$$\gamma : A \otimes A \to A \otimes H, \quad \gamma(a \otimes b) = \sum ab_0 \otimes b_1, a, b \in A,$$

is bijective.

THEOREM 7.10. ([33].) *The following statements are equivalent for a K-Hopf algebra H with bijective antipode and for a right H-comodule algebra A.*

(1) *A is a Galois H-object.*

(2) *The functors F and G defined by*

$$F : K\text{-Mod} \to \mathcal{M}^H_A, \quad F(X) = A \otimes X, X \in K\text{-Mod},$$

$$G : \mathcal{M}_A^H \to K\text{-Mod}, \quad G(M) = M^{\mathrm{co}H}, M \in \mathcal{M}_A^H,$$

are inverse equivalences, or equivalently, for all $X \in K$-Mod *and all* $M \in \mathcal{M}_A^H$, *the natural maps*

$$X \to G\bigl(F(X)\bigr) = (A \otimes X)^{\mathrm{co}H}, \quad x \mapsto 1 \otimes x,$$
$$F\bigl(G(M)\bigr) = A \otimes M^{\mathrm{co}H} \to M, \quad a \otimes m \mapsto m \cdot a,$$

are isomorphisms.

Moreover, if (1) *or* (2) *holds, then* $A^{\mathrm{co}H} = K$.

COROLLARY 7.11. *Let A be a Galois H-object. Then, for any* $M \in \mathcal{M}_A^H$, *the map*

$$\underline{\mathit{Subobjects}}(M) \to \underline{\mathit{Subspaces}}\bigl(M^{\mathrm{co}H}\bigr), \quad N \mapsto N^{\mathrm{co}H},$$

is a lattice isomorphism.

LEMMA 7.12. *Let G be a group, and let M be a K-vector space. Assume that there exists a family $(M_g)_{g \in G}$ of K-subspaces of M such that $M = \sum_{g \in G} M_g$. Then,*

$$\rho : M \to M \otimes_K K[G], \quad x \mapsto x \otimes g,$$

for all $x \in M_g$ is a well-defined map if and only if $M = \bigoplus_{g \in G} M_g$, and in this case M is a right $K[G]$-comodule via the linear map ρ. Conversely, if the K-vector space M is a right $K[G]$-comodule, then M is a G-graded K-vector space, that is, $M = \bigoplus_{g \in G} N_g$ is a direct sum of K-subspaces N_g, $g \in G$.

PROOF. Assume that ρ is a well-defined map, and let $g \in G$. If $m \in M_g \cap \sum_{h \in G \setminus \{g\}} M_h$, then $m = \sum_{h \in G \setminus \{g\}} m_h$ with $m_h \in M_h$. Apply ρ to both sides of this equality, and then apply $1_M \otimes p_g$ to both sides to obtain $m \otimes 1 = 0$, where $p_g : K[G] \to K$ denotes the linear map defined by $p_g(\sum_{h \in G} k_h h) = k_g$. This implies that $m = 0$, hence $M = \bigoplus_{g \in G} M_g$. Conversely, if $M = \bigoplus_{g \in G} M_g$, then clearly ρ is a well-defined linear map. It can be easily checked that the map ρ endows M with a structure of right $K[G]$-comodule, and any right $K[G]$-comodule is a G-graded K-vector space. □

PROPOSITION 7.13. *The following assertions hold for an arbitrary group G and a K-vector space A.*

(1) *A is a right $K[G]$-comodule algebra if and only if A is a G-graded K-algebra.*

(2) *A is a Galois $K[G]$-object if and only if A is a strongly G-graded K-algebra and $A_e = K$.*

COROLLARY 7.14. ([12].) *Let E/F be a G-radical field extension. Then, the assignment $E \to E \otimes F[G/F^*]$, $x \mapsto x \otimes \widehat{g}$, for all $x \in Fg$ and $g \in G$ is a well-defined map which endows E with a structure of $F[G/F^*]$-comodule if and only if $E = \bigoplus_{\widehat{g} \in G/F^*} Fg$. In this case, E is a right $F[G/F^*]$-comodule algebra.*

We are now in a position to present the relationships of Kneser and co-Galois field extensions with Hopf algebras.

THEOREM 7.15. ([12].) *The following assertions are equivalent for a G-radical field extension E/F.*

(1) *E/F is a G-Kneser extension.*

(2) *E is a Galois $F[G/F^*]$-object via the comodule structure given by the map*

$$E \to E \otimes F[G/F^*], \quad x \mapsto x \otimes \widehat{g},$$

for all $x \in Fg$ and $g \in G$.

PROOF. (1) ⇒ (2): By Proposition 7.2(4), E/F is a strongly G/F^*-graded extension via the decomposition $E = \bigoplus_{\widehat{g} \in G/F^*} E_{\widehat{g}}$, with $E_{\widehat{g}} = Fg$ for every $g \in G$. Since $E_{\widehat{1}} = F1 = F$, we deduce by Proposition 7.13(2) that E is a Galois $F[G/F^*]$-object.

(2) ⇒ (1): Again, by Proposition 7.13(2), and using also Corollary 7.14, we deduce that the extension E/F is strongly G/F^*-graded. Now, apply Proposition 7.2(6) to deduce that E/F is G-Kneser. □

COROLLARY 7.16. ([12].) *The following assertions are equivalent for a field extension E/F.*

(1) *E/F is a co-Galois extension.*

(2) *E is a Galois $F[\mathrm{Cog}(E/F)]$-object with respect to the comodule structure given by the linear map $E \to E \otimes F[\mathrm{Cog}(E/F)]$, $x \mapsto x \otimes \widehat{g}$, for all $x \in Fg$ and $g \in T(E/F)$.*

PROOF. The result follows at once from Theorem 7.15, since $\mathrm{Cog}(E/F) = T(E/F)/F^*$, and by definition, an extension E/F is co-Galois if and only if it is $T(E/F)$-Kneser. □

Next, we present the concept of an H-Galois extension, as well as its connection with that of Galois H-object.

DEFINITION. If A is a right H-comodule algebra, then we say that $A^{\mathrm{co}H} \subset A$ is an *H-extension*. The H-extension $A^{\mathrm{co}H} \subset A$ is said to be an *H-Galois extension* if the map $\beta : A \otimes_{A^{\mathrm{co}A}} H \to A \otimes_K H$, $a \otimes b \mapsto ab_0 \otimes b_1$, is bijective.

PROPOSITION 7.17. ([33].) *Let H be a K-Hopf algebra, and suppose that the antipode of H is bijective. Then, the following statements hold for a right H-comodule algebra A.*

(1) *If A is a Galois H-object, then $K = A^{\mathrm{co}H} \subset A$ is an H-Galois extension.*

(2) *If $A^{\mathrm{co}H} \subset A$ is an H-Galois extension and $A^{\mathrm{co}H} = K$, then A is a Galois H-object.*

We end this subsection by examining the connection between classical Galois field extensions and H-Galois extensions.

LEMMA 7.18. ([38].) *Let H be any finite-dimensional K-Hopf algebra, and let A be any K-algebra. Then A is a right H-comodule algebra if and only if A is a left H^*-module algebra. Moreover, in this case we have $A^{H^*} = A^{\mathrm{co}H}$.*

In particular, if E/F is any field extension, and $G \leqslant \mathrm{Gal}(E/F)$ is any finite group of F-automorphisms of E, then it is easily verified that E is a left $F[G]$-module algebra. Consequently, by Lemma 7.18, E is a right $F[G]^*$-comodule algebra and $E^{\mathrm{co}F[G]^*} = E^{F[G]} = \mathrm{Fix}(G)$.

PROPOSITION 7.19. ([12].) *The following assertions are equivalent for a field extension E/F and a finite group G with $G \leqslant \mathrm{Gal}(E/F)$.*

(1) *E/F is a Galois field extension with $G = \mathrm{Gal}(E/F)$.*

(2) *E is a Galois $F[G]^*$-object.*

(3) *$F \subset E$ is an $F[G]^*$-Galois extension and $F = \mathrm{Fix}(G)$.*

Note that the concept of *Galois H-coobject* is defined and investigated in the theory of Hopf algebras and is the formal dual of the Galois H-object (see, e.g., Caenepeel [33, Section 8.7]). It is not clear how this concept is related to that of co-Galois field extension. Also we do not know how it can be expressed via Hopf algebras that field extension is G-co-Galois. Finally, let us mention that a Hopf–Galois correspondence for division F-algebras via corings and Hopf algebras, generalizing the equivalence (3) ⇔ (4) in Proposition 7.3 expressed in a slightly modified terminology, can be found in Masuoka [59].

8. Abstract co-Galois theory

8.1. *Notation and preliminaries*

Throughout this last part of the chapter Γ will denote a fixed profinite group with identity element denoted by 1, and A will always be a fixed subgroup of the Abelian group $\mathbb{Q}/\mathbb{Z}$ such that Γ acts continuously on A endowed with the discrete topology, i.e., A is a discrete Γ-module. For any $r \in \mathbb{Q}$, the coset of r in the quotient group $\mathbb{Q}/\mathbb{Z}$ will be denoted by $\widehat{r}$. The action of $\sigma \in \Gamma$ on $a \in A$ will be denoted by σa. The set of all elements of A invariant under the action of Γ will be denoted by A^{Γ}.

If $n \in \mathbb{N}^*$ and D is an Abelian torsion group, then we shall use the notation $D[n] := \{x \in D \mid nx = 0\}$. For any $p \in \mathbb{P}$ we denote by $D(p)$ the p-primary component of D. Recall that $\mathcal{O}_D$ is the set of all $n \in \mathbb{N}^*$ for which there exists an $x \in D$ of order n, i.e., $D[n]$ has exponent n.

For any topological group T and any subgroup U of T we denote by $\mathbb{L}(T|U)$ (respectively $\overline{\mathbb{L}}(T|U)$) the lattice of all subgroups (respectively closed subgroups) of T lying over U. If $X \subseteq T$, then $\overline{X}$ will denote the closure of X. For a subgroup U of T we shall denote by T/U the set $\{tU \mid t \in T\}$ of all left cosets of U in T.

Recall that a *crossed homomorphism* (or a 1-*cocycle*) of Γ with coefficients in A is a map $f : \Gamma \to A$ such that $f(\sigma\tau) = f(\sigma) + \sigma f(\tau), \sigma, \tau \in \Gamma$. The set of all continuous crossed homomorphisms of Γ with coefficients in A is an Abelian group, which will be denoted by $Z^1_c(\Gamma, A)$. Note that, in fact, $Z^1_c(\Gamma, A)$ is a torsion group. Indeed, since Γ is a profinite group and A is a discrete space, a map $h : \Gamma \to A$ is continuous if and only if h is locally constant, that is, there exists an open normal subgroup Δ (in particular, of finite index) in Γ such that h factorizes through the canonical surjection $\Gamma \to \Gamma/\Delta$. Since A

is a torsion group, it follows now that for any continuous map $h : \Gamma \to A$ there exists an $n \in \mathbb{N}^*$ such that $h(\Gamma) \subseteq (1/n)\mathbb{Z}/\mathbb{Z}$, and then $nh = 0$, i.e., h has finite order.

Elements of $Z^1_c(\Gamma, A)$ will be denoted by f, g, h. Always G, H will denote subgroups of $Z^1_c(\Gamma, A)$ and Δ a subgroup of Γ. For every $a \in A$ we shall denote by f_a the 1-*coboundary* $f_a : \Gamma \to A$, defined as $f_a(\sigma) = \sigma a - a, \sigma \in \Gamma$. The set $B^1(\Gamma, A) := \{f_a \mid a \in A\}$ is a subgroup of $Z^1_c(\Gamma, A)$. The quotient group $Z^1_c(\Gamma, A)/B^1(\Gamma, A)$ is called the first cohomology group of Γ with coefficients in A, and is denoted by $H^1_c(\Gamma, A)$. As in Subsection 5.2, we can consider the *evaluation map*

$$\langle\text{-},\text{-}\rangle : \Gamma \times Z^1_c(\Gamma, A) \to A, \quad \langle\sigma, h\rangle = h(\sigma).$$

For any $\Delta \leqslant \Gamma$, $G \leqslant Z^1_c(\Gamma, A)$, $g \in Z^1_c(\Gamma, A)$, and $\gamma \in \Gamma$ we write

$$\begin{aligned}
\Delta^\perp &= \{h \in Z^1_c(\Gamma, A) \mid \langle\sigma, h\rangle = 0, \forall\sigma \in \Delta\},\\
G^\perp &= \{\sigma \in \Gamma \mid \langle\sigma, h\rangle = 0, \forall h \in G\},\\
g^\perp &= \{\sigma \in \Gamma \mid \langle\sigma, g\rangle = 0\},\\
\gamma^\perp &= \{h \in Z^1_c(\Gamma, A) \mid \langle\gamma, h\rangle = 0\}.
\end{aligned}$$

Then $\Delta^\perp \leqslant Z^1_c(\Gamma, A)$, $G^\perp \leqslant \Gamma$, and $g^\perp = \langle g\rangle^\perp$. Observe that $g^\perp$ is the set of zeroes of the continuous map g from Γ to the discrete group A, hence it is an open subgroup of Γ. Since $G^\perp = \bigcap_{g\in G} g^\perp$, it follows that $G^\perp \in \overline{\mathbb{L}}(\Gamma)$.

The group $Z^1_c(\Gamma, A)$ is clearly a discrete left Γ-module with respect to the following action: $(\sigma h)(\tau) = \sigma h(\sigma^{-1}\tau\sigma), \sigma, \tau \in \Gamma, h \in Z^1_c(\Gamma, A)$. If $\sigma \in \Gamma$ and $G \in \mathbb{L}(Z^1_c(\Gamma, A))$, then $(\sigma G)^\perp = \sigma G^\perp\sigma^{-1}$. For any $\Delta \in \mathbb{L}(\Gamma)$ we denote by

$$\mathrm{res}^\Gamma_\Delta : Z^1_c(\Gamma, A) \to Z^1_c(\Delta, A), \quad h \mapsto h|_\Delta,$$

the restriction map.

The next result lists the main properties of the assignments $(\text{-})^\perp$.

PROPOSITION 8.1. ([15].) *The following assertions hold.*

(1) *The maps*

$$\begin{aligned}
\mathbb{L}(Z^1_c(\Gamma, A)) &\to \overline{\mathbb{L}}(\Gamma), \quad G \mapsto G^\perp,\\
\overline{\mathbb{L}}(\Gamma) &\to \mathbb{L}\big(Z^1_c(\Gamma, A)\big), \quad \Delta \mapsto \Delta^\perp,
\end{aligned}$$

establish a Galois connection between the lattices $\mathbb{L}(Z^1_c(\Gamma, A))$ *and* $\overline{\mathbb{L}}(\Gamma)$, *i.e., they are order-reversing maps and* $X \leqslant X^{\perp\perp}$ *for any element* X *of* $\mathbb{L}(Z^1_c(\Gamma, A))$ *or* $\overline{\mathbb{L}}(\Gamma)$.

(2) $\Delta^\perp = \overline{\Delta}^\perp = \mathrm{Ker}(\mathrm{res}^\Gamma_\Delta)$ *and* $(\mathrm{res}^\Gamma_\Delta(G))^\perp = G^\perp \cap \Delta$ *for any* $\Delta \in \mathbb{L}(\Gamma)$, $G \in Z^1_c(\Gamma, A)$.

(3) $(G_1 + G_2)^\perp = G_1^\perp \cap G_2^\perp$ *and* $\Delta_1^\perp \cap \Delta_2^\perp = \langle\Delta_1 \cup \Delta_2\rangle^\perp$ *for any* $G_1, G_2 \in Z^1_c(\Gamma, A)$, $\Delta_1, \Delta_2 \in \mathbb{L}(\Gamma)$.

The natural continuous action of Γ on the profinite Abelian group $\widehat{Z^1_c(\Gamma, A)}$ induces a canonical continuous 1-cocycle $\eta : \Gamma \to \widehat{Z^1_c(\Gamma, A)}$ that we are going to define below, and which will play a key role in the remaining pages of this chapter. First note that

$\widehat{Z^1_c(\Gamma, A)} := \mathrm{Hom}(Z^1_c(\Gamma, A), \mathbb{Q}/\mathbb{Z}) = \mathrm{Hom}(Z^1_c(\Gamma, A), A)$. Indeed, for any $\varphi \in \widehat{Z^1_c(\Gamma, A)}$ and for any $g \in Z^1_c(\Gamma, A)$, we have $\varphi(g) \in (1/n)\mathbb{Z}/\mathbb{Z}$, where $n = \mathrm{ord}(g)$. On the other hand, as n is the lcm of the orders of $g(\sigma)$ for $\sigma \in \Gamma$, one easily deduces that $(1/n)\mathbb{Z}/\mathbb{Z} \subseteq A$, and hence $\varphi(g) \in A$, as required. The Abelian profinite group $\widehat{Z^1_c(\Gamma, A)}$ becomes a topological Γ-module via the canonical continuous action of the profinite group Γ given by $(\sigma\varphi)(g) = \sigma\varphi(g), \forall\sigma \in \Gamma, \varphi \in \widehat{Z^1_c(\Gamma, A)}, g \in Z^1_c(\Gamma, A)$. We have $\widehat{\widehat{Z^1_c(\Gamma, A)}} = \mathrm{Hom}_\Gamma(\widehat{Z^1_c(\Gamma, A)}, A)$, i.e., any continuous morphism $\chi : \widehat{Z^1_c(\Gamma, A)} \to \mathbb{Q}/\mathbb{Z}$ takes values in A and is also a morphism of Γ-modules. Indeed, the canonical morphism $\alpha : Z^1_c(\Gamma, A) \to \widehat{\widehat{Z^1_c(\Gamma, A)}}$ defined by $\alpha(g)(\varphi) = \varphi(g)$ for $g \in Z^1_c(\Gamma, A)$ and $\varphi \in \widehat{Z^1_c(\Gamma, A)} = \mathrm{Hom}(Z^1_c(\Gamma, A), A)$ is an isomorphism by Pontryagin duality, and hence for $\chi \in \widehat{\widehat{Z^1_c(\Gamma, A)}}, \sigma \in \Gamma$, and $\varphi \in \widehat{Z^1_c(\Gamma, A)}$ we have $\chi(\sigma\varphi) = (\sigma\varphi)(\alpha^{-1}(\chi)) = \sigma\varphi(\alpha^{-1}(\chi)) = \sigma\chi(\varphi)$, as required.

For any subgroup G of $Z^1_c(\Gamma, A)$ set $X_G = Z^1_c(\Gamma, A)/G$. Then observe that $\widehat{X_G} = \mathrm{Hom}(X_G, A)$ is identified with a closed subgroup of $\widehat{Z^1_c(\Gamma, A)}$, stable under the action of Γ, so the quotient $\widehat{Z^1_c(\Gamma, A)}/\widehat{X_G} \cong \widehat{G} = \mathrm{Hom}(G, A)$ is also a topological Γ-module, and $\widehat{\widehat{G}} = \mathrm{Hom}_\Gamma(\widehat{G}, A)$. Now consider the map

$$\eta : \Gamma \to \widehat{Z^1_c(\Gamma, A)}, \quad \eta(\sigma)(g) = \langle\sigma, g\rangle = g(\sigma),\ \sigma \in \Gamma, g \in Z^1_c(\Gamma, A),$$

and for any $G \leqslant Z^1_c(\Gamma, A)$, let $\eta_G : \Gamma \to \widehat{G}$ denote the map obtained from η by composing it with the canonical epimorphism of topological Γ-modules

$$\mathrm{res}_G^{Z^1_c(\Gamma,A)} : \widehat{Z^1_c(\Gamma, A)} \to \widehat{G}, \quad \varphi \mapsto \varphi|_G.$$

PROPOSITION 8.2. ([15].) *For any $G \leqslant Z^1_c(\Gamma, A)$, the map $\eta_G : \Gamma \to \widehat{G}$ defined by $\eta_G(\sigma)(g) = g(\sigma)$, $\sigma \in \Gamma, g \in G$, is a continuous* 1*-cocycle satisfying the following universality property*: *for every $g \in G$ there exists a unique continuous morphism $\chi : \widehat{G} \to A$ such that $\chi \circ \eta_G = g$.*

PROOF. One easily checks that η_G is a continuous 1-cocycle, so $\chi \circ \eta_G \in Z^1_c(\Gamma, A)$ for all $\chi \in \widehat{\widehat{G}} = \mathrm{Hom}_\Gamma(\widehat{G}, A)$. Thus, it is sufficient to show that the canonical morphism $\beta_G : \widehat{\widehat{G}} \to Z^1_c(\Gamma, A), \chi \mapsto \chi \circ \eta_G$, takes values in G and is the inverse of canonical isomorphism $\alpha_G = \alpha|_G : G \to \widehat{\widehat{G}}$ given by Pontryagin duality. The equality $\beta_G \circ \alpha_G = 1_G$ is obvious, so it remains only to check that β_G is injective. Let $\chi \in \widehat{\widehat{G}}$ be such that $\beta_G(\chi) = 0$, and let $g = \alpha_G^{-1}(\chi)$. For all $\sigma \in \Gamma$ we have

$$0 = \beta_G(\chi)(\sigma) = \chi\big(\eta_G(\sigma)\big) = \alpha_G(g)\big(\eta_G(\sigma)\big) = \eta_G(\sigma)(g) = g(\sigma),$$

so $g = 0$, and hence $\chi = \alpha_G(g) = 0$, as desired. □

8.2. *Kneser groups of cocycles*

In this subsection we define the concept of an abstract Kneser group, present the main properties of these groups, and establish an abstract version of the field theoretic Kneser criterion (Theorem 3.2).

LEMMA 8.3. ([15].) *If G is a finite subgroup of $Z_c^1(\Gamma, A)$, then $(\Gamma : G^{\perp}) \leqslant |G|$.*

PROOF. The canonical cocycle $\eta_G : \Gamma \to \widehat{G}$ defined above, induces an injective map $\Gamma / G^{\perp} \to \widehat{G}$, so $(\Gamma : G^{\perp}) \leqslant |\widehat{G}| = |G|$. □

DEFINITION. A subgroup G of $Z_c^1(\Gamma, A)$ is called a *Kneser group* in $Z_c^1(\Gamma, A)$ if the canonical continuous cocycle $\eta_G : \Gamma \to \widehat{G}$ is onto.

From Lemma 8.3 we deduce that a finite subgroup G of $Z_c^1(\Gamma, A)$ is a Kneser group of $Z_c^1(\Gamma, A)$ if and only if $(\Gamma : G^{\perp}) = |G|$. We shall denote by $\mathcal{K}(\Gamma, A)$ the set of all Kneser groups of $Z_c^1(\Gamma, A)$, partially ordered by inclusion, with $\{0\}$ as the least element.

LEMMA 8.4. ([15].) *If $G \in \mathcal{K}(\Gamma, A)$, then $H \in \mathcal{K}(\Gamma, A)$ for any $H \leqslant G$; in other words, $\mathcal{K}(\Gamma, A)$ is a lower subset of the poset $\mathbb{L}(Z_c^1(\Gamma, A))$.*

PROOF. Since η_H is obtained from η_G by composing it with the canonical epimorphism $\operatorname{res}_H^G : \widehat{G} \to \widehat{H}$, $\varphi \mapsto \varphi|_H$, and η_G is onto by assumption, it follows that the cocycle η_H is onto too, so $H \in \mathcal{K}(\Gamma, A)$. □

PROPOSITION 8.5. ([15].) *If $G \in \mathcal{K}(\Gamma, A)$, then the map $\mathbb{L}(G) \to \overline{\mathbb{L}}(\Gamma)$, $H \mapsto H^{\perp}$, is injective. In particular, $H = G \cap H^{\perp\perp}$ for every $H \in \mathbb{L}(G)$.*

PROOF. Let $H_1, H_2 \in \mathbb{L}(G)$ be such that $H_1^{\perp} = H_2^{\perp}$. We have to show that $H_1 = H_2$. Since $(H_1 + H_2)^{\perp} = H_1^{\perp} \cap H_2^{\perp} = H_1^{\perp} = H_2^{\perp}$, we may assume from the beginning that $H_2 \leqslant H_1$. By Lemma 8.4, $H_i \in \mathcal{K}(\Gamma, A)$, $i = 1, 2$, and hence, the map $\Gamma / H_i^{\perp} \to \widehat{H_i}$ induced by the surjective cocycle η_{H_i} is bijective for $i = 1, 2$. As $\eta_{H_2} = \operatorname{res}_{H_2}^{H_1} \circ \eta_{H_1}$ and $H_1^{\perp} = H_2^{\perp}$ by assumption, it follows that $\operatorname{res}_{H_2}^{H_1} : \widehat{H_1} \to \widehat{H_2}$ is an isomorphism, and hence $H_1 = H_2$ by Pontryagin duality. The last part of the statement is now immediate since $(G \cap H^{\perp\perp})^{\perp} = H^{\perp}$ for any $H \in \mathbb{L}(G)$. □

COROLLARY 8.6. ([15].) *If $Z_c^1(\Gamma, A) \in \mathcal{K}(\Gamma, A)$, then the canonical map $\mathbb{L}(Z_c^1(\Gamma, A)) \to \overline{\mathbb{L}}(\Gamma)$ is injective, and $H = H^{\perp\perp}$ for every $H \in \mathbb{L}(Z_c^1(\Gamma, A))$, i.e., every $H \in \mathbb{L}(Z_c^1(\Gamma, A))$ is a closed element of the Galois connection described in Proposition* 8.1(1).

The next statement shows that the property of a subgroup of $Z_c^1(\Gamma, A)$ being Kneser is of finite character.

PROPOSITION 8.7. ([15].) *If $G \leqslant Z_c^1(\Gamma, A)$, then $G \in \mathcal{K}(\Gamma, A) \Leftrightarrow H \in \mathcal{K}(\Gamma, A)$ for every finite subgroup H of G.*

PROOF. By Lemma 8.4, we have only to prove "$\Leftarrow$". By assumption, the continuous cocycle $\eta_H : \Gamma \to \widehat{H}$ is onto for any finite subgroup H of G. We are going to show that the continuous cocycle $\eta_G : \Gamma \to \widehat{G}$ is also onto. Let $\varphi \in \widehat{G}$. Since the family $(\eta_H^{-1}(\varphi|_H))_H$ of nonempty closed subsets of Γ, for H ranging over all finite subgroups of G, has the finite intersection property, it follows by compactness that $S := \bigcap_H \eta_H^{-1}(\varphi|_H) \neq \varnothing$. Consequently, $\eta_G(\sigma) = \varphi$ for all $\sigma \in S$, as η_G is the projective limit of the projective system of maps $(\eta_H)_H$. Thus η_G is onto, and so $G \in \mathcal{K}(\Gamma, A)$, as desired. □

Using Proposition 8.7 and Zorn lemma, it follows that for any $G \in \mathcal{K}(\Gamma, A)$ there exists a maximal Kneser group lying over G.

COROLLARY 8.8. ([15].) $Z_c^1(\Gamma, A^\Gamma) = \mathrm{Hom}_c(\Gamma, A^\Gamma) \in \mathcal{K}(\Gamma, A)$. *In particular, if the action of Γ on A is trivial, then $Z_c^1(\Gamma, A) = \mathrm{Hom}_c(\Gamma, A) \in \mathcal{K}(\Gamma, A)$.*

PROOF. By Proposition 8.7, we have to show that $G \in \mathcal{K}(\Gamma, A)$ for any finite subgroup G of $Z_c^1(\Gamma, A^\Gamma) = \mathrm{Hom}_c(\Gamma, A^\Gamma)$. For any such G it follows that $G^\perp = \bigcap_{g \in G} \mathrm{Ker}(g)$ is an open normal subgroup of Γ, the quotient $\Gamma/G^\perp$ is a finite Abelian group, and G can be embedded into $\mathrm{Hom}(\Gamma/G^\perp, A^\Gamma) \leqslant \mathrm{Hom}(\Gamma/G^\perp, \mathbb{Q}/\mathbb{Z}) = \widehat{\Gamma/G^\perp} \cong \Gamma/G^\perp$. Consequently, $(\Gamma : G^\perp) \leqslant |G| \leqslant (\Gamma : G^\perp)$ by Lemma 8.3, which shows that $G \in \mathcal{K}(\Gamma, A)$. □

The next two results investigate when an internal direct sum of Kneser subgroups of a given subgroup G of $Z_c^1(\Gamma, A)$ is also Kneser.

PROPOSITION 8.9. ([15].) *Let $G \leqslant Z_c^1(\Gamma, A)$, and assume that G is an internal direct sum of a finite family $(G_i)_{1 \leqslant i \leqslant n}$ of finite subgroups. If* $\gcd(|G_i|, |G_j|) = 1$ *for all $i \neq j$ in $\{1, \dots, n\}$, then*

$$G \in \mathcal{K}(\Gamma, A) \quad \Leftrightarrow \quad G_i \in \mathcal{K}(\Gamma, A), \quad \forall i,\ 1 \leqslant i \leqslant n.$$

PROOF. If every G_i is a Kneser group of $Z_c^1(\Gamma, A)$, then, $|G| = \prod_{1 \leqslant i \leqslant n} |G_i| = \prod_{1 \leqslant i \leqslant n} (\Gamma : G_i^\perp)$. Since $G^\perp \leqslant G_i^\perp$, it follows that $(\Gamma : G_i^\perp) \mid (\Gamma : G^\perp)$ for all $i = 1, \dots, n$. But $(\Gamma : G_i^\perp) = |G_i|$ are mutually relatively prime by hypothesis, hence $\prod_{1 \leqslant i \leqslant n} (\Gamma : G_i^\perp) \mid (\Gamma : G^\perp)$, and so, $|G| \mid (\Gamma : G^\perp)$. On the other hand, $(\Gamma : G^\perp) \leqslant |G|$ by Lemma 8.3, which implies that $|G| = (\Gamma : G^\perp)$, i.e., G is a Kneser group. □

REMARK. In general, an internal direct sum of two nonzero Kneser subgroups of $Z_c^1(\Gamma, A)$ is not necessarily Kneser, as the following example shows. Let $\Gamma = \mathcal{D}_6 = \langle \sigma, \tau \mid \sigma^2 = \tau^3 = (\sigma\tau)^2 = 1 \rangle$, and let $A = (1/3)\mathbb{Z}/\mathbb{Z}$ with the action defined by $\sigma a = -a, \tau a = a$ for $a \in A$. The map $Z_c^1(\Gamma, A) \to A \times A$, $g \mapsto (g(\sigma), g(\tau))$, is a group isomorphism. Let $g, h \in Z_c^1(\Gamma, A)$ be defined by $g(\sigma) = 0$, $h(\sigma) = \widehat{1/3}$,

$g(\tau) = h(\tau) = \widehat{1/3}$. Then, it is easily verified that $Z^1_c(\Gamma, A)$ has two independent Kneser subgroups of order 3, namely, $G = \langle g\rangle$ and $H = \langle h\rangle$, whose (internal direct) sum is not Kneser since $|\Gamma| = 6 < 9 = |G \oplus H|$.

The next result is the *local–global principle* for Kneser groups.

COROLLARY 8.10. ([15].) *A subgroup G of $Z^1_c(\Gamma, A)$ is a Kneser group if and only if every one of its p-primary components $G(p)$ is a Kneser group.*

PROOF. For the nontrivial implication, assume that $G(p) \in \mathcal{K}(\Gamma, A)$ for all $p \in \mathbb{P}$. By Proposition 8.7, we have to prove that any finite subgroup H of G is Kneser. Then $H(p) = G \cap G(p)$, so $H(p)$ is a Kneser group of $Z^1_c(\Gamma, A)$ for all $p \in \mathbb{P}$. If

$$\mathbb{I} := \{p \in \mathbb{P} \mid H(p) \neq 0\},$$

then $H = \bigoplus_{p\in\mathbb{I}} H(p)$. Now, observe that $\mathbb{I}$ is finite and $\gcd(|H(p)|, |H(q)|) = 1$ for all $p \neq q \in \mathbb{I}$. Hence H is a Kneser group by Proposition 8.9. □

We are now going to present the main result of this subsection, namely an abstract version of the (field theoretic) Kneser criterion (Theorem 3.2). To do that, we introduce some basic notation which will be used in the sequel.

Let $\mathcal{N}(\Gamma, A)$ denote the set (possibly empty) $\mathbb{L}(Z^1_c(\Gamma, A)) \setminus \mathcal{K}(\Gamma, A)$ of all subgroups of $Z^1_c(\Gamma, A)$ which are not Kneser groups. Clearly, for any $G \in \mathcal{N}(\Gamma, A)$ there exists at least one minimal member H of $\mathcal{N}(\Gamma, A)$ such that $H \subseteq G$. By $\mathcal{N}(\Gamma, A)_{\min}$ we shall denote the set of all minimal members of $\mathcal{N}(\Gamma, A)$. By Proposition 8.7 and Corollary 8.10, if $G \in \mathcal{N}(\Gamma, A)_{\min}$, then necessarily G is a nontrivial finite p-group for some prime number p.

If $p \in \mathbb{P}$, $p > 2$, and $\widehat{1/p} \in A \setminus A^\Gamma$, define the 1-coboundary $\varepsilon_p \in B^1(\Gamma, (1/p)\mathbb{Z}/\mathbb{Z}) \leqslant B^1(\Gamma, A)$ by $\varepsilon_p(\sigma) = \sigma\widehat{1/p} - \widehat{1/p}, \sigma \in \Gamma$. If $\widehat{1/4} \in A \setminus A^\Gamma$, define the map $\varepsilon'_4 : \Gamma \to (1/4)\mathbb{Z}/\mathbb{Z}$ by

$$\varepsilon'_4(\sigma) = \begin{cases} \widehat{1/4} & \text{if } \sigma\widehat{1/4} = -\widehat{1/4},\\ \widehat{0} & \text{if } \sigma\widehat{1/4} = \widehat{1/4}.\end{cases}$$

It is easily checked that $\varepsilon'_4 \in Z^1_c(\Gamma, (1/4)\mathbb{Z}/\mathbb{Z}) \leqslant Z^1_c(\Gamma, A)$. Observe that ε'_4 has order 4 and $\varepsilon_4 := 2\varepsilon'_4$ is the generator of the cyclic group $B^1(\Gamma, (1/4)\mathbb{Z}/\mathbb{Z}) \leqslant \mathrm{Hom}_c(\Gamma, A[2])$ of order 2.

In the sequel we shall use the following notation: $\mathcal{P}(\Gamma, A) = \{p \in \mathcal{P} \mid \widehat{1/p} \in A \setminus A^\Gamma\}$, where, as usual, $\mathcal{P} = (\mathbb{P} \setminus \{2\}) \cup \{4\}$. We shall also use the following notation:

$$\begin{aligned} B_p &= B^1\big(\Gamma, (1/p)\mathbb{Z}/\mathbb{Z}\big) = B^1\big(\Gamma, A[p]\big) = \langle\varepsilon_p\rangle\\ &\cong \mathbb{Z}/p\mathbb{Z} \quad \text{if } 4 \neq p \in \mathcal{P}(\Gamma, A),\\ B_4 &= \langle\varepsilon'_4\rangle \cong \mathbb{Z}/4\mathbb{Z} \quad \text{if } 4 \in \mathcal{P}(\Gamma, A).\end{aligned}$$

We are now in a position to present a crucial result in abstract co-Galois theory; for its very technical proof, see [15].

LEMMA 8.11. ([15].) *In the notation from above* $\mathcal{N}(\Gamma, A)_{\min} = \{B_p \mid p \in \mathcal{P}(\Gamma, A)\}$.

The next statement, which is an equivalent form of Lemma 8.11, is an abstract version of the Kneser criterion (Theorem 3.2) from field theoretic co-Galois theory. Note that the place of the primitive p-th roots of unity ζ_p, p odd prime, from the Kneser criterion is taken in its abstract version by ε_p, while ε_4' corresponds to $1 - \zeta_4$.

THEOREM 8.12 (Abstract Kneser criterion [15]). *A subgroup G of $Z_c^1(\Gamma, A)$ is Kneser if and only if $\varepsilon_p \notin G$ whenever $4 \neq p \in \mathcal{P}(\Gamma, A)$ and $\varepsilon_4' \notin G$ whenever $4 \in \mathcal{P}(\Gamma, A)$.*

PROOF. Assume that $G \in \mathcal{K}(\Gamma, A)$. If $\varepsilon_p \in G$ for some $4 \neq p \in \mathcal{P}(\Gamma, A)$, then $B_p = \langle \varepsilon_p \rangle \leqslant G$, hence $B_p \in \mathcal{K}(\Gamma, A)$, which contradicts Lemma 8.11. Similarly, if $4 \in \mathcal{P}(\Gamma, A)$ and $\varepsilon_4' \in G$ then $B_4 = \langle \varepsilon_4' \rangle \leqslant G$, hence $B_4 \in \mathcal{K}(\Gamma, A)$, which again contradicts Lemma 8.11.

Assume that $G \notin \mathcal{K}(\Gamma, A)$, i.e., $G \in \mathcal{N}(\Gamma, A)$. Then G contains a minimal member of $\mathcal{N}(\Gamma, A)$, i.e., an element of the set $\mathcal{N}(\Gamma, A)_{\min}$. To conclude, apply again Lemma 8.11. □

COROLLARY 8.13. ([15].) *The following assertions are equivalent for $G \leqslant Z_c^1(\Gamma, A)$.*

(1) *$G \in \mathcal{K}(\Gamma, A)$ and $G = G^{\perp\perp}$.*

(2) *$G^{\perp} \nsubseteq \varepsilon_p^{\perp}$ for all $p \in \mathcal{P}(\Gamma, A)$.*

PROOF. Observe that condition (1) is equivalent to $G^{\perp\perp} \in \mathcal{K}(\Gamma, A)$. On the other hand, by Theorem 8.12, $G^{\perp\perp} \in \mathcal{K}(\Gamma, A)$ if and only if $\varepsilon_p \notin G^{\perp\perp}$ whenever $4 \neq p \in \mathcal{P}(\Gamma, A)$ and $\varepsilon_4' \notin G^{\perp\perp}$ whenever $4 \in \mathcal{P}(\Gamma, A)$. Since $\varepsilon_4^{\perp} = \varepsilon_4'^{\perp} = B_4^{\perp}$ if $4 \in \mathcal{P}(\Gamma, A)$, and $\varepsilon \in G^{\perp\perp} \Leftrightarrow G^{\perp} \subseteq \varepsilon^{\perp}$, the result follows. □

COROLLARY 8.14. ([15].) *$Z_c^1(\Gamma, A)$ is a Kneser group of itself if and only if $\mathcal{P}(\Gamma, A) = \varnothing$, i.e., $A[p] \subseteq A^{\Gamma}$ for all $p \in \mathcal{P}$.*

8.3. *Co-Galois groups of cocycles*

In this subsection we define the concept of an abstract co-Galois group and establish various equivalent characterizations for such groups, including a *quasi-purity criterion*, an abstract version of the structure theorem for Kneser groups from field theoretic co-Galois theory, and an analogue of the abstract Kneser criterion (Theorem 8.12) for co-Galois groups.

For a given subgroup G of $Z_c^1(\Gamma, A)$, the lattice $\mathbb{L}(G)$ of all subgroups of G and the lattice $\overline{\mathbb{L}}(\Gamma | G^{\perp})$ of all closed subgroups of Γ lying over $G^{\perp}$ are related through the canonical order-reversing maps $H \mapsto H^{\perp}$ and $\Delta \mapsto G \cap \Delta^{\perp} = G \cap \mathrm{Ker}(\mathrm{res}_{\Delta}^{\Gamma})$. Clearly, these two maps establish a Galois connection, which is induced by the one considered in Proposition 8.1(1).

DEFINITION. A subgroup G of $Z_c^1(\Gamma, A)$ is said to be a *co-Galois group* of $Z_c^1(\Gamma, A)$ if it is a Kneser group of $Z_c^1(\Gamma, A)$ and the maps

$$(\text{-})^{\perp} : \mathbb{L}(G) \to \overline{\mathbb{L}}(\Gamma | G^{\perp}) \quad \text{and} \quad G \cap (\text{-})^{\perp} : \overline{\mathbb{L}}(\Gamma | G^{\perp}) \to \mathbb{L}(G)$$

are lattice anti-isomorphisms inverse to one another.

Some characterizations of co-Galois groups of $Z_c^1(\Gamma, A)$ are given in the next result.

PROPOSITION 8.15. ([15].) *The following statements are equivalent for a Kneser group G of $Z_c^1(\Gamma, A)$.*

(1) $\Delta = (G \cap \Delta^{\perp})^{\perp}$ *for every* $\Delta \in \overline{\mathbb{L}}(\Gamma | G^{\perp})$.
(2) $\mathrm{res}^{\Gamma}_{\Delta}(G) \in \mathcal{K}(\Delta, A)$ *for every* $\Delta \in \overline{\mathbb{L}}(\Gamma | G^{\perp})$.
(3) *The map* $\mathbb{L}(G) \to \overline{\mathbb{L}}(\Gamma | G^{\perp})$, $H \mapsto H^{\perp}$, *is onto.*
(4) *The map* $\overline{\mathbb{L}}(\Gamma | G^{\perp}) \to \mathbb{L}(G)$, $\Delta \mapsto G \cap \Delta^{\perp}$, *is injective.*
(5) *G is a co-Galois group of $Z_c^1(\Gamma, A)$.*

As $\Gamma \in \overline{\mathbb{L}}(\Gamma | G^{\perp})$ for every $G \leqslant Z_c^1(\Gamma, A)$ and $\mathcal{P}(\Delta, A) \subseteq \mathcal{P}(\Gamma, A)$ for all $\Delta \in \overline{\mathbb{L}}(\Gamma)$, the next result follows immediately from Proposition 8.15 and Corollary 8.14.

COROLLARY 8.16. ([15].) *A subgroup G of $Z_c^1(\Gamma, A)$ is co-Galois if and only if* $\mathrm{res}^{\Gamma}_{\Delta}(G)$ *is a Kneser group of $Z_c^1(\Delta, A)$ for every $\Delta \in \overline{\mathbb{L}}(\Gamma | G^{\perp})$. In particular, $Z_c^1(\Gamma, A)$ is a co-Galois group of itself if and only if $Z_c^1(\Gamma, A)$ is a Kneser group of itself.*

DEFINITION. A subgroup D of an Abelian group C is said to be *quasi-n-pure*, where $n \in \mathbb{N}^*$, if $C[n] \subseteq D$. For $\varnothing \neq M \subseteq \mathbb{N}^*$, C is called *quasi-M-pure* if C is quasi-n-pure for all $n \in M$.

A well-established concept in group theory is that of *n-purity*: a subgroup D of an Abelian group C is said to be *n-pure* if $D \cap nC = nD$. There is no connection between the concepts of n-purity and quasi-n-purity; e.g., the subgroup $2\mathbb{Z}/4\mathbb{Z}$ of $\mathbb{Z}_4$ is quasi-2-pure but not 2-pure, and any of the three subgroups of order 2 of the group $\mathbb{Z}_2 \times \mathbb{Z}_2$ is 2-pure but not quasi-2-pure. Notice that the abstract notion of quasi-n-purity goes back to the concept of n-purity from field theoretic co-Galois theory (see Subsection 4.3).

Recall that for any torsion group D we use the notation $\mathcal{O}_T = \{\mathrm{ord}(x) \mid x \in T\}$, and for any G-radical extension E/F we have used in field theoretic co-Galois theory the notation $\mathcal{P}_G := \mathcal{P} \cap \mathcal{O}_{G/F^*}$, where $\mathcal{P} = (\mathbb{P} \setminus \{2\}) \cup \{4\}$. The same notation, but with another meaning will be used in abstract co-Galois theory: if G is a subgroup of $Z_c^1(\Gamma, A)$, then we write $\mathcal{P}_G := \mathcal{P} \cap \mathcal{O}_G$, i.e., $\mathcal{P}_G$ is the set of those $p \in \mathcal{P}$ for which $\exp(G[p]) = p$.

Quasi-$\mathcal{P}_G$-purity plays a basic role in the characterization of co-Galois groups of $Z_c^1(\Gamma, A)$. The next result, with a very technical proof (see [15]), is the abstract version of the general purity criterion (Theorem 4.3) from field theoretic co-Galois theory.

THEOREM 8.17 (Quasi-purity criterion [15]). *The following statements are equivalent for a subgroup G of $Z_c^1(\Gamma, A)$.*

(1) *G is co-Galois.*
(2) *The subgroup A^{Γ} of $A^{G^{\perp}}$ is quasi-$\mathcal{P}_G$-pure.*
(3) *$G^{\perp} \not\subseteq \varepsilon_p^{\perp}$ for all $p \in \mathcal{P}_G \cap \mathcal{P}(\Gamma, A)$.*

Denote by $\mathcal{C}(\Gamma, A)$ the poset of all co-Galois groups of $Z^1(\Gamma, A)$. Recall that $\mathcal{K}(\Gamma, A)$ is the poset of all Kneser groups of $Z^1(\Gamma, A)$. Using the quasi-purity criterion above, one can show that $\mathcal{C}(\Gamma, A)$ has properties similar to $\mathcal{K}(\Gamma, A)$, namely, $\mathcal{C}(\Gamma, A)$ is a lower Γ-poset, for any $G \in \mathcal{C}(\Gamma, A)$ there exists a maximal co-Galois group lying over G, and, moreover, the property of a subgroup of $Z_c^1(\Gamma, A)$ being co-Galois is a property of finite character. Another consequence of the quasi-purity criterion is the following one.

COROLLARY 8.18. ([15].) *Let p be an odd prime number, and let G be a p-subgroup of $Z_c^1(\Gamma, A)$. Then G is co-Galois if and only if G is Kneser.*

REMARKS.
(1) Corollary 8.18 may fail for $p = 2$. Indeed, the simplest example of a Kneser non-co-Galois 2-group is the one corresponding to an action of type D_4 or D_8 (see Lemma 8.22 and the definition preceding it).
(2) In contrast with the property of Kneser groups given in Corollary 8.10, the condition that all p-primary components of G are co-Galois, is, in general, not sufficient to ensure G being co-Galois. To see that, observe that the group corresponding to the action of type D_{pr} is Kneser but not co-Galois, and has all its primary components co-Galois (see again Lemma 8.22 and the definition preceding it).

We are now going to present a result showing that a subgroup $G \leqslant Z_c^1(\Gamma, A)$ is co-Galois if and only if G has a prescribed structure, which is the abstract version of the structure theorem for Kneser groups from field theoretic co-Galois theory (see Theorem 4.11). To do that some notation is needed. For any subgroup G of $Z_c^1(\Gamma, A)$ and for any prime number p, we write

$$\widetilde{G}_p = \begin{cases} G^{\perp\perp}(p) & \text{if either } p \in \mathcal{P}_G, \text{ or } p = 2 \text{ and } 4 \in \mathcal{P}_G, \\ G^{\perp\perp}[2] & \text{if } p = 2, 4 \notin \mathcal{P}_G, \text{ and } G[2] \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

$\widetilde{G} = \bigoplus_{p \in \mathbb{P}} \widetilde{G}_p$, and $\mu_G = \bigcup_{n \in \mathcal{O}_G} (1/n)\mathbb{Z}/\mathbb{Z}$. Observe that μ_G is a subgroup of A, and hence a discrete Γ-submodule of A too. One easily checks that μ_G is the subgroup $\sum_{g \in G} g(\Gamma)$ of $\mathbb{Q}/\mathbb{Z}$ generated by $\bigcup_{g \in G} g(\Gamma)$, and hence it is the smallest subgroup B of A for which $G \leqslant Z_c^1(\Gamma, B)$. Also note that $\mu_G(p) = \mu_{G(p)} = \bigcup_{g \in G(p)} g(\Gamma)$ for all $p \in \mathbb{P}$.

Let $Z_c^1(\Gamma | G^{\perp}, \mu_G) = G^{\perp\perp} \cap Z_c^1(\Gamma, \mu_G)$ denote the subgroup of $Z_c^1(\Gamma, A)$ consisting of those cocycles which are trivial on $G^{\perp}$ and take values in μ_G. Then $G \leqslant Z_c^1(\Gamma | G^{\perp}, \mu_G) \leqslant \widetilde{G} \leqslant G^{\perp\perp}$, which implies that $G^{\perp} = Z_c^1(\Gamma | G^{\perp}, \mu_G)^{\perp} = \widetilde{G}^{\perp}$. Notice also that $\mathcal{P}_G = \mathcal{P}_{Z_c^1(\Gamma | G^{\perp}, \mu_G)} = \mathcal{P}_{\widetilde{G}}$.

THEOREM 8.19. ([15].) *With the notation above, the following assertions are equivalent for a Kneser group G of $Z_c^1(\Gamma, A)$.*

(1) *G is co-Galois.*
(2) $G = Z^1_c(\Gamma | G^\perp, \mu_G)$.
(3) $G = \widetilde{G}$.

PROOF. (1) ⇒ (3): If G is co-Galois, then $\widetilde{G}$ is also co-Galois by the quasi-purity criterion (Theorem 8.17) since $\mathcal{P}_G = \mathcal{P}_{\widetilde{G}}$ and $G^\perp = \widetilde{G}^\perp$. Therefore, by the definition of the concept of a co-Galois group, we have $H = \widetilde{G} \cap H^{\perp\perp}$ for any $H \in \mathbb{L}(\widetilde{G})$. In particular, we deduce that $G = \widetilde{G} \cap G^{\perp\perp} = \widetilde{G}$, as desired.

(3) ⇒ (2) is trivial.

(2) ⇒ (1): Assume that $G = Z^1_c(\Gamma | G^\perp, \mu_G)$ and G is not co-Galois. Then, by the quasi-purity criterion, there exists $p \in \mathcal{P}_G \cap \mathcal{P}(\Gamma, A)$ such that $G^\perp \subseteq \varepsilon_p^\perp$. Therefore, $\varepsilon_p \in Z^1_c(\Gamma | G^\perp, \mu_G) = G$ for $p \neq 4$, and $\varepsilon'_4 \in Z^1_c(\Gamma | G^\perp, \mu_G) = G$ for $p = 4$. By the abstract Kneser criterion (Theorem 8.12), we deduce that G is not a Kneser group, contrary to our hypothesis. □

COROLLARY 8.20. ([15].) *For any $G, H \in \mathcal{C}(\Gamma, A)$ we have $H \leqslant G$ if and only if $G^\perp \leqslant H^\perp$. In particular, the map $\mathcal{C}(\Gamma, A) \to \overline{\mathbb{L}}(\Gamma)$, $G \mapsto G^\perp$, is injective.*

PROOF. Let $G, H \in \mathcal{C}(\Gamma, A)$ be such that $G^\perp \leqslant H^\perp$, and prove that $H \leqslant G$. By the definition of the groups $\widetilde{G}$ and $\widetilde{H}$, and using Theorem 8.19, it suffices to show that $\mathcal{P}_H \subseteq \mathcal{P}_G$ and $H[2] \neq \{0\} \Rightarrow G[2] \neq \{0\}$. Let $p \in \mathcal{P}_H \cup \{2\}$ and $h \in H$ be such that $\operatorname{ord}(h) = p$. Since $H \in \mathcal{C}(\Gamma, A)$, we have $(\Gamma : h^\perp) = p$, and moreover, there exists only one proper subgroup (of index 2) lying over $h^\perp$ if $p = 4$. Since $G \in \mathcal{C}(\Gamma, A)$ and $G^\perp \leqslant H^\perp \leqslant h^\perp$, it follows that $G \cap h^{\perp\perp}$ is a cyclic subgroup of G of order p, and hence either $p \in \mathcal{P}_G$ or $p = 2$ and $G[2] \neq \{0\}$, as desired. The injectivity of the canonical map $\mathcal{C}(\Gamma, A) \to \overline{\mathbb{L}}(\Gamma)$ is now obvious. □

COROLLARY 8.21. ([15].) *The following assertions are equivalent for a co-Galois G group of $Z^1_c(\Gamma, A)$.*

(1) *G is stable under the action of Γ, i.e., G is a Γ-submodule of $Z^1_c(\Gamma, A)$.*
(2) $G^\perp \lhd \Gamma$.
(3) $\mu_G^{G^\perp} = \mu_G$.

In this case, we have $G \cong Z^1_c(\Gamma / G^\perp, \mu_G)$.

According to Lemma 8.11, the Kneser groups are precisely those subgroups of $Z^1_c(\Gamma, A)$ which do not contain some particular cyclic groups, namely the minimal subgroups B_p which are not Kneser, $p \in \mathcal{P}(\Gamma, A)$. We are now going to present a similar characterization for co-Galois groups. To do that we will first describe effectively the minimal subgroups of $Z^1_c(\Gamma, A)$ which are Kneser but not co-Galois. A special class of actions which are introduced below plays a major role in this description.

DEFINITION. Let Γ be a finite group, and let A be a finite subgroup of $\mathbb{Q}/\mathbb{Z}$ on which the group Γ acts. One says that the action of Γ on A, or the Γ-module A, is

(1) of type D$_4$ if $\Gamma = \mathcal{D}_4 = \langle \sigma, \tau \mid \sigma^2 = \tau^2 = (\sigma\tau)^2 = 1 \rangle \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, $A = (1/4)\mathbb{Z}/\mathbb{Z}$, and $\sigma \widehat{1/4} = -\widehat{1/4}$, $\tau \widehat{1/4} = \widehat{1/4}$;

(2) of type D_8 if $\Gamma = \mathcal{D}_8 = \langle \sigma, \tau \mid \sigma^2 = \tau^4 = (\sigma\tau)^2 = 1\rangle \cong \mathbb{Z}/4\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$, $A = (1/4)\mathbb{Z}/\mathbb{Z}$, and $\sigma\widehat{1/4} = -\widehat{1/4}$, $\tau\widehat{1/4} = \widehat{1/4}$;

(3) of type D_{pr} if $\Gamma = \langle \sigma, \tau \mid \sigma^r = \tau^p = \sigma\tau\sigma^{-1}\tau^{-u} = 1\rangle \cong \mathbb{Z}/p\mathbb{Z} \rtimes_u \mathbb{Z}/r\mathbb{Z}$, $A = (1/pr)\mathbb{Z}/\mathbb{Z}$, and $\sigma\widehat{1/pr} = u\widehat{1/pr}$, $\tau\widehat{1/pr} = \widehat{1/pr}$, where $p \in \mathbb{P}$, $p > 2$, $r \in \mathbb{N}$, $r > 1$, $r \mid (p-1)$, and $u \in (\mathbb{Z}/pr\mathbb{Z})^*$ is such that the order of u mod p in $(\mathbb{Z}/p\mathbb{Z})^*$ is r and u mod $l = 1$ mod l for all $l \in \mathcal{P}$, $l \mid r$.

In the definition above we used the following notation: for any integers $k, m \in \mathbb{Z}$, k mod m denotes the congruence class $k + m\mathbb{Z}$ of k modulo m; if $n \in \mathbb{N}^*$ is a divisor of m, then we write $k + m\mathbb{Z}$ mod n instead of k mod n; finally, $(\mathbb{Z}/m\mathbb{Z})^*$ denotes the group of units of the ring $\mathbb{Z}/m\mathbb{Z}$.

Let $\mathcal{M}(\Gamma, A)$ denote the set (possibly empty) $\mathbb{L}(Z^1_c(\Gamma, A)) \setminus \mathcal{C}(\Gamma, A)$ of all subgroups of $Z^1_c(\Gamma, A)$ which are not co-Galois groups. Clearly, for any $G \in \mathcal{M}(\Gamma, A)$ there exists at least one minimal member H of $\mathcal{M}(\Gamma, A)$ such that $H \subseteq G$. By $\mathcal{M}(\Gamma, A)_{\min}$ we shall denote the set of all minimal members of $\mathcal{M}(\Gamma, A)$, and call them *minimal non-co-Galois groups*. Observe that whenever $G \in \mathcal{M}(\Gamma, A)_{\min}$, then necessarily G is a nontrivial finite group.

LEMMA 8.22. ([15].) *The following conditions are equivalent for a Kneser group G of $Z^1_c(\Gamma, A)$.*

(1) $G \in \mathcal{M}(\Gamma, A)_{\min}$.

(2) *$G^\perp \lhd \Gamma$ and the action of $\Gamma/G^\perp$ on μ_G is of one of the types* $\mathrm{D}_4, \mathrm{D}_8$, *or* D_{pr} *defined above.*

COROLLARY 8.23. ([15].) *Any Kneser minimal non-co-Galois group of $Z^1_c(\Gamma, A)$ is isomorphic either to $\mathbb{Z}/4\mathbb{Z}$, or to $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z}$, or to $\mathbb{Z}/pr\mathbb{Z}$ for an odd prime number p and a divisor $r \neq 1$ of $p - 1$.*

The next result, providing an analogue of the abstract Kneser criterion (Theorem 8.12) for co-Galois groups of cocycles, is an immediate consequence of Lemma 8.22.

THEOREM 8.24. ([15].) *A Kneser subgroup G of $Z^1_c(\Gamma, A)$ is co-Galois if and only if G contains no H for which $H^\perp \lhd \Gamma$ and the action of $\Gamma/H^\perp$ on μ_H is of one of the types* $\mathrm{D}_4, \mathrm{D}_8$, *or* D_{pr}.

As it follows from Lemma 8.22, the fact that all the p-primary components of a subgroup G of $Z^1_c(\Gamma, A)$ are co-Galois does not imply that the whole group G is co-Galois. The next result provides a supplementary lattice theoretic condition which ensures such an implication, obtaining in this way a *local–global principle* for co-Galois groups.

THEOREM 8.25. ([15].) *Let G be a subgroup of $Z^1_c(\Gamma, A)$, and let*

$$\theta : \overline{\mathbb{L}}(\Gamma | G^\perp) \to \prod_{p \in \mathbb{P}} \overline{\mathbb{L}}\big(\Gamma | G(p)^\perp\big), \quad \Delta \mapsto \big(\overline{\langle \Delta \cup G(p)^\perp \rangle}\big)_{p \in \mathbb{P}}.$$

Then, the following statements are equivalent.

(1) *G is co-Galois.*
(2) *$G(p)$ is co-Galois for all prime numbers p, and the order-preserving map θ is a lattice isomorphism.*
(3) *G is Kneser, $G(2)$ is co-Galois, and $\varDelta = \varGamma$ whenever $\varDelta \in \overline{\mathbb{L}}(\varGamma | G^{\perp})$ is such that $\theta(\varDelta) = \theta(\varGamma)$.*

Finally, we consider the case when G is stable under the action of $\varGamma$. Then, the local–global principle for co-Galois groups has the following simple formulation.

PROPOSITION 8.26. ([15].) *The following assertions are equivalent for a $\varGamma$-submodule G of $Z^1_c(\varGamma, A)$.*
(1) *G is co-Galois.*
(2) *$G(p)$ is co-Galois for all prime numbers p.*
(3) *G is Kneser, and $G(2)$ is co-Galois.*

PROOF. The implication (1) $\Rightarrow$ (2) is trivial, while the implication (2) $\Rightarrow$ (3) follows at once from Corollary 8.10.

(3) $\Rightarrow$ (1): Assuming that the $\varGamma$-module G is Kneser but not co-Galois, we have only to show that $G(2)$ is not co-Galois. Let H be a minimal non-co-Galois subgroup of G. According to Lemma 8.22, $H^{\perp} \lhd \varGamma$ and the action of $\varGamma / H^{\perp}$ on μ_H is the one described in definition just before Lemma 8.22. If the action is of type D_4 or of type D_8, then $H \leqslant G(2)$, and hence $G(2)$ is not co-Galois, as desired. Now assume that the action is of type D_{pr}. Then $(\varGamma : H^{\perp} G(p)^{\perp}) = p$. On the other hand, $G(p)^{\perp} \lhd \varGamma$ since $G(p)$ is a $\varGamma$-submodule of G. Hence $H^{\perp} G(p)^{\perp} \lhd \varGamma$, and so, $\mathbb{Z}/p\mathbb{Z}$ is a quotient of $\varGamma / H^{\perp} \cong \mathbb{Z}/p\mathbb{Z} \rtimes_u \mathbb{Z}/r\mathbb{Z}$, which is a contradiction. □

8.4. *Kummer groups of cocycles*

In this subsection we introduce four types of *Kummer groups of cocycles*; these are precisely the abstract group theoretic correspondents of the various types of Kummer field extensions studied in Galois theory and co-Galois theory. We show that all of them are co-Galois groups of cocycles.

DEFINITION. Let $G \leqslant Z^1_c(\varGamma, A)$, and let $n \in \mathbb{N}^*$.
(1) G is said to be a *classical n-Kummer group* if $nG = \{0\}$ and $A[n] \subseteq A^{\varGamma}$.
(2) G is said to be a *generalized n-Kummer group* if $nG = \{0\}$ and $A^{G^{\perp}}[n] \subseteq A^{\varGamma}$.
(3) G is said to be an *n-Kummer group with few cocycles* if $nG = \{0\}$ and $A^{G^{\perp}}[n] \subseteq A[2]$.
(4) G is said to be an *n-quasi-Kummer group* if $nG = \{0\}$ and $A[p] \subseteq A^{\varGamma}$ for every $p \in \mathcal{P}_n$.

We say that G is a *classical Kummer group* (respectively a *generalized Kummer group, Kummer group with few cocycles, quasi-Kummer group*) if G is a classical m-Kummer group (respectively a generalized m-Kummer group, m-Kummer group with few cocycles, m-quasi-Kummer group) for some $m \in \mathbb{N}^*$.

Since $A[2] \subseteq \{\widehat{0}, \widehat{1/2}\} \subseteq A^{\varGamma}$, any n-Kummer group with few cocycles is a generalized n-Kummer group. Clearly, any classical n-Kummer group is both a generalized n-Kummer group and an n-quasi-Kummer group.

PROPOSITION 8.27. ([13].) *Any generalized Kummer group and any quasi-Kummer group is co-Galois. In particular, any classical Kummer group and any Kummer group with few cocycles is co-Galois.*

PROOF. Let $G \leqslant Z_c^1(\varGamma, A)$. If G is a generalized Kummer group, then $nG = \{0\}$ and $A^{G^{\perp}}[n] \subseteq A^{\varGamma}$ for some $n \in \mathbb{N}^*$. If $p \in \mathcal{P}_G$, then $p \mid n$, and hence $A^{G^{\perp}}[p] \subseteq A^{G^{\perp}}[n] \subseteq A^{\varGamma}$, i.e., the subgroup $A^{\varGamma}$ of $A^{G^{\perp}}$ is quasi-$\mathcal{P}_G$-pure. Thus G is a co-Galois group of $Z_c^1(\varGamma, A)$ by the quasi-purity criterion (Theorem 8.17).

If G is a quasi-Kummer group, then there exists $n \in \mathbb{N}^*$ with $nG = \{0\}$ and $A[p] \subseteq A^{\varGamma}$ for every $p \in \mathcal{P}_n$. Since $\mathcal{P}_G \subseteq \mathcal{P}_n$, we have $A^{G^{\perp}}[p] \subseteq A[p] \subseteq A^{\varGamma}$ for every $p \in \mathcal{P}_G$, which shows that the subgroup $A^{\varGamma}$ of $A^{G^{\perp}}$ is quasi-$\mathcal{P}_G$-pure. Again by Theorem 8.17, G is a co-Galois group of $Z_c^1(\varGamma, A)$. □

COROLLARY 8.28. ([13].) *Let $G \leqslant Z_c^1(\varGamma, A)$ be any of the four types of Kummer groups of cocycles defined above. Then the maps*

$$(\text{-})^{\perp} : \mathbb{L}(G) \to \overline{\mathbb{L}}(\varGamma | G^{\perp}) \quad \textit{and} \quad G \cap (\text{-})^{\perp} : \overline{\mathbb{L}}(\varGamma | G^{\perp}) \to \mathbb{L}(G)$$

are lattice anti-isomorphisms, inverse to one another.

The next two results are the abstract analogs of Propositions 5.27 and 5.29 from field theoretic Kummer theory.

PROPOSITION 8.29. ([13].) *Let $G \leqslant Z_c^1(\varGamma, A)$ be a co-Galois group of bounded order such that $A[\exp(G)] \subseteq A^{G^{\perp}}$. Then G is a quasi-Kummer group.*

PROPOSITION 8.30. ([13].) *Any generalized Kummer group $G \leqslant Z_c^1(\varGamma, A)$ with $A[\exp(G)] \subseteq A^{G^{\perp}}$ is a classical Kummer group.*

8.5. *Field theoretic co-Galois theory via abstract co-Galois theory*

The aim of this subsection is two-fold: firstly, to present a field theoretic co-Galois theory ↔ abstract co-Galois theory dictionary, and secondly, using this dictionary, to show how some basic results of field theoretic co-Galois theory, like the Kneser criterion, the general purity criterion, the uniqueness of the Kneser group of a G-co-Galois field extension, etc., can be very easily deduced form their abstract versions.

The abstract correspondents for subgroups of $Z_c^1(\varGamma, A)$ of Kneser field extensions and co-Galois field extensions are those of Kneser groups of cocycles and co-Galois groups of cocycles, respectively. If we move now via the maps $(\text{-})^{\perp}$ from subgroups of $Z_c^1(\varGamma, A)$ to subgroups of the given profinite group $\varGamma$, then one can define as follows the abstract versions for the latter of the concepts of radical, simple radical, Kneser, and co-Galois field extension.

DEFINITION. Let $\varDelta$ be a subgroup of $\varGamma$. We say that $\varDelta$ is *G-radical* if $\varDelta = G^{\perp}$ for some $G \leqslant Z_c^1(\varGamma, A)$. A *radical subgroup* of $\varGamma$ is a subgroup which is G-radical for some $G \leqslant Z_c^1(\varGamma, A)$. $\varDelta$ is called *simple radical* if there exists a $g \in Z_c^1(\varGamma, A)$ such that $\varDelta = g^{\perp}$.

$\varDelta$ is said to be *G-Kneser* if $\varDelta$ is G-radical and G is a Kneser group of $Z_c^1(\varGamma, A)$. $\varDelta$ is said to be a *Kneser* group if $\varDelta$ is G-Kneser for some $G \leqslant Z_c^1(\varGamma, A)$.

$\varDelta$ is said to be *co-Galois* if there exists a co-Galois group G of $Z_c^1(\varGamma, A)$ such that $\varDelta = G^{\perp}$. $\varDelta$ is said to be *strongly co-Galois* if $\varDelta = \varDelta^{\perp\perp}$ and $\varDelta^{\perp}$ is a co-Galois group of $Z_c^1(\varGamma, A)$.

Observe that any radical subgroup of $\varGamma$ is necessarily closed, and any simple radical subgroups of $\varGamma$ is open. If $\varDelta$ is co-Galois, then the co-Galois group G of $Z_c^1(\varGamma, A)$ for which $\varDelta = G^{\perp}$ is uniquely determined by Corollary 8.20, and we say in this case that $\varDelta$ is *G-co-Galois*.

LEMMA 8.31. ([13].) *A radical subgroup $\varDelta$ of $\varGamma$ is strongly co-Galois if and only if $\varDelta^{\perp}$ is a Kneser group of $Z_c^1(\varGamma, A)$.*

PROOF. If $\varDelta$ is a radical subgroup of $\varGamma$, then there exists $H \leqslant Z_c^1(\varGamma, A)$ such that $\varDelta = H^{\perp}$, and so $\varDelta^{\perp\perp} = (H^{\perp})^{\perp\perp} = H^{\perp} = \varDelta$. If $G := \varDelta^{\perp}$ is not a co-Galois group of $Z_c^1(\varGamma, A)$, then we are going to show that G is not Kneser. As G is not co-Galois, it follows by the quasi-purity criterion (Theorem 8.17) that there exists $p \in \mathcal{P}(\varGamma, A) \cap \mathcal{P}_G$ such that $\varDelta = G^{\perp} \leqslant \varepsilon_p^{\perp}$, and hence $\varepsilon_p^{\perp\perp} \leqslant \varDelta^{\perp} = G$. Consequently, $\varepsilon_p \in \varepsilon_p^{\perp\perp} \leqslant G$ if $p \neq 4$, and $\varepsilon_4' \in \varepsilon_4^{\perp\perp} \leqslant G$ if $p = 4$. By the abstract Kneser criterion (Theorem 8.12) we deduce that G is not Kneser. □

We establish now a dictionary relating the basic notions of field theoretic co-Galois theory to their analogs in the group theoretic abstract co-Galois theory; this will allow us to recover at the end of this subsection some of the main results of the former theory from the latter one.

In the sequel, $\varOmega/F$ denotes a fixed Galois extension with the (profinite) Galois group $\varGamma := \mathrm{Gal}(\varOmega/F)$. In particular, we can take as $\varOmega$ an algebraic separable closure $\widetilde{F}^{\mathrm{sep}}$ of the base field F, in which case $\varGamma$ is the absolute Galois group of F. If $\varOmega = \widetilde{F}^{\mathrm{sep}}$ and the characteristic $\mathrm{Char}(F)$ of F is p, then the multiplicative torsion group $\mu(\varOmega)$ of all roots of unity contained in $\varOmega$ is isomorphic in a non-canonical way to the additive group $\mathbb{Q}/\mathbb{Z}$ if $p = 0$, respectively to its subgroup $\bigoplus_{q \in \mathbb{P}\setminus\{p\}} (\mathbb{Q}/\mathbb{Z})(q)$ for $p \neq 0$. Thus, the group $A := \mu(\varOmega)$ is isomorphic to a uniquely determined subgroup of $\mathbb{Q}/\mathbb{Z}$, and the canonical action of $\varGamma$ on $\varOmega$ induces a continuous action of the profinite group $\varGamma$ on the discrete group A.

The exact sequence $\{1\} \to A \to \varOmega^* \to \varOmega^*/A \to \{1\}$ of topological discrete $\varGamma$-modules yields the exact sequence of cohomology groups in low dimensions

$$\{1\} \to A^{\varGamma} \to \varOmega^{*\varGamma} \to (\varOmega^*/A)^{\varGamma} \to H_c^1(\varGamma, A) \to H_c^1(\varGamma, \varOmega^*).$$

Since $H^1_c(\Gamma, \Omega^*) = \{1\}$ by Hilbert theorem 90, we obtain a canonical epimorphism of Abelian torsion groups

$$\psi : T(\Omega/F) \to Z^1_c(\Gamma, A), \quad x \mapsto \big(\sigma \in \Gamma \mapsto (\sigma x)x^{-1} \in A\big),$$

whose kernel is F^*, where

$$T(\Omega/F) = \big\{x \in \Omega^* \mid (\sigma x)x^{-1} \in A, \forall \sigma \in \Gamma\big\} = \big\{x \in \Omega^* \mid \exists n \in \mathbb{N}^*, x^n \in F\big\}.$$

The quotient $T(\Omega/F)/F^*$ is exactly the torsion subgroup of the quotient group Ω^*/F^*, that is, the co-Galois group $\operatorname{Cog}(\Omega/F)$ of the field extension Ω/F. Thus, the epimorphism ψ induces a canonical isomorphism

$$\varphi : \operatorname{Cog}(\Omega/F) \overset{\sim}{\to} Z^1_c(\Gamma, A)$$

(see also Theorem 5.5), which identifies in a canonical way the subgroups $G \leqslant Z^1_c(\Gamma, A)$ investigated in the frame of abstract co-Galois theory with the subgroups $\mathbb{G}/F^* := \varphi^{-1}(G) \leqslant \operatorname{Cog}(\Omega/F)$ investigated in the frame of field theoretic co-Galois theory (see Corollary 5.7). In particular, for every intermediate field E of Ω/F, the restriction of ψ to $T(E/F) = T(\Omega/F) \cap E$ induces an isomorphism from the torsion group $\operatorname{Cog}(E/F) := T(E/F)/F^*$ of E^*/F^* onto the subgroup $\Gamma_E^{\perp}$ of $Z^1_c(\Gamma, A)$, where $\Gamma_E := \operatorname{Gal}(\Omega/E)$.

The lattice $\mathbb{I}(\Omega/F)$ of all intermediate fields of the extension Ω/F, the lattice $\mathbb{L}(T(\Omega/F)|F^*)$ of all subgroups of $T(\Omega/F)$ lying over F^*, the lattice $\overline{\mathbb{L}}(\Gamma)$ of all closed subgroups of Γ, and the lattice $\mathbb{L}(Z^1_c(\Gamma, A))$ of all subgroups of $Z^1_c(\Gamma, A)$ are related as shown in the commutative diagram:

$$\begin{tikzcd}
\mathbb{L}(T(\Omega/F)|F^*) \arrow[r, shift left] \arrow[d] & \mathbb{I}(\Omega/F) \arrow[l, shift left] \arrow[d] \\
\mathbb{L}(Z^1_c(\Gamma, A)) \arrow[r, shift left] & \overline{\mathbb{L}}(\Gamma) \arrow[l, shift left]
\end{tikzcd}$$

where the left vertical arrow is the lattice isomorphism induced by ψ, the right vertical arrow is the canonical lattice anti-isomorphism $E \mapsto \Gamma_E$ (with inverse $\Delta \mapsto E^{\Delta}$) given by the fundamental theorem of infinite Galois theory, the horizontal top arrows are the sup-semilattice morphism $\mathbb{G} \mapsto F(\mathbb{G})$ and the inf-semilattice morphism $E \mapsto T(E/F)$, while the horizontal bottom arrows are the sup-semilattice anti-morphism $G \mapsto G^{\perp}$ and the inf-semilattice anti-morphism $\Delta \mapsto \Delta^{\perp}$ defined in Proposition 8.1. Note that the commutativity of the diagram above follows at once from Proposition 5.11.

The next result is essentially a reformulation of the corresponding results from Subsection 5.2 involving the lattices and the maps above.

PROPOSITION 8.32. ([13].) *Let E be an intermediate field of the given Galois extension Ω/F, let $\Gamma_E = \operatorname{Gal}(\Omega/E)$, let $A = \mu(\Omega)$, let $\mathbb{G} \in \mathbb{L}(T(\Omega/F)|F^*)$, and let $G = \psi(\mathbb{G})$, where ψ is the canonical group epimorphism $\psi : T(\Omega/F) \to Z^1_c(\Gamma, A)$ defined above. Then, the following statements hold.*

1. *The extension E/F is $\mathbb{G}$-radical if and only if the subgroup Γ_E of Γ is G-radical. In particular, E/F is a radical extension* (*respectively a simple radical extension*) *if and only if Γ_E is a radical group* (*respectively a simple radical group*) *of Γ.*

(2) *The extension E/F is $\mathbb{G}$-Kneser if and only if the subgroup Γ_E of Γ is G-Kneser. In particular, E/F is a Kneser extension if and only if Γ_E is a Kneser group of Γ.*

(3) *The extension $F(\mathbb{G})/F$ is $\mathbb{G}$-Kneser if and only if G is a Kneser group of $Z_c^1(\Gamma, A)$.*

(4) *The extension E/F is $\mathbb{G}$-co-Galois if and only if the subgroup Γ_E of Γ is co-Galois. In this case, G is the unique co-Galois group of $Z_c^1(\Gamma, A)$ for which $\Gamma_E = G^{\perp}$.*

(5) *The extension $F(\mathbb{G})/F$ is $\mathbb{G}$-co-Galois if and only if G is a co-Galois group of $Z_c^1(\Gamma, A)$.*

(6) *The extension E/F is co-Galois if and only if the subgroup Γ_E of Γ is strongly co-Galois.*

Proof. (1) is a reformulation of Proposition 5.11, (2) is a reformulation of Corollary 5.12(1), and (4) is a reformulation of Corollary 5.12(2). The uniqueness of G is assured by Corollary 8.20.

(6) Write $\mathbb{H} = T(E/F)$ and $H = \psi(\mathbb{H}) = \Gamma_E^{\perp}$. By (2), the extension E/F is co-Galois, i.e., $\mathbb{H}$-Kneser, if and only if $\Gamma_E = H^{\perp} = \Gamma_E^{\perp\perp}$ and $\Gamma_E^{\perp}$ is a Kneser group of $Z_c^1(\Gamma, A)$. By Lemma 8.31, this means precisely that Γ_E is strongly-co-Galois. □

Corollary 8.33. ([13].) *Let Ω/F be a Galois extension, $\Gamma := \mathrm{Gal}(\Omega/F)$, $A := \mu(\Omega)$, and let $F^* \leqslant \mathbb{G} \leqslant T(\Omega/F)$ be such that $E := F(\mathbb{G})$ is a $\mathbb{G}$-co-Galois extension of F. If $G := \psi(\mathbb{G}) \leqslant Z_c^1(\Gamma, A)$, then, the following assertions are equivalent.*

(1) *G is a Γ-submodule of $Z_c^1(\Gamma, A)$, i.e., it is stable under the action of Γ.*

(2) *E/F is a Galois extension.*

(3) *$\sigma x \in E$ for all $\sigma \in \Gamma$ and $x \in \mathbb{G}$.*

Proof. First, observe that $\Gamma_E := \mathrm{Gal}(\Omega/E) = G^{\perp}$ by Proposition 8.32(4). Now, by Corollary 8.21, G is a Γ-submodule of $Z_c^1(\Gamma, A)$ if and only if $G^{\perp}$ is a normal subgroup of Γ if and only if E/F is a Galois extension. □

The connection between various types of Kummer field extensions and their abstract correspondents is given by the next result.

Proposition 8.34. ([13].) *Let E/F be an arbitrary separable algebraic extension, let $\Omega := \widetilde{F}^{\mathrm{sep}}$, $\Gamma = \mathrm{Gal}(\Omega/F)$, $A = \mu(\Omega)$, and let $n \in \mathbb{N}^*$ be such that* $\gcd(n, e(F)) = 1$. *Then, the extension E/F is a classical n-Kummer extension (respectively a generalized n-Kummer extension, an n-Kummer extension with few roots of unity, an n-quasi-Kummer extension) if and only if there exists a unique classical n-Kummer group (respectively a generalized n-Kummer group, an n-Kummer group with few cocycles, an n-quasi-Kummer group) G, $G \leqslant Z_c^1(\Gamma, A)$, such that $\Gamma_E :=$* $\mathrm{Gal}(\Omega/E) = G^{\perp}$.

Proof. We may assume that $E \subseteq \Omega$. If E/F is a classical n-Kummer extension, then there exists a group $\mathbb{G} \in \mathbb{L}(T(E/F)|F^*)$ such that $E = F(\mathbb{G})$, $\mathbb{G}^n \subseteq F^*$, and $A[n] = \mu_n(\Omega) \subseteq \mu_n(F) \subseteq A^{\Gamma}$. Let $G = \psi(\mathbb{G})$. Then $nG = \{0\}$, and so G is a classical n-Kummer group. By Proposition 8.32(1) we have $\Gamma_E := \mathrm{Gal}(\Omega/E) = G^{\perp}$. Now observe that G is co-Galois by Proposition 8.27, so the uniqueness of G follows from Proposition 8.32(4).

Conversely, assume that there exists a classical n-Kummer group G of $Z_c^1(\Gamma, A)$ such that $\mathrm{Gal}(\Omega/E) = G^{\perp}$. If we denote $\mathbb{G} = \psi^{-1}(G)$, then $E = F(\mathbb{G})$ by Proposition 8.32(1). Since clearly $\mathbb{G}^n = 1$ and $\mu_n(\Omega) = A[n] \subseteq A^{\Gamma} \subseteq F$, we deduce that E/F is a classical n-Kummer extension.

The cases of generalized n-Kummer extensions, n-Kummer extensions with few roots of unity, and n-quasi-Kummer extensions follow in the same manner as above from the following simple facts: $A[n] = \mu_n(\Omega)$ (in particular $A[2] = \{-1, 1\}$) and $A^{\mathrm{Gal}(\Omega/L)} = \{x \in \mu(\Omega) \mid \sigma x = x, \forall \sigma \in \mathrm{Gal}(\Omega/L)\} = \mu(L)$ for any intermediate field L of the given Galois extension Ω/F. □

By Proposition 8.34, all the types of Kummer groups of cocycles defined in Subsection 8.4 are abstract versions of corresponding field extensions from field theoretic Kummer theory. So, the counterexamples from field theoretic Kummer theory, converted into Kummer groups of cocycles via Proposition 8.34, show that, except the obvious inclusions indicated just before Proposition 8.27, no other inclusions between these four types of Kummer groups of cocycles do exist.

The results above permit us now to retrieve easily most of the results of field theoretic co-Galois theory from the basic results of abstract co-Galois theory. We will illustrate this by presenting only three of them.

KNESER CRITERION (Theorem 3.2). *Let E/F be an arbitrary separable $\mathbb{G}$-radical extension. For any $n \in \mathbb{N}^*$, let $\zeta_n \in \Omega := \widetilde{F}^{\mathrm{sep}}$ denote a primitive n-th root of unity. Then, the following assertions are equivalent.*

(1) *E/F is a $\mathbb{G}$-Kneser extension.*

(2) *$\zeta_p \in \mathbb{G} \Rightarrow \zeta_p \in F$ for every odd prime p, and $1 \pm \zeta_4 \in \mathbb{G} \Rightarrow \zeta_4 \in F$.*

PROOF. We may assume that $E \subseteq \Omega$. Set $\Gamma = \mathrm{Gal}(\Omega/F)$, $A = \mu(\Omega)$, and let $\psi : T(\Omega/F) \to Z_c^1(\Gamma, A)$ be the canonical group epimorphism defined above. Then $A^{\Gamma} = \mu(F)$ and $\mathcal{P}(\Gamma, A) = \{p \in \mathcal{P} \mid \zeta_p \notin F\}$. By assumption, $E = F(\mathbb{G})$, with $F^* \leqslant \mathbb{G} \leqslant T(\Omega/F)$. Setting $G := \psi(\mathbb{G}) \leqslant Z_c^1(\Gamma, A)$ we have $\Gamma_E = G^{\perp}$ by Proposition 8.32(1). Consequently, by Proposition 8.32(3), the extension E/F is $\mathbb{G}$-Kneser if and only if G is a Kneser group of $Z_c^1(\Gamma, A)$.

For every odd prime $p \neq \mathrm{Char}(F)$, $\varepsilon_p := \psi(\zeta_p) \in Z_c^1(\Gamma, A)$ is the coboundary assigning to any $\sigma \in \Gamma$ the p-th root of unity $(\sigma\zeta_p)\zeta_p^{-1} \in A[p]$. Obviously, $\varepsilon_p \in G$ if and only if $\zeta_p \in \mathbb{G}$. Observe that if $p = \mathrm{Char}(F) > 2$, then $\zeta_p \in A[p] = \{1\} \subseteq A^{\Gamma}$.

Assume that $\mathrm{Char}(F) \neq 2$. Since $1 - \zeta_4 \in T(\Omega/F)$, we can consider the continuous cocycle $\psi(1 - \zeta_4) \in Z_c^1(\Gamma, A)$, which by definition works as follows:

$$\psi(1-\zeta_4)(\sigma) = \sigma(1-\zeta_4)\cdot(1-\zeta_4)^{-1} = (1-\sigma\zeta_4)\cdot(1-\zeta_4)^{-1}, \quad \forall \sigma \in \Gamma.$$

Since for any $\sigma \in \Gamma$, we have either $\sigma\zeta_4 = \zeta_4$ or $\sigma\zeta_4 = -\zeta_4$, we deduce that

$$\psi(1-\zeta_4)(\sigma) = \begin{cases} \zeta_4 & \text{if } \sigma\zeta_4 = -\zeta_4, \\ 1 & \text{if } \sigma\zeta_4 = \zeta_4. \end{cases}$$

Thus, $\psi(1 - \zeta_4)$ is nothing else than the multiplicative version of the cocycle ε_4' defined in Subsection 8.2 just before Lemma 8.11 and appearing in the statement of the abstract Kneser criterion (Theorem 8.12). A simple calculation shows that $\psi(1+\zeta_4) = (\psi(1-\zeta_4))^{-1}$ in the multiplicative group $Z_c^1(\Gamma, A)$, so $\varepsilon_4' \in G \Leftrightarrow 1 \pm \zeta_4 \in \mathbb{G}$. Observe that if $\mathrm{Char}(F) = 2$, then $\zeta_4 \in A[4] = \{1\} \subseteq A^\Gamma$. To finish the proof it remains to apply Proposition 8.32(3) and the abstract Kneser criterion (Theorem 8.12). □

COROLLARY 8.35. ([13].) *Let E/F be a separable $\mathbb{G}$-radical extension, which is not $\mathbb{G}$-Kneser. Assume that the extension E/F is minimal with respect to the property not being $\mathbb{G}$-Kneser, that is, for any proper subgroup $\mathbb{G}'$ of $\mathbb{G}$, the extension $F(\mathbb{G}')/F$ is $\mathbb{G}'$-Kneser. Then, the extension E/F is cyclic having either the form $E = F(\zeta_p)$ with $p \neq \mathrm{Char}(F)$ an odd prime number and $\zeta_p \notin F$, or the form $F(\zeta_4)$ with $\mathrm{Char}(F) \neq 2$ and $\zeta_4 \notin F$.*

PROOF. With $\Omega = \widetilde{F}^{\mathrm{sep}}$, Γ, and A as above, let E/F be a subextension of Ω/F satisfying the minimality condition from the statement. Using the canonical group epimorphism $\psi : T(\Omega/F) \to Z_c^1(\Gamma, A)$ as well as Proposition 8.32, we deduce that $G = \psi(\mathbb{G})$ is a minimal non-Kneser group of $Z_c^1(\Gamma, A)$. According to Lemma 8.11, it follows that either $G = \langle \varepsilon_p \rangle \cong \mathbb{Z}/p\mathbb{Z}$ for some odd prime number $p \neq \mathrm{Char}(F)$ such that $\zeta_p \notin F$, or $G = \langle \varepsilon_4' \rangle \cong \mathbb{Z}/4\mathbb{Z}$, with $\mathrm{Char}(F) \neq 2$ and $\zeta_4 \notin F$. Consequently, $\mathbb{G} = F^*\langle \zeta_p \rangle$ in the former case and $\mathbb{G} = F^*\langle 1+\zeta_4 \rangle$ in the latter one. The result now follows easily. □

Observe that the inverse implication in Corollary 8.35 may fail. Indeed, $F(\zeta_4)/F$ is $F^*\langle \zeta_4 \rangle$-co-Galois, in particular Kneser, whenever $\mathrm{Char}(F) \neq 2$ and $\zeta_4 \notin F$. Also, for every odd prime p, if $e(F)$ is prime with $p(p-1)$, $\zeta_p \notin F$, and $\zeta_{p-1} \in F$, then there exists $\theta \in E := F(\zeta_p)$ such that $E = F(\theta)$ and $\theta^{p-1} \in F$; therefore E/F is an $F^*\langle \theta \rangle$-co-Galois extension, in particular Kneser.

GENERAL PURITY CRITERION (Theorem 4.3). *A separable $\mathbb{G}$-radical extension E/F is $\mathbb{G}$-co-Galois if and only if it is $\mathcal{P}_{\mathbb{G}}$-pure.*

PROOF. We may assume that $E \subseteq \Omega := \widetilde{F}^{\mathrm{sep}}$. Set $\Gamma := \mathrm{Gal}(\Omega/F)$ and $A := \mu(\Omega)$. Since E/F is a $\mathbb{G}$-radical extension, we have $E = F(\mathbb{G})$ with $F^* \leqslant \mathbb{G} \leqslant T(\Omega/F)$. If $G := \psi(\mathbb{G}) \leqslant Z_c^1(\Gamma, A)$, then $\Gamma_E := \mathrm{Gal}(\Omega/E) = G^\perp$ by Proposition 8.32(1), so E/F is $\mathbb{G}$-co-Galois if and only if G is a co-Galois subgroup of $Z_c^1(\Gamma, A)$ by Proposition 8.32(5). Since for any $p \in \mathcal{P}_{\mathbb{G}}$ we have $A^\Gamma[p] = \mu_p(F)$ and $A^{G^\perp}[p] = \mu_p(E)$, we deduce that the $\mathcal{P}_{\mathbb{G}}$-purity of the extension E/F is equivalent to the quasi-$\mathcal{P}_G$-purity of the embedding $A^\Gamma \leqslant A^{G^\perp}$. The result follows now at once by applying Theorem 8.17. □

UNIQUENESS OF THE KNESER GROUP (Theorem 4.13). *If E/F is an algebraic separable extension which is simultaneously $\mathbb{G}$-co-Galois and $\mathbb{H}$-co-Galois, then $\mathbb{G} = \mathbb{H}$.*

PROOF. Apply Corollary 8.20 and Proposition 8.32(4). □

References

[1] M. Acosta de Orozco, W.Y. Vélez, The lattice of subfields of a radical extension, J. Number Theory 15 (1982) 388–405.

[2] M. Acosta de Orozco, W.Y. Vélez, The torsion group of a field defined by radicals, J. Number Theory 19 (1984) 283–294.

[3] T. Albu, Kummer extensions with few roots of unity, J. Number Theory 41 (1992) 322–358.

[4] T. Albu, Some examples in Cogalois Theory with applications to elementary Field Arithmetic, J. Algebra Appl. 1 (2002) 1–29.

[5] T. Albu, Infinite field extensions with Cogalois correspondence, Comm. Algebra 30 (2002) 2335–2353.

[6] T. Albu, Infinite field extensions with Galois–Cogalois correspondence (I), Rev. Roumaine Math. Pures Appl. 47 (2002) 1–20.

[7] T. Albu, Infinite field extensions with Galois–Cogalois correspondence (II), Rev. Roumaine Math. Pures Appl. 47 (2002) 149–161.

[8] T. Albu, Field extensions with the unique subfield property, and G-Cogalois extensions, Turkish J. Math. 26 (2002) 433–445.

[9] T. Albu, On radical field extensions of prime exponent, J. Algebra Appl. 1 (2002) 365–373.

[10] T. Albu, Corrigendum and Addendum to my paper concerning Kummer extensions with few roots of unity, J. Number Theory 99 (2003) 222–224.

[11] T. Albu, Cogalois Theory, Monographs Textbooks, Pure Appl. Math., vol. 252, Dekker, New York, 2003.

[12] T. Albu, Infinite Cogalois Theory, Clifford extensions, and Hopf algebras, J. Algebra Appl. 2 (2003) 119–136.

[13] T. Albu, Field theoretic Cogalois Theory via Abstract Cogalois Theory, J. Pure Appl. Algebra 208 (2007) 101–106.

[14] T. Albu, Ş.A. Basarab, Lattice-isomorphic groups, and infinite Abelian G-Cogalois field extensions, J. Algebra Appl. 1 (2002) 243–253.

[15] T. Albu, Ş.A. Basarab, An Abstract Cogalois Theory for profinite groups, J. Pure Appl. Algebra 200 (2005) 227–250.

[16] T. Albu, F. Nicolae, Kneser field extensions with Cogalois correspondence, J. Number Theory 52 (1995) 299–318.

[17] T. Albu, F. Nicolae, G-Cogalois field extensions and primitive elements, in: M. Behara, R. Fritsch, R.G. Lintz (Eds.), Symposia Gaussiana, Conference A: Mathematics and Theoretical Physics, de Gruyter, Berlin, 1995, pp. 233–240.

[18] T. Albu, F. Nicolae, Heckesche Systeme idealer Zahlen und Knesersche Körpererweiterungen, Acta Arith. 73 (1995) 43–50.

[19] T. Albu, F. Nicolae, Finite radical field extensions and crossed homomorphisms, J. Number Theory 60 (1996) 291–309.

[20] T. Albu, F. Nicolae, M. Ţena, Some remarks on G-Cogalois field extensions, Rev. Roumaine Math. Pures Appl. 41 (1996) 145–153.

[21] T. Albu, M. Ţena, Infinite Cogalois Theory, Math. Rep. (Bucur.) 3 (53) (2001) 105–132.

[22] E. Artin, Galoissche Theorie, Teubner, Leipzig, 1959.

[23] R. Baer, The significance of the system of subgroups for the structure of the group, Amer. J. Math. 61 (1939) 1–44.

[24] A. Baker, H.M. Stark, On a fundamental inequality in number theory, Ann. of Math. 94 (1971) 190–199.

[25] F. Barrera-Mora, On subfields of radical extensions, Comm. Algebra 27 (1999) 4641–4649.

[26] F. Barrera-Mora, M. Rzedowski-Calderón, G. Villa-Salvador, On Cogalois extensions, J. Pure Appl. Algebra 76 (1991) 1–11.

[27] F. Barrera-Mora, M. Rzedowski-Calderón, G. Villa-Salvador, Allowable groups and Cogalois extensions, J. Pure Appl. Algebra 104 (1995) 123–147.

[28] F. Barrera-Mora, W.Y. Vélez, Some results on radical extensions, J. Algebra 162 (1993) 295–301.

[29] E. Becker, R. Grobe, M. Niermann, Radicals of binomial ideals, J. Pure Appl. Algebra 117–118 (1997) 41–79.

[30] A. Besicovitch, On the linear independence of fractional powers of integers, J. London Math. Soc. 15 (1940) 3–6.

[31] N. Bourbaki, Algèbre, Chapitres 4 à 7, Masson, Paris, 1981.

[32] T. Brzeziński, R. Wisbauer, Corings and Comodules, Cambridge Univ. Press, Cambridge, 2003.

[33] S. Caenepeel, Brauer Groups, Hopf Algebras and Galois Theory, Kluwer Acad. Publ., Dordrecht, 1998.

[34] J.W.S. Cassels, A. Fröhlich, Algebraic Number Theory, Academic Press, London, 1967.

[35] E.C. Dade, Compounding Clifford's theory, Ann. of Math. 91 (1970) 236–290.

[36] E.C. Dade, Isomorphisms of Clifford extensions, Ann. of Math. 92 (1970) 375–433.

[37] E.C. Dade, Group-graded rings and modules, Math. Z. 174 (1980) 241–262.

[38] S. Dăscălescu, C. Năstăsescu, Ş. Raianu, Hopf Algebras: An Introduction, Dekker, New York, 2001.

[39] D.S. Dummit, On the torsion in quotients of the multiplicative groups in Abelian extensions, in: J.M. De Koninck, C. Levesque (Eds.), Number Theory, Proceedings of the International Number Theory Conference, Université Laval, 1987, De Gruyter, Berlin, 1989.

[40] E.E. Enochs, J.R. Garcia Rozas, L. Oyonarte, Compact coGalois groups, Math. Proc. Cambridge Philos. Soc. 128 (2000) 233–244.

[41] E.E. Enochs, J.R. Garcia Rozas, L. Oyonarte, Covering morphisms, Comm. Algebra 28 (2000) 3823–3835.

[42] L. Gaal, Classical Galois Theory with Examples, Markham, Chicago, IL, 1971.

[43] D. Gay, W.Y. Vélez, On the degree of the splitting field of an irreducible binomial, Pacific J. Math. 78 (1978) 117–120.

[44] D. Gay, W.Y. Vélez, The torsion group of a radical extension, Pacific J. Math. 92 (1981) 317–327.

[45] C. Greither, D.K. Harrison, A Galois correspondence for radical extensions of fields, J. Pure Appl. Algebra 43 (1986) 257–270.

[46] C. Greither, B. Pareigis, Hopf Galois theory for separable field extensions, J. Algebra 106 (1987) 239–258.

[47] F. Halter-Koch, Über Radikalerweiterungen, Acta Arith. 36 (1980) 43–58.

[48] F. Halter-Koch, Körper über denen alle algebraischen Erweiterungen der Kummerschen Theorie genügen, J. Algebra 64 (1980) 391–398.

[49] H. Hasse, Zahlentheorie, Akademie-Verlag, Berlin, 1963.

[50] H. Hasse, Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper, Teil II: Reziprozitätsgesetz, Physica-Verlag, Würzburg, 1965.

[51] E. Hecke, Eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen (Zweite Mitteilung), Math. Z. 4 (1920) 11–51.

[52] E. Hecke, Vorlesungen über die Theorie der algebraischen Zahlen, Chelsea, New York, 1948.

[53] I. Kaplansky, Fields and Rings, Univ. Chicago Press, Chicago, IL, 1972.

[54] G. Karpilovsky, Topics in Field Theory, North-Holland, Amsterdam, 1989.

[55] M. Kneser, Lineare Abhängigkeit von Wurzeln, Acta Arith. 26 (1975) 307–308.

[56] P. Lam-Estrada, F. Barrera-Mora, G.D. Villa-Salvador, On Kneser extensions, J. Algebra 201 (1998) 703–717.

[57] S. Lang, Algebra, Addison–Wesley, Reading, MA, 1965.

[58] H. Leptin, Ein Darstellungssatz für kompakte total unzusammenhängende Gruppen, Arch. Math. (Basel) 6 (1955) 371–373.

[59] A. Masuoka, Cogalois theory for field extensions, J. Math. Soc. Japan 41 (1989) 577–592.

[60] W. May, Multiplicative groups under field extensions, Canad. J. Math. 31 (1979) 436–440.

[61] S. Montgomery, Hopf Algebras and Their Actions on Rings, CBMS Reg. Conference Ser. Math., vol. 82, Amer. Math. Soc., Providence, RI, 1993.

[62] L.J. Mordell, On the linear independence of algebraic numbers, Pacific J. Math. 3 (1953) 625–630.

[63] J. Neukirch, Algebraische Zahlentheorie, Springer-Verlag, Berlin, 1992.

[64] M. Norris, W.Y. Vélez, Structure theorems for radical extensions of fields, Acta Arith. 38 (1980) 111–115.

[65] P. Ribenboim, Algebraic Numbers, Wiley–Interscience, New York, 1972.

[66] L. Ribes, P. Zalesskii, Profinite Groups, Springer-Verlag, Berlin, 2000.

[67] I. Richards, An application of Galois theory to elementary arithmetic, Adv. Math. 13 (1974) 268–273.

[68] A. Schinzel, On linear dependence of roots, Acta Arith. 28 (1975) 161–175.

[69] A. Schinzel, Abelian binomials, power residues, and exponential congruences, Acta Arith. 32 (1977) 245–274.

[70] A. Schinzel, Selected Topics on Polynomials, Ann Arbor, The Univ. Michigan Press, 1982.

[71] R. Schmidt, Subgroup Lattices of Groups, de Gruyter, Berlin, 1994.

[72] J.-P. Serre, Cohomologie Galoisienne, Lecture Notes in Math., vol. 5, Springer-Verlag, Berlin, 1964.

[73] C.L. Siegel, Algebraische Abhängigkeit von Wurzeln, Acta Arith. 21 (1972) 59–64.

[74] D. Ştefan, Cogalois extensions via strongly graded fields, Comm. Algebra 27 (1999) 5687–5702.

[75] B. Stenström, Rings of Quotients, Springer-Verlag, Berlin, 1975.

[76] M. Suzuki, Structure of a Group and the Structure of its Lattice of Subgroups, Ergeb. Math. Grenzgeb., vol. 10, Springer-Verlag, Berlin, 1956.

[77] M.E. Sweedler, Hopf Algebras, Benjamin, New York, 1969.

[78] M. Ţena, Abelian G-Cogalois and cyclotomic field extensions of the rational field, Bull. Soc. Sci. Math. Roumaine 42 (90) (1999) 151–158.

[79] H.D. Ursell, The degrees of radical extensions, Canad. Math. Bull. 17 (1974) 615–617.

[80] W.Y. Vélez, On normal binomials, Acta Arith. 36 (1980) 113–124.

[81] W.Y. Vélez, Correction to the paper "Structure theorems for radical extensions of fields", Acta Arith. 42 (1983) 427–428.

[82] W.Y. Vélez, Several results on radical extensions, Arch. Math. (Basel) 45 (1985) 342–349.

[83] J.S. Wilson, Profinite Groups, Clarendon Press, Oxford, 1998.

[84] J.-P. Zhou, On the degree of extensions generated by finitely many algebraic numbers, J. Number Theory 34 (1990) 133–141.

Section 2A
Category Theory

This page intentionally left blank

Operads and PROPs

Martin Markl[1]

Mathematical Institute of the Academy, Žitná 25, 115 67 *Prague* 1, *Czech Republic*
E-mail: *markl@math.cas.cz*

Dedicated to the memory of Jakub Jan Ryba (1765–1815)

Contents

Abstract

We review definitions and basic properties of operads, PROPs and algebras over these structures.

[1] Supported by the grant GA ČR 201/05/2117 and by the Academy of Sciences of the Czech Republic, Institutional Research Plan No. AV0Z10190503.

HANDBOOK OF ALGEBRA, VOL. 5
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(07)05002-4

This page intentionally left blank

Operads involve an abstraction of the family $\{Map(X^n, X)\}_{n\geqslant 0}$ of composable functions of several variables together with an action of permutations of variables. As such, they were originally studied as a tool in homotopy theory, specifically for iterated loop spaces and homotopy invariant structures, but the theory of operads has recently received new inspiration from homological algebra, category theory, algebraic geometry and mathematical physics. The name *operad* and the formal definition appear first in the early 1970s in J.P. May's book [86], but a year or more earlier, M. Boardman and R. Vogt [9] described the same concept under the name categories of *operators in standard form*, inspired by the PROPs and PACTs of Adams and Mac Lane [67]. As pointed out in [62], also Lambek's definition of multicategory [60] (late 1960s) was almost equivalent to what is called today a colored or many-sorted operad. Another important precursor was the associahedron K that appeared in J.D. Stasheff's 1963 paper [106] on homotopy associativity of H-spaces. We do not, however, aspire to write an account on the *history* of operads and their applications here – we refer to the introduction of [83], to [89], [114], or to the report [105] instead.

Operads are important not in and of themselves but, like PROPs, through their representations, more commonly called *algebras* over operads or *operad algebras*. If an operad is thought of as a kind of algebraic theory, then an algebra over an operad is a model of that theory. Algebras over operads involve most of 'classical' algebras (associative, Lie, commutative associative, Poisson, &c.), loop spaces, moduli spaces of algebraic curves, vertex operator algebras, &c. *Colored* or *many-sorted* operads then describe diagrams of homomorphisms of these objects, homotopies between homomorphisms, modules, &c.

PROPs generalize operads in the sense that they admit operations with several inputs and *several* outputs. Therefore various bialgebras (associative, Lie, infinitesimal) are PROPic algebras. PROPs were also used to encode 'profiles' of structures in formal differential geometry [92,93].

By the *renaissance* of operads we mean the first half of the nineties of the last century when several papers which stimulated the rebirth of interest in operads appeared [31,34, 41,45,47,49,72]. Let us mention the most important new ideas that emerged during this period.

First of all, operads were recognized as the underlying combinatorial structure of the moduli space of stable algebraic curves in complex geometry, and of compactifications of configuration spaces of points of affine spaces in real geometry. In mathematical physics, several very important concepts such as vertex operator algebras or various string theories were interpreted as algebras over operads. On the algebraic side, the notion of *Koszulness* of operads was introduced and studied, and the relation between resolutions of operads and deformations of their algebras was recognized. See [63] for an autochthonous account of the renaissance. Other papers that later became influential then followed in a rapid succession [30,33,32,73,75].

Let us list some of the most important outcomes of the renaissance of operads. The choice of the material for this incomplete catalog has been of course influenced by the author's personal expertise and inclination towards algebra, geometry and topics that are commonly called mathematical physics. We will therefore not be able to pay as much attention to other aspects of operads, such as topology, category theory and homotopy theory, as they deserve.

Complex geometry. Applications involve moduli spaces of stable complex algebraic curves of genus zero [34], enumerative geometry, Frobenius manifolds, quantum cohomology and cohomological field theory [55,71]. The moduli space of genus zero curves exhibits an additional symmetry that leads to a generalization called *cyclic* operads [32]. *Modular* operads [33] then describe the combinatorial structure of the space of curves of arbitrary genus.

Real geometry. Compactifications of configuration spaces of points in real smooth manifolds are operads in the category of smooth manifolds with corners or modules over these operads [76]. This fact is crucial for the theory of configuration spaces with summable labels [96]. The cacti operad [117] lies behind the Chas–Sullivan product on the free loop space of a smooth manifold [13], see also [14]. Tamarkin's proof of the formality of Hochschild cochains of the algebra of functions on smooth manifolds [110] explained in [40] uses obstruction theory for operad algebras and the affirmative answer to the Deligne conjecture [17,56].

Mathematical physics. The formality mentioned in the previous item implies the existence of the deformation quantization of Poisson manifolds [54]. We must not forget to mention the operadic interpretation of vertex operator algebras [46], string theory [49] and Connes–Kreimer's approach to renormalization [15]. Operads and multicategories are important also for Beilinson–Drinfeld's theory of chiral algebras [6].

Algebra. Operadic cohomology [1,26,31,34,83] provides a uniform treatment of all 'classical' cohomology theories, such as the Hochschild cohomology of associative algebras, Harrison cohomology of associative commutative algebras, Chevalley–Eilenberg cohomology of Lie algebras, &c. Minimal models for operads [75] offer a conceptual understanding of strong homotopy algebras, their homomorphisms and homotopy invariance [80]. Operads serve as a natural language for various types of 'multialgebras' [64,65]. Relation between Koszulness of operads and properties of posets was studied in [27]. Also the concept of the operadic distributive law turned out to be useful [26,74].

Model structures. It turned out [8,31,39,100] that algebras over a reasonable (possibly colored) operad form a model category that generalizes the classical model structures of the categories of dg commutative associative algebras and dg Lie algebras [95,107]. Operads, in a reasonable monoidal model category, themselves form a model category [7,31] such that algebras over cofibrant operads are homotopy invariant, see also [104]. Minimal operads mentioned in the previous item are particular cases of cofibrant dg-operads and the classical W-construction [9] is a functorial cofibrant replacement in the category of topological operads [116]. The above model structures are important for various constructions in the homology theory of (free or based) loop spaces [14,43] and formulations of 'higher' Deligne conjecture [44].

Topology. Operads as gadgets organizing homotopy coherent structures are important in the brave new algebra approach to topological Hochschild cohomology and algebraic

K-theory, see [22,23,90,115], or [21] for a historical background. A description of a localized category of integral and p-adic homotopy types by E_∞-operads was given in [69, 70]. An operadic approach to partial algebras and their completions was applied in [58] to mixed Tate motives over the rationals. See also an overview [88].

Category theory. Operads and multicategories were used as a language in which to propose a definition of weak ω-category [3–5,61]. Operads themselves can be viewed as special kinds of algebraic theory (as can multicategories, if one allows many-sorted theories), see [85]. There are also 'categorical' generalizations of operads, e.g. the globular operads of [2] and T-categories of [11]. An interesting presentation of PROP-like structures in enriched monoidal categories can be found in [91].

Graph complexes. Each cyclic operad $\mathcal{P}$ determines a graph complex [33,77]. As observed earlier by M. Kontsevich [52], these graph complexes are, for some specific choices of $\mathcal{P}$, closely related to some very interesting objects such as moduli spaces of Riemann surfaces, automorphisms of free groups or primitives in the homology of certain infinite-dimensional Lie algebras, see also [83, II.5.5]. In the same vein, complexes of directed graphs are related to PROPs [84,111–113] and directed graphs with back-in-time edges are tied to wheeled PROPs introduced in [93].

Deformation theory and homotopy invariant structures in algebra. A concept of homotopy invariant structures in algebra parallel to the classical one in topology [9,10] was developed in [80]. It was explained in [56,73,79] how cofibrant resolutions of operads or PROPs determine a cohomology theory governing deformations of related algebras. In [81], deformations were identified with solutions of the Maurer–Cartan equation of a certain strongly homotopy Lie algebra constructed in a very explicit way from a cofibrant resolution of the underlying operad or PROP.

Terminology. As we already observed, operads are abstractions of families of composable functions. Given functions $f : X^{\times n} \to X$ and $g_i : X^{\times k_i} \to X$, $1 \leqslant i \leqslant n$, one may consider the simultaneous composition

$$f(g_1, \dots, g_n) : X^{\times(k_1+\cdots+k_n)} \to X. \tag{I}$$

One may also consider, for $f : X^{\times n} \to X$, $g : X^{\times m} \to X$ and $1 \leqslant i \leqslant n$, the individual compositions

$$f(id, \dots, id, g, id, \dots, id) : X^{\times(m+m-1)} \to X, \tag{II}$$

with g at the i-th place. While May's original definition of an operad [86] was an abstraction of type (I) compositions, there exists an alternative approach based on type (II) compositions. This second point of view was formalized in the 1963 papers by Gerstenhaber [29] and Stasheff [106]. A definition that included the symmetric group action was formulated much later in the author's paper [75] in which the two approaches were also compared.

In the presence of operadic units, these approaches are equivalent. There are, however, situations where one needs also non-unital versions, and then the two approaches lead to *different* structures – a non-unital structure of the second type always determines a non-unital structure of the first type, but *not vice versa*! It turns out that more common are non-unital structures of the second type; they describe, for example, the underlying combinatorial structure of the moduli space of stable complex curves.

We will therefore call the non-unital versions of the first type of operads *non-unital May's operads*, while the second version simply *non-unital operads*. We opted for this terminology, which was used already in the first version of [75], after a long hesitation, being aware that it might not be universally welcome. Note that non-unital operads are sometimes called *(Markl's) pseudo-operads* [75,83].

Outline of the chapter. In the first three sections we review the basic definitions of (unital and non-unital) operads and operad algebras, and give examples that illustrate these notions. Section 4 describes free operads and their relation to rooted trees. In Section 5 we explain that operads can be defined as algebras over the monad of rooted trees. In the following two sections we show that, replacing rooted trees by other types of trees, one obtains two important generalizations – cyclic and modular operads. In the last two sections, PROPs and their versions are recalled; this article is the first expository text where these structures are systematically treated.

Sections 1–3 are based on the classical book [86] by J.P. May and the author's article [75]. Sections 4–7 follow the seminal paper [34] by V. Ginzburg and M.M. Kapranov, and papers [32,33] by E. Getzler and M.M. Kapranov. The last two sections are based on the preprint [84] of A.A. Voronov and the author, and on an e-mail message [53] from M. Kontsevich. We were also influenced by T. Leinster's concept of biased versus unbiased definitions [61]. At some places, our exposition follows the monograph [83] by S. Shnider, J.D. Stasheff and the author.

1. Operads

Although operads, operad algebras and most of related structures can be defined in an arbitrary symmetric monoidal category with countable coproducts, we decided to follow the choice of [58] and formulate precise definitions only for the category $\mathtt{Mod}_{\mathbf{k}} = (\mathtt{Mod}_{\mathbf{k}}, \otimes)$ of modules over a commutative unital ring $\mathbf{k}$, with the monoidal structure given by the tensor product $\otimes = \otimes_{\mathbf{k}}$ over $\mathbf{k}$. The reason for such a decision was to give, in Section 4, a clean construction of free operads. In a general monoidal category, this construction involves the unordered $\odot$-product [83, Definition II.1.38] so the free operad is then a double colimit, see [83, Section II.1.9]. Our choice also allows us to write formulas involving maps in terms of elements, which is sometimes a welcome simplification. We believe that the reader can easily reformulate our definitions for the case of other monoidal categories or consult [83,87].

As usual, the symmetric group Σ_n is, for $n \geqslant 1$, the automorphism group of the set $\{1, \dots, n\}$, with the group multiplication given by the composition, $\sigma' \cdot \sigma'' := \sigma'(\sigma'')$, for $\sigma', \sigma'' \in \Sigma_n$. Let $\mathbf{k}[\Sigma_n]$ denote the $\mathbf{k}$-group ring of Σ_n.

Definition 1 *(May operad).* An *operad* in the category of $\mathbf{k}$-modules is a collection $\mathcal{P} = \{\mathcal{P}(n)\}_{n \geqslant 0}$ of right $\mathbf{k}[\Sigma_n]$-modules, together with $\mathbf{k}$-linear maps (operadic compositions)

$$\gamma : \mathcal{P}(n) \otimes \mathcal{P}(k_1) \otimes \cdots \otimes \mathcal{P}(k_n) \to \mathcal{P}(k_1 + \cdots + k_n), \tag{1}$$

for $n \geqslant 1$ and $k_1, \dots, k_n \geqslant 0$, and a unit map $\eta : \mathbf{k} \to \mathcal{P}(1)$. These data fulfill the following axioms.

Associativity. Let $n \geqslant 1$ and let $m_1, \dots, m_n$ and $k_1, \dots, k_m$, where $m := m_1 + \cdots + m_n$, be non-negative integers. Then the following diagram, in which $g_s := m_1 + \cdots + m_{s-1}$ and $h_s = k_{g_s+1} + \cdots + k_{g_{s+1}}$, for $1 \leqslant s \leqslant n$, commutes.[1]

$$\begin{tikzcd}
\Big(\mathcal{P}(n) \otimes \bigotimes_{s=1}^{n} \mathcal{P}(m_s)\Big) \otimes \bigotimes_{r=1}^{m} \mathcal{P}(k_r) \arrow[r, "\gamma\otimes id"] \arrow[dd, "\text{shuffle}"'] & \mathcal{P}(m) \otimes \bigotimes_{r=1}^{m} \mathcal{P}(k_r) \arrow[d, "\gamma"] \\
 & \mathcal{P}(k_1 + \cdots + k_m) \\
\mathcal{P}(n) \otimes \bigotimes_{s=1}^{n} \Big(\mathcal{P}(m_s) \otimes \bigotimes_{q=1}^{m_s} \mathcal{P}(k_{g_s+q})\Big) \arrow[r, "id\otimes(\bigotimes_{s=1}^{n} \gamma)"] & \mathcal{P}(n) \otimes \bigotimes_{s=1}^{n} \mathcal{P}(h_s) \arrow[u, "\gamma"']
\end{tikzcd}$$

Equivariance. Let $n \geqslant 1$, let $k_1, \dots, k_n$ be non-negative integers and let $\sigma \in \Sigma_n$, $\tau_1 \in \Sigma_{k_1}, \dots, \tau_n \in \Sigma_{k_n}$ be permutations. Let $\sigma(k_1, \dots, k_n) \in \Sigma_{k_1+\cdots+k_n}$ denote the permutation that permutes the n blocks $(1, \dots, k_1), \dots, (k_1+\cdots+k_{n-1}+1, \dots, k_1+\cdots+k_n)$ as σ permutes $(1, \dots, n)$ and let $\tau_1 \oplus \cdots \oplus \tau_n \in \Sigma_{k_1+\cdots+k_n}$ be the block sum of permutations. Then the following diagrams commute.

$$\begin{tikzcd}
\mathcal{P}(n) \otimes \mathcal{P}(k_1) \otimes \cdots \otimes \mathcal{P}(k_n) \arrow[r, "\sigma\otimes\sigma^{-1}"] \arrow[d, "\gamma"'] & \mathcal{P}(n) \otimes \mathcal{P}(k_{\sigma(1)}) \otimes \cdots \otimes \mathcal{P}(k_{\sigma(n)}) \arrow[d, "\gamma"] \\
\mathcal{P}(k_1 + \cdots + k_n) \arrow[r, "\sigma(k_{\sigma(1)},\dots,k_{\sigma(n)})"] & \mathcal{P}(k_{\sigma(1)} + \cdots + k_{\sigma(n)})
\end{tikzcd}$$

$$\begin{tikzcd}
\mathcal{P}(n) \otimes \mathcal{P}(k_1) \otimes \cdots \otimes \mathcal{P}(k_n) \arrow[r, "id\otimes\tau_1\otimes\cdots\otimes\tau_n"] \arrow[d, "\gamma"'] & \mathcal{P}(n) \otimes \mathcal{P}(k_1) \otimes \cdots \otimes \mathcal{P}(k_n) \arrow[d, "\gamma"] \\
\mathcal{P}(k_1 + \cdots + k_n) \arrow[r, "\tau_1\oplus\cdots\oplus\tau_n"] & \mathcal{P}(k_1 + \cdots + k_n)
\end{tikzcd}$$

Unitality. For each $n \geqslant 1$, the following diagrams commute.

$$\begin{tikzcd}
\mathcal{P}(n) \otimes \mathbf{k}^{\otimes n} \arrow[r, "\cong"] \arrow[d, "id\otimes\eta^{\otimes n}"'] & \mathcal{P}(n) \\
\mathcal{P}(n) \otimes \mathcal{P}(1)^{\otimes n} \arrow[ur, "\gamma"] &
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbf{k} \otimes \mathcal{P}(n) \arrow[r, "\cong"] \arrow[d, "\eta\otimes id"'] & \mathcal{P}(n) \\
\mathcal{P}(1) \otimes \mathcal{P}(n) \arrow[ur, "\gamma"] &
\end{tikzcd}$$

[1] The meaning of "shuffle" in the diagram below is explained in a remark following this definition.

In the associativity diagram above, "shuffle" means switching the various factors in the tensor product

$$\mathcal{P}(n) \otimes \big(\mathcal{P}(m_1) \otimes \cdots \otimes \mathcal{P}(m_n)\big) \otimes \big(\mathcal{P}(k_1) \otimes \cdots \otimes \mathcal{P}(k_m)\big)$$

suitably. For instance, if $n = 2$, "shuffle" is the natural isomorphism of

$$\mathcal{P}(2) \otimes \big(\mathcal{P}(m_1) \otimes \mathcal{P}(m_2)\big) \otimes \big(\mathcal{P}(k_1) \otimes \cdots \otimes \mathcal{P}(k_{m_1+m_2})\big)$$

with

$$\mathcal{P}(2) \otimes \big(\mathcal{P}(m_1) \otimes \mathcal{P}(k_1) \otimes \cdots \otimes \mathcal{P}(k_{m_1})\big) \otimes \big(\mathcal{P}(m_2) \otimes \mathcal{P}(k_{m_1+1}) \otimes \cdots \otimes \mathcal{P}(k_{m_1+m_2})\big).$$

A straightforward modification of the above definition makes sense in any symmetric monoidal category $(\mathcal{M}, \odot, \mathbf{1})$ such as the categories of differential graded modules, simplicial sets, topological spaces, &c., see [83, Definition II.1.4] or [87, Definition 1]. We then speak about *differential graded* operads, *simplicial* operads, *topological* operads, &c.

EXAMPLE 2. All properties axiomatized by Definition 1 can be read from the *endomorphism operad* $\mathcal{E}\mathrm{nd}_V = \{\mathcal{E}\mathrm{nd}_V(n)\}_{n\geqslant 0}$ of a $\mathbf{k}$-module V. It is defined by setting $\mathcal{E}\mathrm{nd}_V(n)$ to be the space of $\mathbf{k}$-linear maps $V^{\otimes n} \to V$. The operadic composition of $f \in \mathcal{E}\mathrm{nd}_V(n)$ with $g_1 \in \mathcal{E}\mathrm{nd}_V(k_1), \dots, g_n \in \mathcal{E}\mathrm{nd}_V(k_n)$ is given by the usual composition of multilinear maps as

$$\gamma(f, g_1, \dots, g_n) := f(g_1 \otimes \cdots \otimes g_n),$$

the symmetric group acts by[2]

$$\gamma\sigma(f, g_1, \dots, g_n) := f(g_{\sigma^{-1}(1)} \otimes \cdots \otimes g_{\sigma^{-1}(n)}), \quad \sigma \in \Sigma_n,$$

and the unit map $\eta : \mathbf{k} \to \mathcal{E}\mathrm{nd}_V(1)$ is given by $\eta(1) := id_V : V \to V$. The endomorphism operad can be constructed over an object of an arbitrary symmetric monoidal category with an internal hom-functor, as it was done in [83, Definition II.1.7].

One often considers operads $\mathcal{A}$ such that $\mathcal{A}(0) = 0$ (the trivial $\mathbf{k}$-module). We will indicate that $\mathcal{A}$ is of this type by writing $\mathcal{A} = \{\mathcal{A}(n)\}_{n\geqslant 1}$.

EXAMPLE 3. Let us denote by $\mathcal{A}ss = \{\mathcal{A}ss(n)\}_{n\geqslant 1}$ the operad with $\mathcal{A}ss(n) := \mathbf{k}[\Sigma_n]$, $n \geqslant 1$, and the operadic composition defined as follows. Let $id_n \in \Sigma_n$, $id_{k_1} \in \Sigma_{k_1}, \dots, id_{k_n} \in \Sigma_{k_n}$ be the identity permutations. Then

$$\gamma(id_n, id_{k_1}, \dots, id_{k_n}) := id_{k_1+\cdots+k_n} \in \Sigma_{k_1+\cdots+k_n}.$$

The above formula determines $\gamma(\sigma, \tau_1, \dots, \tau_n)$ for general $\sigma \in \Sigma_n$, $\tau_1 \in \Sigma_{k_1}, \dots, \tau_n \in \Sigma_{k_n}$ by the equivariance axiom. The unit map $\eta : \mathbf{k} \to \mathcal{A}ss(1)$ is given by $\eta(1) := id_1$.

[2] I.e. the action of Σ_n on a linear map $f \in \mathrm{Lin}(V^{\otimes n}, V)$ is given by $f(\sigma)(v_1 \otimes \cdots \otimes v_n) := f(v_{\sigma^{-1}(1)} \otimes \cdots \otimes v_{\sigma^{-1}(n)})$.

EXAMPLE 4. Let us give an example of a topological operad. For $k \geqslant 1$, the *little k-discs operad* $\mathcal{D}_k = \{\mathcal{D}_k(n)\}_{n \geqslant 0}$ is defined as follows [83, Section II.4.1]. Let

$$\mathbb{D}^k := \left\{(x_1, \dots, x_k) \in \mathbb{R}^k;\ x_1^2 + \cdots + x_k^2 \leqslant 1\right\}$$

be the standard closed disc in $\mathbb{R}^k$. A little k-disc is then a linear embedding $d : \mathbb{D}^k \hookrightarrow \mathbb{D}^k$ which is the restriction of a linear map $\mathbb{R}^k \to \mathbb{R}^k$ given by the composition of a translation and a contraction. The n-th space $\mathcal{D}_k(n)$ of the little k-disc operad is the space of all n-tuples $(d_1, \dots, d_n)$ of little k-discs such that the images of $d_1, \dots, d_n$ have mutually disjoint interiors. The operad structure is obvious – the symmetric group Σ_n acts on $\mathcal{D}_k(n)$ by permuting the labels of the little discs and the structure map γ is given by composition of embeddings. The unit is the identity embedding $id : \mathbb{D}^k \hookrightarrow \mathbb{D}^k$.

EXAMPLE 5. The collection of normalized singular chains $C_*(\mathcal{T}) = \{C_*(\mathcal{T}(n))\}_{n \geqslant 0}$ of a topological operad $\mathcal{T} = \{\mathcal{T}(n)\}_{n \geqslant 0}$ is an operad in the category of differential graded $\mathbb{Z}$-modules. For a ring R, the singular homology $H_*(\mathcal{T}(n); R) = H_*(C_*(\mathcal{T}(n)) \otimes_{\mathbb{Z}} R)$ forms an operad $H_*(\mathcal{T}; R)$ in the category of graded R-modules, see [58, Section I.5] for details.

DEFINITION 6. Let $\mathcal{P} = \{\mathcal{P}(n)\}_{n \geqslant 0}$ and $\mathcal{Q} = \{\mathcal{Q}(n)\}_{n \geqslant 0}$ be two operads. A *homomorphism* $f : \mathcal{P} \to \mathcal{Q}$ is a sequence $f = \{f(n) : \mathcal{P}(n) \to \mathcal{Q}(n)\}_{n \geqslant 0}$ of equivariant maps which commute with the operadic compositions and preserve the units.

An operad $\mathcal{R} = \{\mathcal{R}(n)\}_{n \geqslant 0}$ is a *suboperad* of $\mathcal{P}$ if $\mathcal{R}(n)$ is, for each $n \geqslant 0$, a Σ_n-submodule of $\mathcal{P}(n)$ and if all structure operations of $\mathcal{R}$ are the restrictions of those of $\mathcal{P}$. Finally, an *ideal* in the operad $\mathcal{P}$ is the collection $\mathcal{I} = \{\mathcal{I}(n)\}_{n \geqslant 0}$ of Σ_n-invariant subspaces $\mathcal{I}(n) \subset \mathcal{P}(n)$ such that

$$\gamma_{\mathcal{P}}(f \otimes g_1 \otimes \cdots \otimes g_n) \in \mathcal{I}(k_1 + \cdots + k_n)$$

if either $f \in \mathcal{I}(n)$ or $g_i \in \mathcal{I}(k_i)$ for some $1 \leqslant i \leqslant n$.

EXAMPLE 7. Given an operad $\mathcal{P} = \{\mathcal{P}(n)\}_{n \geqslant 0}$, let $\widehat{\mathcal{P}} = \{\widehat{\mathcal{P}}(n)\}_{n \geqslant 0}$ be the collection defined by $\widehat{\mathcal{P}}(n) := \mathcal{P}(n)$ for $n \geqslant 1$ and $\widehat{\mathcal{P}}(0) := 0$. Then $\widehat{\mathcal{P}}$ is a suboperad of $\mathcal{P}$. The correspondence $\mathcal{P} \mapsto \widehat{\mathcal{P}}$ is a full embedding of the category of operads $\mathcal{P}$ with $\mathcal{P}(0) \cong \mathbf{k}$ into the category of operads $\mathcal{A}$ with $\mathcal{A}(0) = 0$. Operads satisfying $\mathcal{P}(0) \cong \mathbf{k}$ have been called *unital* while operads with $\mathcal{A}(0) = 0$ *non-unital* operads. We will not use this terminology because non-unital operads will mean something different in this chapter, see Section 2.

An example of an operad $\mathcal{A}$ which is not of the form $\widehat{\mathcal{P}}$ for some operad $\mathcal{P}$ with $\mathcal{P}(0) \cong \mathbf{k}$ can be constructed as follows. Observe first that operads $\mathcal{P}$ with the property that

$$\mathcal{P}(0) \cong \mathbf{k} \quad \text{and} \quad \mathcal{P}(n) = 0 \quad \text{for } n \geqslant 2$$

are the same as augmented associative algebras. Indeed, the space $\mathcal{P}(1)$ with the operation $\circ_1 : \mathcal{P}(1) \otimes \mathcal{P}(1) \to \mathcal{P}(1)$ is clearly a unital associative algebra, augmented by the composition

$$\mathcal{P}(1) \xrightarrow{\cong} \mathcal{P}(1) \otimes \mathbf{k} \xrightarrow{\cong} \mathcal{P}(1) \otimes \mathcal{P}(0) \xrightarrow{\circ_1} \mathcal{P}(0) \cong \mathbf{k}.$$

Now take an arbitrary unital associative algebra A and define the operad $\mathcal{A} = \{\mathcal{A}(n)\}_{n\geqslant 1}$ by

$$\mathcal{A}(n) := \begin{cases} A, & \text{for } n = 1,\\ 0, & \text{for } n \neq 1,\end{cases}$$

with $\circ_1 : \mathcal{A}(1) \otimes \mathcal{A}(1) \to \mathcal{A}(1)$ the multiplication of A. It follows from the above considerations that $\mathcal{A} = \widehat{\mathcal{P}}$ for some operad $\mathcal{P}$ with $\mathcal{P}(0) \cong \mathbf{k}$ if and only if A admits an augmentation. Therefore any unital associative algebra that does not admit an augmentation produces the desired example.

Example 8. Kernels, images, &c., of homomorphisms between operads in the category of $\mathbf{k}$-modules are defined componentwise. For example, if $f : \mathcal{P} \to \mathcal{Q}$ is such homomorphism, then $\mathrm{Ker}(f) = \{\mathrm{Ker}(f)(n)\}_{n\geqslant 0}$ is the collection with

$$\mathrm{Ker}(f)(n) := \mathrm{Ker}\big(f(n) : \mathcal{P}(n) \to \mathcal{Q}(n)\big), \quad n \geqslant 0.$$

It is clear that $\mathrm{Ker}(f)$ is an ideal in $\mathcal{P}$.

Also quotients are defined componentwise. If $\mathcal{I}$ is an ideal in $\mathcal{P}$, then the collection $\mathcal{P}/\mathcal{I} = \{(\mathcal{P}/\mathcal{I})(n)\}_{n\geqslant 0}$ with $(\mathcal{P}/\mathcal{I})(n) := \mathcal{P}(n)/\mathcal{I}(n)$ for $n \geqslant 0$, has a natural operad structure induced by the structure of $\mathcal{P}$. The canonical projection $\mathcal{P} \to \mathcal{P}/\mathcal{I}$ has the expected universal property. The kernel of this projection equals $\mathcal{I}$.

Sometimes it suffices to consider operads without the symmetric group action. This notion is formalized by:

Definition 9 *(May non-Σ operad)*. A *non-Σ operad* in the category of $\mathbf{k}$-modules is a collection $\underline{\mathcal{P}} = \{\underline{\mathcal{P}}(n)\}_{n\geqslant 0}$ of $\mathbf{k}$-modules, together with operadic compositions

$$\underline{\gamma} : \underline{\mathcal{P}}(n) \otimes \underline{\mathcal{P}}(k_1) \otimes \cdots \otimes \underline{\mathcal{P}}(k_n) \to \underline{\mathcal{P}}(k_1 + \cdots + k_n),$$

for $n \geqslant 1$ and $k_1, \ldots, k_n \geqslant 0$, and a unit map $\underline{\eta} : \mathbf{k} \to \underline{\mathcal{P}}(1)$ that fulfill the associativity and unitality axioms of Definition 1.

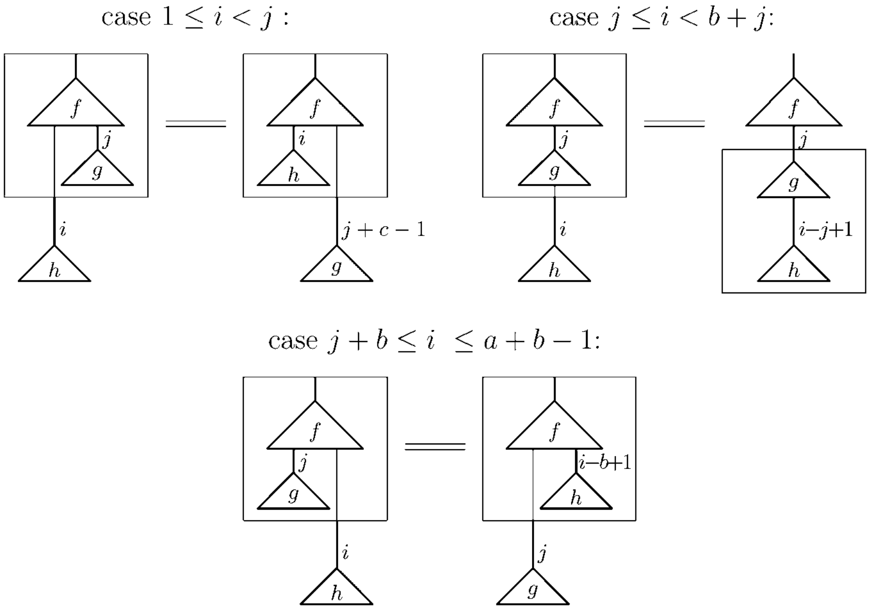
Each operad can be considered as a non-Σ operad by forgetting the Σ_n-actions. On the other hand, given a non-Σ operad $\underline{\mathcal{P}}$, there is an associated operad $\Sigma[\underline{\mathcal{P}}]$ with $\Sigma[\underline{\mathcal{P}}](n) := \underline{\mathcal{P}}(n) \otimes \mathbf{k}[\Sigma_n]$, $n \geqslant 0$, with the structure operations induced by the structure operations of $\underline{\mathcal{P}}$. Operads of this form are sometimes called *regular* operads.

Example 10. Consider the operad $\mathcal{C}om = \{\mathcal{C}om(n)\}_{n\geqslant 1}$ such that $\mathcal{C}om(n) := \mathbf{k}$ with the trivial Σ_n-action, $n \geqslant 1$, and the operadic compositions (1) given by the canonical identifications

$$\mathcal{C}om(n) \otimes \mathcal{C}om(k_1) \otimes \cdots \otimes \mathcal{C}om(k_n) \cong \mathbf{k}^{\otimes(n+1)} \xrightarrow{\cong} \mathbf{k} \cong \mathcal{C}om(k_1 + \cdots + k_n).$$

The operad $\mathcal{C}om$ is obviously not regular. Observe also that $\mathcal{C}om \cong \widehat{\mathcal{E}nd_{\mathbf{k}}}$, where $\widehat{\mathcal{E}nd_{\mathbf{k}}}$ is the endomorphism operad of the ground ring without the initial component, see Example 7 for the notation.

Let $\underline{\mathcal{A}ss}$ denote the operad $\mathcal{C}om$ considered as a non-Σ operad. Its symmetrization $\Sigma[\underline{\mathcal{A}ss}]$ then equals the operad $\mathcal{A}ss$ introduced in Example 3.


case $1 \leq i < j$:

case $j \leq i < b + j$:

case $j + b \leq i \ \leq a + b - 1$:

Fig. 1. Flow charts explaining the associativity in a Markl operad.

As we already observed, there is an alternative approach to operads. For the purposes of comparison, in the rest of this section and in the following section we will refer to operads viewed from this alternative perspective as to Markl operads. See also the remarks on the terminology in the introduction.

DEFINITION 11. A *Markl operad* in the category of **k**-modules is a collection $\mathcal{S} = \{\mathcal{S}(n)\}_{n \geqslant 0}$ of right $\mathbf{k}[\Sigma_n]$-modules, together with **k**-linear maps ($\circ_i$-compositions)

$$\circ_i : \mathcal{S}(m) \otimes \mathcal{S}(n) \to \mathcal{S}(m + n - 1),$$

for $1 \leqslant i \leqslant m$ and $n \geqslant 0$. These data fulfill the following axioms.

Associativity. For each $1 \leqslant j \leqslant a$; $1 \leqslant i \leqslant a + b - 1$; $b, c \geqslant 0$, $f \in \mathcal{S}(a)$, $g \in \mathcal{S}(b)$ and $h \in \mathcal{S}(c)$,

$$(f \circ_j g) \circ_i h = \begin{cases} (f \circ_i h) \circ_{j+c-1} g, & \text{for } 1 \leqslant i < j, \\ f \circ_j (g \circ_{i-j+1} h), & \text{for } j \leqslant i < b + j, \\ (f \circ_{i-b+1} h) \circ_j g, & \text{for } j + b \leqslant i \leqslant a + b - 1, \end{cases}$$

see Fig. 1.

Equivariance. For each $1 \leqslant i \leqslant m$, $n \geqslant 0$, $\tau \in \Sigma_m$ and $\sigma \in \Sigma_n$, let $\tau \circ_i \sigma \in \Sigma_{m+n-1}$ be given by inserting the permutation σ at the i-th place in τ. Let $f \in \mathcal{S}(m)$ and $g \in \mathcal{S}(n)$. Then

$$(f\tau) \circ_i (g\sigma) = (f \circ_{\tau(i)} g)(\tau \circ_i \sigma).$$

Unitality. There exists $e \in \mathcal{S}(1)$ such that

$$f \circ_i e = f \quad \text{and} \quad e \circ_1 g = g \tag{2}$$

for each $1 \leqslant i \leqslant m, n \geqslant 0$, $f \in \mathcal{S}(m)$ and $g \in \mathcal{S}(n)$.

EXAMPLE 12. All axioms in Definition 11 can be read from the endomorphism operad $\mathcal{E}nd_V = \{\mathcal{E}nd_V(n)\}_{n \geqslant 0}$ of a $\mathbf{k}$-module V reviewed in Example 2, with $\circ_i$-operations given by

$$f \circ_i g := f\big(id_V^{\otimes i-1} \otimes g \otimes id_V^{\otimes m-i}\big),$$

for $f \in \mathcal{E}nd_V(m)$, $g \in \mathcal{E}nd_V(n)$, $1 \leqslant i \leqslant m$ and $n \geqslant 0$.

The following proposition shows that Definition 1 describes the same objects as Definition 11.

PROPOSITION 13. *The category of May operads is isomorphic to the category of Markl operads.*

PROOF. Given a Markl operad $\mathcal{S} = \{\mathcal{S}(n)\}_{n \geqslant 0}$ as in Definition 11, define a May operad $\mathcal{P} = \mathrm{May}(\mathcal{S})$ by $\mathcal{P}(n) := \mathcal{S}(n)$ for $n \geqslant 0$, with the γ-operations given by

$$\gamma(f, g_1, \dots, g_n) := \big(\dots\big((f \circ_n g_n) \circ_{n-1} g_{n-1}\big)\dots\big) \circ_1 g_1 \tag{3}$$

where $f \in \mathcal{P}(n)$, $g_i \in \mathcal{P}(k_i)$, $1 \leqslant i \leqslant n$, $k_1, \dots, k_n \geqslant 0$. The unit morphism $\eta : \mathbf{k} \to \mathcal{P}(1)$ is defined by $\eta(1) := e$. It is easy to verify that $\mathrm{May}(-)$ extends to a functor from the category of Markl operads the category of May operads.

On the other hand, given a May operad $\mathcal{P}$, one can define a Markl operad $\mathcal{S} = \mathrm{Mar}(\mathcal{P})$ by $\mathcal{S}(n) := \mathcal{P}(n)$ for $n \geqslant 0$, with the $\circ_i$-operations:

$$f \circ_i g := \gamma(f, \underbrace{e, \dots, e}_{i-1}, g, \underbrace{e, \dots, e}_{m-i}), \tag{4}$$

for $f \in \mathcal{S}(m)$, $g \in \mathcal{S}(n)$, $m \geqslant 1$, $n \geqslant 0$, where $e := \eta(1) \in \mathcal{P}(1)$. It is again obvious that $\mathrm{Mar}(-)$ extends to a functor that the functors $\mathrm{May}(-)$ and $\mathrm{Mar}(-)$ are mutually inverse isomorphisms between the category of Markl operads and the category of May operads. This involves, of course, checking that the functors $\mathrm{May}(-)$ and $\mathrm{Mar}(-)$ take the respective associativity, unitality, and equivariance conditions into each other. This is left as a potential exercise for the reader. □

The equivalence between May and Markl operads implies that an operad can be defined by specifying $\circ_i$-operations and a unit. This is sometimes simpler that to define the γ-operations directly, as illustrated by:

EXAMPLE 14. Let Σ be a Riemann sphere, that is, a nonsingular complex projective curve of genus 0. By a puncture or a parametrized hole we mean a point p of Σ together with a holomorphic embedding of the standard closed disc $U = \{z \in \mathbb{C};\ |z| \leqslant 1\}$ to Σ centered at the point. Thus a puncture is a holomorphic embedding $u : \tilde{U} \to \Sigma$, where $\tilde{U} \subset \mathbb{C}$ is an open neighborhood of U and $u(0) = p$. We say that two punctures $u_1 : \tilde{U}_1 \to \Sigma$ and $u_2 : \tilde{U}_2 \to \Sigma$ are disjoint, if

$$u_1(\mathring{U}) \cap u_2(\mathring{U}) = \emptyset,$$

where $\mathring{U} := \{z \in \mathbb{C};\ |z| < 1\}$ is the interior of U.

Let $\widehat{\mathfrak{M}}_0(n)$ be the moduli space of Riemann spheres Σ with $n+1$ disjoint punctures $u_i : \tilde{U}_i \to \Sigma$, $0 \leqslant i \leqslant n$, modulo the action of complex projective automorphisms. The topology of $\widehat{\mathfrak{M}}_0(n)$ is a very subtle thing and we are not going to discuss this issue here; see [46]. The constructions below will be made only 'up to topology'.

Renumbering the holes $u_1, \dots, u_n$ defines on each $\widehat{\mathfrak{M}}_0(n)$ a natural right Σ_n-action and the Σ-module $\widehat{\mathfrak{M}}_0 = \{\widehat{\mathfrak{M}}_0(n)\}_{n \geqslant 0}$ forms a topological operad under sewing Riemannian spheres at punctures. Let us describe this operadic structure using the $\circ_i$-formalism. Thus, let Σ represent an element $x \in \widehat{\mathfrak{M}}_0(m)$ and Δ represent an element $y \in \widehat{\mathfrak{M}}_0(n)$. For $1 \leqslant i \leqslant m$, let $u_i : \tilde{U}_i \to \Sigma$ be the i-th puncture of Σ and let $u_0 : \tilde{U}_0 \to \Delta$ be the 0-th puncture of Δ.

There certainly exists some $0 < r < 1$ such that both $\tilde{U}_0$ and $\tilde{U}_i$ contain the disc $U_{1/r} := \{z \in \mathbb{C};\ |z| < 1/r\}$. Let now $\Sigma_r := \Sigma \setminus u_i(U_r)$ and $\Delta_r := \Delta \setminus u_0(U_r)$. Define finally

$$\Xi := (\Sigma_r \sqcup \Delta_r)/\sim,$$

where the relation $\sim$ is given by

$$\Sigma_r \ni u_i(\xi) \sim u_0(1/\xi) \in \Delta_r,$$

for $r < |\xi| < 1/r$. It is immediate to see that Ξ is a well-defined punctured Riemann sphere, with $n+m$ punctures induced in the obvious manner from those of Σ and Δ, and that the class of the punctured surface Ξ in the moduli space $\widehat{\mathfrak{M}}_0(m+n-1)$ does not depend on the representatives Σ, Δ and on r. We define $x \circ_i y$ to be the class of Ξ.

The unit $e \in \widehat{\mathfrak{M}}_0(1)$ can be defined as follows. Let $\mathbb{CP}^1$ be the complex projective line with homogeneous coordinates $[z, w]$, $z, w \in \mathbb{C}$, [118, Example I.1.6]. Let $0 := [0,1] \in \mathbb{CP}^1$ and $\infty := [1,0] \in \mathbb{CP}^1$. Recall that we have two canonical isomorphisms $p_\infty : \mathbb{CP}^1 \setminus \infty \to \mathbb{C}$ and $p_0 : \mathbb{CP}^1 \setminus 0 \to \mathbb{C}$ given by

$$p_\infty\big([z,w]\big) := z/w \quad \text{and} \quad p_0\big([z,w]\big) := w/z.$$

Then $p_\infty^{-1} : \mathbb{C} \to \mathbb{CP}^1$ (respectively $p_0^{-1} : \mathbb{C} \to \mathbb{CP}^1$) is a puncture at 0 (respectively at ∞). We define $e \in \widehat{\mathfrak{M}}_0(1)$ to be the class of $(\mathbb{CP}^1, p_0^{-1}, p_\infty^{-1})$.

It is not hard to verify that the above constructions make the collection $\widehat{\mathfrak{M}}_0 = \{\widehat{\mathfrak{M}}_0(n)\}_{n \geqslant 0}$ into a Markl operad. By Proposition 13, $\widehat{\mathfrak{M}}_0$ is a also May operad.

In the rest of this chapter, we will consider May and Markl operads as two versions of the same object which we will call simply a (unital) operad.

2. Non-unital operads

It turns out that the combinatorial structure of the moduli space of stable genus zero curves is captured by a certain non-unital version of operads. Let $\mathcal{M}_{0,n+1}$ be the moduli space of $(n+1)$-tuples $(x_0, \dots, x_n)$ of distinct numbered points on the complex projective line $\mathbb{CP}^1$ modulo projective automorphisms, that is, transformations of the form

$$\mathbb{CP}^1 \ni [\xi_1, \xi_2] \mapsto [a\xi_1 + b\xi_2, c\xi_1 + d\xi_2] \in \mathbb{CP}^1,$$

where $a, b, c, d \in \mathbb{C}$ with $ad - bc \neq 0$.

The moduli space $\mathcal{M}_{0,n+1}$ has, for $n \geqslant 2$, a canonical compactification $\overline{\mathcal{M}}_0(n) \supset \mathcal{M}_{0,n+1}$ introduced by A. Grothendieck and F.F. Knudsen [16,50]. The space $\overline{\mathcal{M}}_0(n)$ is the moduli space of stable $(n+1)$-pointed curves of genus 0:

DEFINITION 15. A *stable $(n+1)$-pointed curve of genus* 0 is an object

$$(C; x_0, \dots, x_n),$$

where C is a (possibly reducible) algebraic curve with at most nodal singularities and $x_0, \dots, x_n \in C$ are distinct smooth points such that:

(i) each component of C is isomorphic to $\mathbb{CP}^1$,

(ii) the graph of intersections of components of C (i.e. the graph whose vertices correspond to the components of C and edges to the intersection points of the components) is a tree,

(iii) each component of C has at least three special points, where a special point means either one of the x_i, $0 \leqslant i \leqslant n$, or a singular point of C (the stability).

It can be easily seen that a stable curve $(C; x_0, \dots, x_n)$ admits no infinitesimal automorphisms that fix marked points $x_0, \dots, x_n$, therefore $(C; x_0, \dots, x_n)$ is 'stable' in the usual sense. Observe also that $\overline{\mathcal{M}}_0(0) = \overline{\mathcal{M}}_0(1) = \emptyset$ (there are no stable curves with less than three marked points) and that $\overline{\mathcal{M}}_0(2) =$ the point corresponding to the three-pointed stable curve $(\mathbb{CP}^1; \infty, 1, 0)$. The space $\mathcal{M}_{0,n+1}$ forms an open dense part of $\overline{\mathcal{M}}_0(n)$ consisting of marked curves $(C; x_0, \dots, x_n)$ such that C is isomorphic to $\mathbb{CP}^1$.

Let us try to equip the collection $\overline{\mathcal{M}}_0 = \{\overline{\mathcal{M}}_0(n)\}_{n \geqslant 2}$ with an operad structure as in Definition 1. For $C = (C, x_1, \dots, x_n) \in \overline{\mathcal{M}}_0(n)$ and $C_i = (C_i, y_1^i, \dots, y_{k_i}^i) \in \overline{\mathcal{M}}_0(k_i)$, $1 \leqslant i \leqslant n$, let

$$\gamma(C, C_1, \dots, C_n) \in \overline{\mathcal{M}}_0(k_1 + \dots + k_n) \tag{5}$$

be the stable marked curve obtained from the disjoint union $C \sqcup C^1 \sqcup \dots \sqcup C^n$ by identifying, for each $1 \leqslant i \leqslant n$, the point $x_i \in C$ with the point $y_0^i \in C_i$, introducing a nodal singularity, and relabeling the remaining marked points accordingly. The symmetric group acts on $\overline{\mathcal{M}}_0(n)$ by

$$(C, x_0, x_1, \dots, x_n) \mapsto (C, x_0, x_{\sigma(1)}, \dots, x_{\sigma(n)}), \quad \sigma \in \Sigma_n.$$

We have defined the γ-compositions and the symmetric group action, but there is no room for the identity, because $\overline{\mathcal{M}}_0(1)$ is empty! The above structure is, therefore, a non-unital operad in the sense of the following definition (which is formulated, as all definitions in this article, for the monoidal category of **k**-modules).

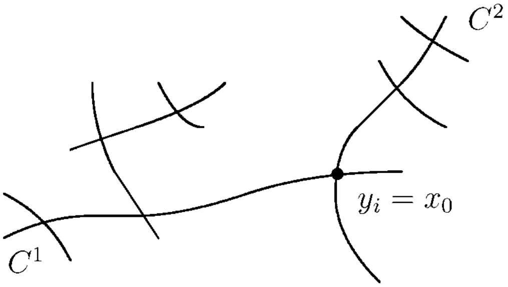


Fig. 2. The $\circ_i$-compositions in $\overline{\mathcal{M}}_0 = \{\overline{\mathcal{M}}_0(n)\}_{n\geqslant 2}$.

Definition 16. A *May non-unital operad* in the category of $\mathbf{k}$-modules is a collection $\mathcal{P} = \{\mathcal{P}(n)\}_{n\geqslant 0}$ of $\mathbf{k}[\Sigma_n]$-modules, together with operadic compositions

$$\gamma : \mathcal{P}(n) \otimes \mathcal{P}(k_1) \otimes \cdots \otimes \mathcal{P}(k_n) \to \mathcal{P}(k_1 + \cdots + k_n),$$

for $n \geqslant 1$ and $k_1, \ldots, k_n \geqslant 0$, that fulfill the associativity and equivariance axioms of Definition 1.

We may as well define on the collection $\overline{\mathcal{M}}_0 = \{\overline{\mathcal{M}}_0(n)\}_{n\geqslant 2}$ operations

$$\circ_i : \overline{\mathcal{M}}_0(m) \times \overline{\mathcal{M}}_0(n) \to \overline{\mathcal{M}}_0(m+n-1) \tag{6}$$

for $m, n \geqslant 2$, $1 \leqslant i \leqslant m$, by

$$(C^1; y_0, \ldots, y_m) \times (C^2; x_0, \ldots, x_n)$$
$$\mapsto (C; y_0, \ldots, y_{i-1}, x_0, \ldots, x_n, y_{i+1}, \ldots, y_m)$$

where C is the quotient of the disjoint union $C^1 \sqcup C^2$ given by identifying x_0 with y_i at a nodal singularity, see Fig. 2. The collection $\overline{\mathcal{M}}_0 = \{\overline{\mathcal{M}}_0(n)\}_{n\geqslant 2}$ with $\circ_i$-operations (6) is an example of another version of non-unital operads, recalled in:

Definition 17. A *non-unital Markl operad* in the category of $\mathbf{k}$-modules is a collection $\mathcal{P} = \{\mathcal{P}(n)\}_{n\geqslant 0}$ of $\mathbf{k}[\Sigma_n]$-modules, together with operadic compositions

$$\circ_i : \mathcal{S}(m) \otimes \mathcal{S}(n) \to \mathcal{S}(m+n-1),$$

for $1 \leqslant i \leqslant m$ and $n \geqslant 0$, that fulfill the associativity and equivariance axioms of Definition 11.

A we saw in Proposition 13, in the presence of operadic units, May operads are the same as Markl operads. Surprisingly, the non-unital versions of these structures are *radically different* – Markl operads capture more information than May operads! This is made precise in the following:

Proposition 18. *The category of non-unital Markl operads is a subcategory of the category of non-unital May operads.*

PROOF. It is easy to see that (3) defines, as in the proof of Proposition 13, a functor $\psi\mathrm{May}(-)$ which is an embedding of the category of non-unital Markl operads into the category of non-unital May operads. □

Observe that formula (4), inverse to (3), does not make sense without units. The relation between various versions of operads discussed so far is summarized in the following diagram of categories and their inclusions:

$$\begin{tikzcd}
\text{May operads} \arrow[d, hook] \arrow[rr, bend left, "\text{Mar}"] & & \text{Markl operads} \arrow[ll, "\text{May}"'] \arrow[d, hook] \\
\text{non-unital May operads} & & \text{non-unital Markl operads} \arrow[ll, hook, "\psi\text{May}"']
\end{tikzcd}$$

The following example shows that non-unital Markl operads form a proper sub-category of the category of non-unital May operads.

EXAMPLE 19. We describe a non-unital May operad $\mathcal{V} = \{\mathcal{V}(n)\}_{n \geqslant 0}$ which is not of the form $\psi\mathrm{May}(\mathcal{S})$ for some non-unital Markl operad $\mathcal{S}$. Let

$$\mathcal{V}(n) := \begin{cases} \mathbf{k} & \text{for } n = 2 \text{ or } 4, \\ 0 & \text{otherwise.} \end{cases}$$

The only non-trivial γ-composition is $\gamma : \mathcal{V}(2) \otimes \mathcal{V}(2) \otimes \mathcal{V}(2) \to \mathcal{V}(4)$, given as the canonical isomorphism

$$\mathcal{V}(2) \otimes \mathcal{V}(2) \otimes \mathcal{V}(2) \cong \mathbf{k}^{\otimes 3} \xrightarrow{\cong} \mathbf{k} \cong \mathcal{V}(4).$$

Suppose that $\mathcal{V} = \mathrm{May}(\mathcal{S})$ for some non-unital Markl operad $\mathcal{S}$. Then, according to (3), for $f, g_1, g_2 \in \mathcal{V}(2)$,

$$\gamma(f, g_1, g_2) = (f \circ_2 g_2) \circ_1 g_1.$$

Since $(f \circ_2 g) \in \mathcal{V}(3) = 0$, this would imply that γ is trivial, which is not true.

Proposition 21 below shows that the Markl non-unital operads rather than the May non-unital operads are the true non-unital versions of operads. We will need the following definition in which $\mathcal{K} = \{\mathcal{K}(n)\}_{n \geqslant 1}$ is the trivial (unital) operad with $\mathcal{K}(1) := \mathbf{k}$ and $\mathcal{K}(n) = 0$, for $n \neq 1$.

DEFINITION 20. An *augmentation* of an operad $\mathcal{P}$ in the category of $\mathbf{k}$-modules is a homomorphism $\epsilon : \mathcal{P} \to \mathcal{K}$. Operads with an augmentation are called *augmented* operads. The kernel

$$\bar{\mathcal{P}} := \mathrm{Ker}(\epsilon : \mathcal{P} \to \mathcal{K})$$

is called the *augmentation ideal*.

The following proposition was proved in [75].

PROPOSITION 21. *The correspondence $\mathcal{P} \mapsto \overline{\mathcal{P}}$ is an isomorphism between the category of augmented operads and the category of Markl non-unital operads.*

PROOF. The $\circ_i$-operations of $\mathcal{P}$ obviously restrict to $\overline{\mathcal{P}}$, making it a non-unital Markl operad. It is simple to describe a functorial inverse $\mathcal{S} \mapsto \widetilde{\mathcal{S}}$ of the correspondence $\mathcal{P} \mapsto \overline{\mathcal{P}}$. For a Markl non-unital operad $\mathcal{S}$, denote by $\widetilde{\mathcal{S}}$ the collection

$$\widetilde{\mathcal{S}}(n) := \begin{cases} \mathcal{S}(n), & \text{for } n \neq 1, \\ \mathcal{S}(1) \oplus \mathbf{k}, & \text{for } n = 1. \end{cases} \tag{7}$$

The $\circ_i$-operations of $\widetilde{\mathcal{S}}$ are uniquely determined by requiring that they extend the $\circ_i$-operations of $\mathcal{S}$ and satisfy (2), with the unit $e := 0 \oplus 1_{\mathbf{k}} \in \mathcal{S}(1) \oplus \mathbf{k} = \widetilde{\mathcal{S}}(1)$. Informally, $\widetilde{\mathcal{S}}$ is obtained from the Markl non-unital operad $\mathcal{S}$ by adjoining a unit. □

Observe that if $\mathcal{S}$ were a May, not Markl, non-unital operad, the construction of $\widetilde{\mathcal{S}}$ described in the above proof would not make sense, because we would not know how to define

$$\gamma(f, \underbrace{e, \dots, e}_{i-1}, g, \underbrace{e, \dots, e}_{m-i})$$

for $f \in \mathcal{S}(m)$, $g \in \mathcal{S}(n)$, $m \geqslant 2$, $n \geqslant 0$, $1 \leqslant i \leqslant m$. Proposition 21 should be compared to the obvious statement that the category of augmented unital associative algebras is isomorphic to the category of (non-unital) associative algebras. In the following proposition, $\mathtt{Oper}$ denotes the category of $\mathbf{k}$-linear operads and $\psi\mathtt{Oper}$ the category of $\mathbf{k}$-linear Markl non-unital operads.

PROPOSITION 22. *Let $\mathcal{P}$ be an augmented operad and $\mathcal{Q}$ an arbitrary operad in the category of* $\mathbf{k}$*-modules. Then there exists a natural isomorphism*

$$\mathrm{Mor}_{\mathtt{Oper}}(\mathcal{P}, \mathcal{Q}) \cong \mathrm{Mor}_{\psi\mathtt{Oper}}(\overline{\mathcal{P}}, \psi\mathrm{May}(\mathcal{Q})). \tag{8}$$

The proof is simple and we leave it to the reader. Combining (8) with the isomorphism of Proposition 21 one obtains a natural isomorphism

$$\mathrm{Mor}_{\mathtt{Oper}}(\widetilde{\mathcal{S}}, \mathcal{Q}) \cong \mathrm{Mor}_{\psi\mathtt{Oper}}\big(\mathcal{S}, \psi\mathrm{May}(\mathcal{Q})\big) \tag{9}$$

which holds for each Markl non-unital operad $\mathcal{S}$ and operad $\mathcal{Q}$. Isomorphism (9) means that $\widetilde{\ }: \psi\mathtt{Oper} \to \mathtt{Oper}$ and $\psi\mathrm{May} : \mathtt{Oper} \to \psi\mathtt{Oper}$ are adjoint functors. This adjunction will be used in the construction of free operads in Section 4.

In the rest of this chapter, non-unital Markl operads will be called simply non-unital operads. This will not lead to confusion, since all non-unital operads referred to in the rest of this chapter will be Markl.

3. Operad algebras

As we already remarked, operads are important through their representations called operad algebras or simply algebras.

DEFINITION 23. Let V be a $\mathbf{k}$-module and $\mathcal{E}nd_V$ the endomorphism operad of V recalled in Example 2. A $\mathcal{P}$-*algebra* is a homomorphism of operads $\rho : \mathcal{P} \to \mathcal{E}nd_V$.

The above definition admits an obvious generalization for an arbitrary symmetric monoidal category with an internal hom-functor. The last assumption is necessary for the existence of the 'internal' endomorphism operad, see [83, Definition II.1.20]. Definition 23 can, however, be unwrapped into the form given in [58, Definition 2.1] that makes sense in an arbitrary symmetric monoidal category without the internal hom-functor assumption:

PROPOSITION 24. *Let $\mathcal{P}$ be an operad. A $\mathcal{P}$-algebra is the same as a $\mathbf{k}$-module V together with maps*

$$\alpha : \mathcal{P}(n) \otimes V^{\otimes n} \to V, \quad n \geqslant 0, \tag{10}$$

that satisfy the following axioms.

Associativity. For each $n \geqslant 1$ and non-negative integers $k_1, \ldots, k_n$, the following diagram commutes.

$$\begin{tikzcd}
\left(\mathcal{P}(n) \otimes \bigotimes_{s=1}^{n} \mathcal{P}(k_s)\right) \otimes \bigotimes_{s=1}^{n} V^{\otimes k_s} \arrow[r, "\gamma\otimes id"] \arrow[dd, "\text{shuffle}"'] & \mathcal{P}(k_1 + \cdots + k_n) \otimes V^{\otimes(k_1+\cdots+k_n)} \arrow[d, "\alpha"] \\
 & V \\
\mathcal{P}(n) \otimes \bigotimes_{s=1}^{n}\left(\mathcal{P}(k_s) \otimes V^{\otimes k_s}\right) \arrow[r, "id\otimes(\bigotimes_{s=1}^{n}\alpha)"] & \mathcal{P}(n) \otimes V^{\otimes n} \arrow[u, "\alpha"']
\end{tikzcd}$$

Equivariance. For each $n \geqslant 1$ and $\sigma \in \Sigma_n$, the following diagram commutes.

$$\begin{tikzcd}
\mathcal{P}(n) \otimes V^{\otimes n} \arrow[rr, "\sigma\otimes\sigma^{-1}"] \arrow[dr, "\alpha"'] & & \mathcal{P}(n) \otimes V^{\otimes n} \arrow[dl, "\alpha"] \\
 & V &
\end{tikzcd}$$

Unitality. For each $n \geqslant 1$, the following diagram commutes.

$$\begin{tikzcd}
\mathbf{k} \otimes V \arrow[r, "\cong"] \arrow[d, "\eta\otimes id"'] & V \\
\mathcal{P}(1) \otimes V \arrow[ur, "\alpha"] &
\end{tikzcd}$$

We leave as an exercise to formulate a version of Proposition 24 that would use $\circ_i$-operations instead of γ-operations.

EXAMPLE 25. In this example we verify, using Proposition 24, that algebras over the operad $\mathcal{C}om = \{\mathcal{C}om(n)\}_{n\geqslant 1}$ recalled in Example 10 are ordinary commutative associative algebras. To simplify the exposition, let us agree that v's with various subscripts denote elements of V. Since $\mathcal{C}\mathrm{om}(n) = \mathbf{k}$ for $n \geqslant 1$, the structure map (10) determines, for each $n \geqslant 1$, a linear map $\mu_n : V^{\otimes n} \to V$ by

$$\mu_n(v_1, \dots, v_n) := \alpha(1_n, v_1, \dots, v_n),$$

where 1_n denotes in this example the unit $1_n \in \mathbf{k} = \mathcal{C}\mathrm{om}(n)$. The associativity of Proposition 24 says that

$$\begin{aligned}&\mu_n\bigl(\mu_{k_1}(v_1, \dots, v_{k_1}), \dots, \mu_{k_n}(v_{k_1+\cdots+k_{n-1}+1}, \dots, v_{k_1+\cdots+k_n})\bigr)\\ &\quad = \mu_{k_1+\cdots+k_n}(v_1, \dots, v_{k_1+\cdots+k_n}),\end{aligned} \tag{11}$$

for each $n, k_1, \dots, k_n \geqslant 1$. The equivariance of Proposition 24 means that each μ_n is fully symmetric

$$\mu_n(v_1, \dots, v_n) = \mu_n(v_{\sigma(1)}, \dots, v_{\sigma(n)}), \quad \sigma \in \Sigma_n, \tag{12}$$

and the unitality implies that μ_1 is the identity map,

$$\mu_1(v) = v. \tag{13}$$

The above structure can be identified with a commutative associative multiplication on V. Indeed, the bilinear map $\cdot := \mu_2 : V \otimes V \to V$ is clearly associative:

$$(v_1 \cdot v_2) \cdot v_3 = v_1 \cdot (v_2 \cdot v_3) \tag{14}$$

and commutative:

$$v_1 \cdot v_2 = v_2 \cdot v_1. \tag{15}$$

On the other hand, $\mu_1(v) := v$ and

$$\mu_n(v_1, \dots, v_n) := \bigl(\cdots (v_1 \cdot v_2) \ \cdots \ v_{n-1}\bigr) \cdot v_n \quad \text{for } n \geqslant 2,$$

defines multilinear maps $\{\mu_n : V^{\otimes n} \to V\}$ satisfying (11)–(13). It is equally easy to verify that algebras over the operad $\mathcal{A}\mathrm{ss}$ introduced in Example 3 are ordinary associative algebras.

Following Leinster [61], one could say that (11)–(13) is an *unbiased* definition of associative commutative algebras, while (14), (15) is a definition of the same object *biased* towards bilinear operations. Operads therefore provide unbiased definitions of algebras.

EXAMPLE 26. Let us denote by $\mathrm{U}\mathcal{C}\mathrm{om}$ the endomorphism operad $\mathcal{E}\mathrm{nd}_{\mathbf{k}}$ of the ground ring $\mathbf{k}$. It is easy to verify that $\mathrm{U}\mathcal{C}\mathrm{om}$-algebras are *unital* commutative associative algebras. We leave it to the reader to describe the operad $\mathrm{U}\mathcal{A}\mathrm{ss}$ governing unital associative operads.

Algebras over a non-Σ operad $\underline{\mathcal{P}}$ are defined as algebras, in the sense of Definition 23, over the symmetrization $\Sigma[\underline{\mathcal{P}}]$ of $\underline{\mathcal{P}}$. Algebras over non-unital operads as discussed in Section 2 are defined by appropriate obvious modifications of Definition 23.

EXAMPLE 27. Let Y be a topological space with a base point $*$ and $\mathbb{S}^k$ the k-dimensional sphere, $k \geqslant 1$. The k-fold loop space $\Omega^k Y$ is the space of all continuous maps $\mathbb{S}^k \to Y$ that send the south pole of $\mathbb{S}^k$ to the base point of Y. Equivalently, $\Omega^k Y$ is the space of all continuous maps $\lambda : (\mathbb{D}^k, \mathbb{S}^{k-1}) \to (Y, *)$ from the standard closed k-dimensional disc $\mathbb{D}^k$ to Y that map the boundary $\mathbb{S}^{k-1}$ of $\mathbb{D}^k$ to the base point of Y. Let us show, following Boardman and Vogt [10], that $\Omega^k Y$ is a natural topological algebra over the little k-discs operad $\mathcal{D}_k = \{\mathcal{D}_k(n)\}_{n\geqslant 0}$ recalled in Example 4.

The action $\alpha : \mathcal{D}_k(n) \times (\Omega^k Y)^{\times n} \to \Omega^k Y$ is, for $n \geqslant 0$, defined as follows. Given $\lambda_i : (\mathbb{D}^k, \mathbb{S}^{k-1}) \to (Y, *) \in \Omega^k Y$, $1 \leqslant i \leqslant n$, and little k-discs $d = (d_1, \dots, d_n) \in \mathcal{D}_k(n)$ as in Example 4, then

$$\alpha(d, \lambda_1, \dots, \lambda_n) : (\mathbb{D}^k, \mathbb{S}^{k-1}) \to (Y, *) \in \Omega^k Y$$

is the map defined to be $\lambda_i : \mathbb{D}^k \to Y$ (suitably rescaled) on the image of d_i, and to be $*$ on the complement of the images of the maps d_i, $1 \leqslant i \leqslant n$.

Therefore each k-fold loop space is a $\mathcal{D}_k$-space. The following classical theorem is a certain form of the inverse statement.

THEOREM 28 *(Boardman–Vogt [10], May [86]). A path-connected $\mathcal{D}_k$-algebra X has the weak homotopy type of a k-fold loop space.*

The connectedness assumption in the above theorem can be weakened by assuming that the $\mathcal{D}_k$-action makes the set $\pi_0(X)$ of path components of X a group.

EXAMPLE 29. The non-unital operad $\overline{\mathcal{M}}_0$ of stable pointed curves of genus 0 (also called the *configuration (non-unital) operad*) recalled on p. 100 is a non-unital operad in the category of smooth complex projective varieties. It therefore makes sense, as explained in Example 5, to consider its homology operad $H_*(\overline{\mathcal{M}}_0, \mathbf{k}) = \{H_*(\overline{\mathcal{M}}_0(n), \mathbf{k})\}_{n\geqslant 2}$.

An algebra over this non-unital operad is called a (tree level) *cohomological conformal field theory* or a *hyper-commutative algebra* [55]. It consist of a family

$$\big\{(\text{-}, \dots, \text{-}) : V^{\otimes n} \to V\big\}_{n\geqslant 2}$$

of linear operations which are totally symmetric, that is

$$(v_{\sigma(1)}, \dots, v_{\sigma(n)}) = (v_1, \dots, v_n),$$

for each permutation $\sigma \in \Sigma_n$. Moreover, we require the following form of associativity:

$$\sum_{(S,T)} \big((u, v, x_i;\ i \in S), w, x_j;\ j \in T\big)$$
$$= \sum_{(S,T)} \big(u, (v, w, x_i;\ i \in S), x_j;\ j \in T\big), \tag{16}$$

where $u, v, w, x_1, \dots, x_n \in V$ and (S, T) runs over disjoint decompositions $S \sqcup T = \{1, \dots, n\}$. For $n = 0$, (16) means the (usual) associativity of the bilinear operation $(-, -)$, i.e. $((u, v), w) = (u, (v, w))$. For $n = 1$ we get

$$\big((u, v), w, x\big) + \big((u, v, x), w\big) = \big(u, (v, w, x)\big) + \big(u, (v, w), x\big).$$

EXAMPLE 30. In this example, $\mathbf{k}$ is a field of characteristic 0. The non-unital operad $\overline{\mathcal{M}}_0(\mathbb{R}) = \{\overline{\mathcal{M}}_0(\mathbb{R})(n)\}_{n\geqslant 2}$ of real points in the configuration operad $\overline{\mathcal{M}}_0$ is called the *mosaic non-unital operad* [19]. Algebras over the homology $H_*(\overline{\mathcal{M}}_0(\mathbb{R}), \mathbf{k}) = \{H_*(\overline{\mathcal{M}}_0(\mathbb{R})(n), \mathbf{k})\}_{n\geqslant 2}$ of this operad were recently identified [25] with 2-*Gerstenhaber algebras*, which are structures (V, μ, τ) consisting of a commutative associative product $\mu : V \otimes V \to V$ and an anti-symmetric degree $+1$ ternary operation $\tau : V \otimes V \otimes V \to V$ which satisfies the generalized Jacobi identity

$$\sum_\sigma \operatorname{sgn}(\sigma) \cdot \tau\bigl(\tau(x_{\sigma(1)}, x_{\sigma(2)}, x_{\sigma(3)}), x_{\sigma(4)}, x_{\sigma(5)}\bigr) = 0,$$

where the summation runs over all (3, 2)-unshuffles $\sigma(1) < \sigma(2) < \sigma(3)$, $\sigma(4) < \sigma(5)$. Moreover, the ternary operation τ is tied to the multiplication μ by the distributive law

$$\tau\bigl(\mu(s,t), u, v\bigr) = \mu\bigl(\tau(s,u,v), t\bigr) + (-1)^{(1+|u|+|v|)|s|} \cdot \mu\bigl(s, \tau(t,u,v)\bigr),$$

$s, t, u, v \in V$, saying that the assignment $s \mapsto \tau(s, u, v)$ is a degree $(1 + |u| + |v|)$-derivation of the associative commutative algebra (V, μ), for each $u, v \in V$.

4. Free operads and trees

The purpose of this section is three-fold. First, we want to study free operads because each operad is a quotient of a free one. The second reason why we are interested in free operads is that their construction involves trees. Indeed, it turns out that rooted trees provide 'pasting schemes' for operads and that, replacing trees by other types of graphs, one can introduce several important generalizations of operads, such as cyclic operads, modular operads, and PROPs. The last reason is that the free operad functor defines a monad which provides an unbiased definition of operads as algebras over this monad. Everything in this section is written for $\mathbf{k}$-linear operads, but the constructions can be generalized into an arbitrary symmetric monoidal category with countable coproducts $(\mathcal{M}, \odot, \mathbf{1})$ whose monoidal product $\odot$ is distributive over coproducts, see [83, Section II.1.9].

Recall that a *Σ-module* is a collection $E = \{E(n)\}_{n\geqslant 0}$ in which each $E(n)$ is a right $\mathbf{k}[\Sigma_n]$-module. There is an obvious forgetful functor $\Box : \mathtt{Oper} \to \Sigma\text{-}\mathtt{mod}$ from the category $\mathtt{Oper}$ of $\mathbf{k}$-linear operads to the category $\Sigma\text{-}\mathtt{mod}$ of Σ-modules.

DEFINITION 31. The *free operad functor* is the left adjoint [38, §II.7] $\Gamma : \Sigma\text{-}\mathtt{mod} \to \mathtt{Oper}$ to the forgetful functor $\Box : \mathtt{Oper} \to \Sigma\text{-}\mathtt{mod}$. This means that there exists a functorial isomorphism

$$\mathrm{Mor}_{\mathtt{Oper}}\bigl(\Gamma(E), \mathcal{P}\bigr) \cong \mathrm{Mor}_{\Sigma\text{-}\mathtt{mod}}\bigl(E, \Box(\mathcal{P})\bigr)$$

for an arbitrary Σ-module E and operad $\mathcal{P}$. The operad $\Gamma(E)$ is the *free operad* generated by the Σ-module E. Similarly, the *free non-unital operad functor* is a left adjoint $\Psi : \Sigma\text{-}\mathtt{mod} \to \psi\mathtt{Oper}$ of the obvious forgetful functor $\Box_\psi : \psi\mathtt{Oper} \to \Sigma\text{-}\mathtt{mod}$, that is

$$\mathrm{Mor}_{\psi\mathtt{Oper}}\bigl(\Psi(E), \mathcal{S}\bigr) \cong \mathrm{Mor}_{\Sigma\text{-}\mathtt{mod}}\bigl(E, \Box_\psi(\mathcal{S})\bigr),$$

where E is a Σ-module and $\mathcal{S}$ a non-unital operad. The non-unital operad $\Psi(E)$ is the *free non-unital operad* generated by the Σ-module E.

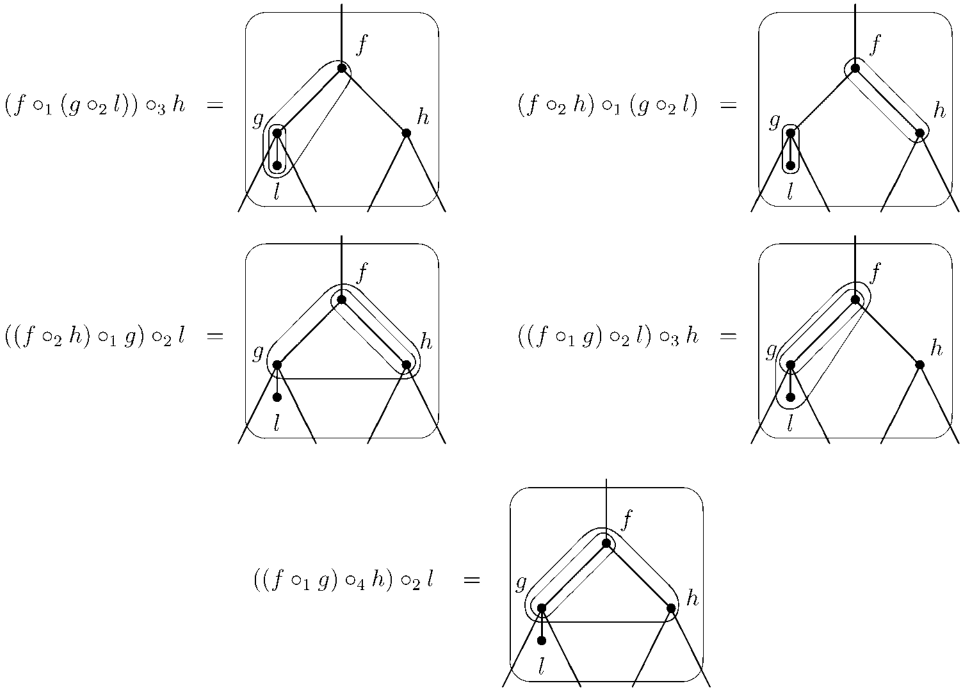

$(f \circ_1 (g \circ_2 l)) \circ_3 h =$
$(f \circ_2 h) \circ_1 (g \circ_2 l) =$
$((f \circ_2 h) \circ_1 g) \circ_2 l =$
$((f \circ_1 g) \circ_2 l) \circ_3 h =$
$((f \circ_1 g) \circ_4 h) \circ_2 l =$
f
g
h
l

Fig. 3. Flow diagrams in non-unital operads.

Let $\widetilde{\ }\colon \Psi\mathtt{Oper} \to \mathtt{Oper}$ be the functor of 'adjoining the unit' considered in the proof of Proposition 21 on p. 103. The functorial isomorphism (9) implies that one may take

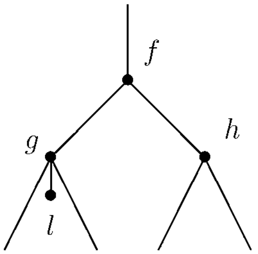
$$\Gamma := \widetilde{\Psi}, \tag{17}$$

which means that the free operad $\Gamma(E)$ can be obtained from the free non-unital operad $\Psi(E)$ by formally adjoining the unit.

Let us indicate how to construct the free non-unital operad $\Psi(E)$, a precise description will be given later in this section. The free non-unital operad $\Psi(E)$ must be built up from all formal $\circ_i$-compositions of elements of E modulo the axioms listed in Definition 11. For instance, given $f \in E(2)$, $g \in E(3)$, $h \in E(2)$ and $l \in E(0)$, the component $\Psi(E)(5)$ must contain the following five compositions

$$\begin{gathered}\big(f \circ_1 (g \circ_2 l)\big) \circ_3 h, \quad (f \circ_2 h) \circ_1 (g \circ_2 l), \quad \big((f \circ_2 h) \circ_1 g\big) \circ_2 l,\\ \big((f \circ_1 g) \circ_2 l\big) \circ_3 h \quad \text{and} \quad \big((f \circ_1 g) \circ_4 h\big) \circ_2 l.\end{gathered} \tag{18}$$

The elements in (18) can be depicted by the 'flow diagrams' of Fig. 3. Nodes of these diagrams are decorated by elements f, g, h and l of E in such a way that an element of $E(n)$ decorates a node with n input lines, $n \geqslant 0$. Thin 'amoebas' indicate the nesting which specifies the order in which the $\circ_i$-operations are performed. The associativity of Definition 11 however says that the result of the composition does not depend on the order, therefore the amoebas can be erased and the common value of the compositions represented by



(19)

Let us look more closely how diagram (19) determines an element of the (still hypothetical) free non-unital operad $\Psi(E)$. The crucial fact is that the underlying graph of (19) is a planar rooted tree. Recall that a *tree* is a finite connected simply connected graph without loops and multiple edges. For a tree T we denote, as usual, by $\mathrm{Vert}(T)$ the set of vertices and $\mathrm{Edg}(T)$ the set of edges of T. The number of edges adjacent to a vertex $v \in \mathrm{Vert}(T)$ is called the *valence* of v and denoted $\mathrm{val}(v)$. We assume that one is given a subset

$$\mathrm{ext}(T) \subset \big\{v \in \mathrm{Vert}(T);\ \mathrm{val}(v) = 1\big\}$$

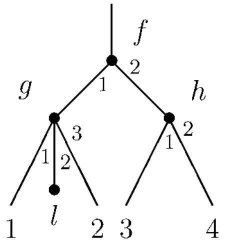
of *external* vertices, the remaining vertices are *internal*. Let us denote

$$\mathrm{vert}(T) := \mathrm{Vert}(T) \setminus \mathrm{ext}(T)$$

the set of all internal vertices. Henceforth, we will assume that our trees have at least one internal vertex. This excludes at this stage the *exceptional tree* consisting of two external vertices connected by an edge.

Edges adjacent to external vertices are the *legs* of T. A tree is *rooted* if one of its legs, called the *root*, is marked and all other edges are oriented, pointing to the root. The legs different from the root are the *leaves* of T. For example, the tree in (19) has 4 internal vertices decorated f, g, h and l, and 4 leaves. Finally, the *planarity* means that an embedding of T into the plane is specified. In our pictures, the root will always be placed on the top. By a vertex we will always mean an internal one.

The planarity and a choice of the root of the underlying tree of (19) specifies a total order of the set $\mathrm{in}(v)$ of input edges of each vertex $v \in \mathrm{vert}(T)$ as well as a total order of the set $\mathrm{Leaf}(T)$ of the leaves of T, by numbering from the left to the right:



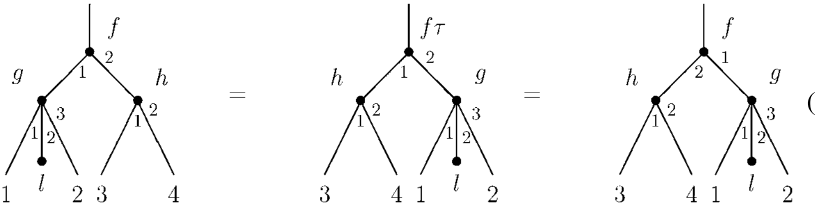
(20)

This tells us that l should be inserted into the second input of g, g into the first input of f and h into the second input of f. Using 'abstract variables' v_1, v_2, v_3 and v_4, the element represented by (20) can also be written as the 'composition' $f(g(v_1, l, v_2), h(v_3, v_4))$.

Now we need to take into account also the symmetric group action. If τ is the generator of Σ_2, then the obvious equality

$$f\big(g(v_1, l, v_2), h(v_3, v_4)\big) = f\tau\big(h(v_3, v_4), g(v_1, l, v_2)\big)$$

of 'abstract compositions' coming from the equivariance of Definition 11 translates into the following equality of flow diagrams:



(21)

Relation (21) shows that the equivariance of Definition 11 violates the linear orders induced by the planar embedding of T. This leads us to the conclusion that the flow diagrams describing elements of free non-unital operads are (abstract, non-planar) rooted, leaf-labeled decorated trees.

Let us describe, after these motivational remarks, a precise construction of $\Psi(E)$. The first subtlety one needs to understand is how to decorate vertices of non-planar trees. To this end, we need to explain how each Σ-module $E = \{E(n)\}_{n\geqslant 0}$ naturally extends into a functor (denoted again E) from the category $\mathtt{Set}_f$ of finite sets and their bijections to the category of $\mathbf{k}$-modules. If X and Y are finite sets, denote by

$$\mathrm{Bij}(Y, X) := \{\vartheta : X \xrightarrow{\cong} Y\} \tag{22}$$

the set of all isomorphisms between X and Y (notice the unexpected direction of the arrow!). It is clear that $\mathrm{Bij}(Y, X)$ is a natural left Aut_Y- right Aut_X-bimodule, where $\mathrm{Aut}_X := \mathrm{Bij}(X, X)$ and $\mathrm{Aut}_Y := \mathrm{Bij}(Y, Y)$ are the sets of automorphisms with group structure given by composition. For a finite set $S \in \mathtt{Set}_f$ of cardinality n and a Σ-module $E = \{E(n)\}_{n\geqslant 0}$ define $E(S)$ to be

$$E(S) := E(n) \times_{\Sigma_n} \mathrm{Bij}\big([n], S\big) \tag{23}$$

where, as usual, $[n] := \{1, \dots, n\}$ and, of course, $\Sigma_n = \mathrm{Aut}_{[n]}$.

Let us recall that a *(leaf-) labeled rooted n-tree* is a rooted tree T together with a specified bijection $\ell : \mathrm{Leaf}(T) \xrightarrow{\sim} [n]$. Let $\mathtt{Tree}_n$ be the category of labeled rooted n-trees and their bijections. For $T \in \mathtt{Tree}_n$ define

$$E(T) := \bigotimes_{v\in \mathrm{vert}(T)} E\big(\mathrm{in}(v)\big) \tag{24}$$

where $\mathrm{in}(v)$ is, as before, the set of all input edges of a vertex $v \in \mathrm{vert}(T)$. It is easy to verify that $E \mapsto E(T)$ defines a functor from the category $\mathtt{Tree}_n$ to the category of $\mathbf{k}$-modules.

Recall that the colimit of a covariant functor $F : \mathcal{D} \to \mathtt{Mod}_{\mathbf{k}}$ is the quotient

$$\operatorname*{colim}_{x\in\mathcal{D}} F(x) = \bigoplus_{x\in\mathcal{D}} F(x)/\sim,$$

where $\sim$ is the equivalence generated by

$$F(y) \ni a \sim F(f)(a) \in F(z),$$

for each $a \in F(y)$, $y, z \in \mathcal{D}$ and $f \in \mathrm{Mor}_{\mathcal{D}}(y, z)$. Define finally

$$\Psi(E)(n) := \operatorname*{colim}_{T \in \mathtt{Tree}_n} E(T), \quad n \geqslant 0. \tag{25}$$

The following theorem was proved in [83, II.1.9].

THEOREM 32. *There exists a natural non-unital operad structure on the Σ-module*

$$\Psi(E) = \bigl\{\Psi(E)(n)\bigr\}_{n \geqslant 0},$$

with the $\circ_i$-operations given by the grafting of trees and the symmetric group re-labeling the leaves, such that $\Psi(E)$ is the free non-unital operad generated by the Σ-module E.

One could simplify (25) by introducing $\mathcal{T}\mathrm{ree}(n)$ as the set of *isomorphism* classes of n-trees from $\mathtt{Tree}_n$ and defining $\Psi(E)$ by the formula

$$\Psi(E)(n) = \bigoplus_{[T] \in \mathcal{T}\mathrm{ree}(n)} E(T), \quad n \geqslant 0, \tag{26}$$

which does not involve the colimit. The drawback of (26) is that it assumes a choice of a representative $[T]$ of each isomorphism class in $\mathcal{T}\mathrm{ree}(n)$, while (25) is functorial and admits simple generalizations to other types of operads and PROPs. See [83, Section II.1.9] for other representations of the free non-unital operad functor.

Having constructed the free non-unital operad $\Psi(E)$, we may use (17) to define the free operad $\Gamma(E)$. This is obviously equivalent to enlarging, in (25) for $n = 1$, the category $\mathtt{Tree}_n$ by the *exceptional rooted tree* | with one leg and no internal vertex. If we denote this enlarged category of trees and their isomorphisms (which, however, differs from $\mathtt{Tree}_n$ only at $n = 1$) by $\mathtt{UTree}_n$, we may represent the free operad as

$$\Gamma(E)(n) := \operatorname*{colim}_{T \in \mathtt{UTree}_n} E(T), \quad n \geqslant 0. \tag{27}$$

If E is a Σ-module such that $E(0) = E(1) = 0$, then (26) reduces to a summation over *reduced trees*, that is trees whose all vertices have at least two input edges. By simple combinatorics, the number of isomorphism classes of reduced trees in $\mathtt{Tree}_n$ is finite for each $n \geqslant 0$. This implies the following proposition that says that operads are relatively small objects.

PROPOSITION 33. *Let $E = \{E(n)\}_{n \geqslant 0}$ be a Σ-module such that*

$$E(0) = E(1) = 0$$

and such that $E(n)$ are finite-dimensional for $n \geqslant 2$. Then the spaces $\Psi(E)(n)$ and $\Gamma(E)(n)$ are finite-dimensional for each $n \geqslant 0$.

We close this section by showing how the free operad functor can be used to define operads. It follows from general principles that any operad $\mathcal{P}$ is a quotient $\mathcal{P} = \Gamma(E)/(R)$, where E and R are Σ-modules and (R) is the operadic ideal (see Definition 6) generated by R in $\Gamma(E)$.

EXAMPLE 34. The commutative associative operad $\mathcal{C}\mathrm{om}$ recalled in Example 10 is generated by the Σ-module

$$E_{\mathcal{C}\mathrm{om}}(n) := \begin{cases} \mathbf{k} \cdot \mu, & \text{if } n = 2, \\ 0 & \text{if } n \neq 2, \end{cases}$$

where $\mathbf{k} \cdot \mu$ is the trivial representation of Σ_2. The ideal of relations is generated by

$$R_{\mathcal{C}\mathrm{om}} := \mathrm{Span}_{\mathbf{k}}\{\mu(\mu \otimes id) - \mu(id \otimes \mu)\} \subset \Gamma(E_{\mathcal{C}\mathrm{om}})(3),$$

where $\mu(\mu \otimes id) - \mu(id \otimes \mu)$ is the obvious shorthand for $\gamma(\mu, \mu, e) - \gamma(\mu, e, \mu)$, with e the unit of $\Gamma(E_{\mathcal{C}\mathrm{om}})$.

Similarly, the operad $\mathcal{A}\mathrm{ss}$ for associative algebras reviewed in Example 3 is generated by the Σ-module $E_{\mathcal{A}\mathrm{ss}}$ such that

$$E_{\mathcal{A}\mathrm{ss}}(n) := \begin{cases} \mathbf{k}[\Sigma_2], & \text{if } n = 2, \\ 0 & \text{if } n \neq 2. \end{cases}$$

The ideal of relations is generated by the $\mathbf{k}[\Sigma_3]$-closure $R_{\mathcal{A}\mathrm{ss}}$ of the associativity

$$\alpha(\alpha \otimes id) - \alpha(id \otimes \alpha) \in \Gamma(E_{\mathcal{A}\mathrm{ss}})(3), \tag{28}$$

where α is a generator of the regular representation $E_{\mathcal{A}\mathrm{ss}}(2) = \mathbf{k}[\Sigma_2]$.

EXAMPLE 35. The operad $\mathcal{L}\mathrm{ie}$ governing Lie algebras is the quotient

$$\mathcal{L}\mathrm{ie} := \Gamma(E_{\mathcal{L}\mathrm{ie}})/(R_{\mathcal{L}\mathrm{ie}}),$$

where $E_{\mathcal{L}\mathrm{ie}}$ is the Σ-module

$$E_{\mathcal{L}\mathrm{ie}}(n) := \begin{cases} \mathbf{k} \cdot \beta & \text{if } n = 2, \\ 0 & \text{if } n \neq 2, \end{cases}$$

with $\mathbf{k} \cdot \beta$ is the signum representation of Σ_2. The ideal of relations $(R_{\mathcal{L}\mathrm{ie}})$ is generated by the Jacobi identity:

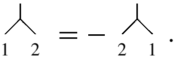
$$\beta(\beta \otimes id) + \beta(\beta \otimes id)c + \beta(\beta \otimes id)c^2 = 0, \tag{29}$$

in which $c \in \Sigma_3$ is the cyclic permutation $(1, 2, 3) \mapsto (2, 3, 1)$.

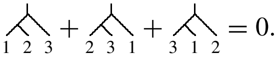
EXAMPLE 36. We show how to describe the presentations of the operads $\mathcal{A}\mathrm{ss}$ and $\mathcal{L}\mathrm{ie}$ given in Examples 34 and 35 in a simple graphical language. The generator α of $E_{\mathcal{A}\mathrm{ss}}$ is an operation with two inputs and one output, so we depict it as . The associativity (28) then reads

= ,

therefore $\mathcal{A}\mathrm{ss} = \Gamma(⅄)/(⅄ = ⅄)$. Also the operad for $\mathcal{L}\mathrm{ie}$ algebras is generated by one bilinear operation ⅄, but this time the operation is anti-symmetric:



The Jacobi identity (29) reads:



The kind of description used in the above examples is ‘tautological’ in the sense that it just says that the operad $\mathcal{P}$ governing a certain type of algebras is generated by operations of these algebras, with an appropriate symmetry, modulo the axioms satisfied by these operations. It does not say directly anything about the properties of the individual spaces $\mathcal{P}(n)$, $n \geqslant 0$. Describing these individual components may be a very nontrivial task, see for example the formula for the Σ_n-modules $\mathcal{L}\mathrm{ie}(n)$ given in [83, p. 50]. The operads in Examples 34 and 35 are quadratic in the sense of the following:

DEFINITION 37. An operad $\mathcal{P}$ is *quadratic* if it has a presentation $\mathcal{P} = \Gamma(E)/(R)$, where $E = \mathcal{P}(2)$ and $R \subset \Gamma(E)(3)$.

Quadratic operads form a very important class of operads. Each quadratic operad $\mathcal{P}$ has a *quadratic dual* $\mathcal{P}^!$ [34], [83, Definition II.3.37] which is a quadratic operad defined, roughly speaking, by dualizing the generators of $\mathcal{P}$ and replacing the relations of $\mathcal{P}$ by their annihilator in the dual space. For example, $\mathcal{A}\mathrm{ss}^! = \mathcal{A}\mathrm{ss}$, $\mathcal{C}\mathrm{om}^! = \mathcal{L}\mathrm{ie}$ and $\mathcal{L}\mathrm{ie}^! = \mathcal{C}\mathrm{om}$. A quadratic operad $\mathcal{P}$ is *Koszul* if it has the homotopy type of the bar construction of its quadratic dual [34], [83, Definition II.3.40]. For quadratic Koszul operads, there is a deep understanding of the derived category of the corresponding algebras. The operads $\mathcal{A}\mathrm{ss}$, $\mathcal{C}\mathrm{om}$ and $\mathcal{L}\mathrm{ie}$ above, as well as most quadratic operads one encounters in everyday life, are Koszul.

5. Unbiased definitions

In this section, we review the definition of a triple (monad) and give, in Theorem 40, a description of unital and non-unital operads in terms of algebras over a triple. The relevant triples come from the endofunctors Ψ and Γ recalled in Section 4. Let $\mathrm{End}(\mathcal{C})$ be the strict symmetric monoidal category of endofunctors on a category $\mathcal{C}$ where multiplication is the composition of functors.

DEFINITION 38. A *triple* (also called a *monad*) T on a category $\mathcal{C}$ is an associative and unital monoid (T, μ, υ) in $\mathrm{End}(\mathcal{C})$. The multiplication $\mu : TT \to T$ and unit morphism $\upsilon : \mathrm{id} \to T$ satisfy the axioms given by commutativity of the diagrams in Fig. 4.

Triples arise naturally from pairs of adjoint functors. Given an adjoint pair [38, II.7]

$$\begin{tikzcd} \mathsf{A} \arrow[r, bend left, "G"] & \mathsf{B} \arrow[l, bend left, "F"] \end{tikzcd}$$

$$\begin{tikzcd}
TTT \arrow[r, "T\mu"] \arrow[d, "\mu T"'] & TT \arrow[d, "\mu"] \\
TT \arrow[r, "\mu"] & T
\end{tikzcd}
\qquad
\begin{tikzcd}
T \arrow[rr, "T\upsilon"] \arrow[dr, "id"'] & & TT \arrow[dl, "\mu"] \\
& T &
\end{tikzcd}
\qquad
\begin{tikzcd}
T \arrow[rr, "\upsilon T"] \arrow[dr, "id"'] & & TT \arrow[dl, "\mu"] \\
& T &
\end{tikzcd}$$

Fig. 4. Associativity and unit axioms for a triple.

$$\begin{tikzcd}
T(T(A)) \arrow[r, "T(\alpha)"] \arrow[d, "\mu"'] & T(A) \arrow[d, "\alpha"] \\
T(A) \arrow[r, "\alpha"] & A
\end{tikzcd}
\qquad
\begin{tikzcd}
A \arrow[rr, "\upsilon_A"] \arrow[dr, "id"'] & & T(A) \arrow[dl, "\alpha"] \\
& A &
\end{tikzcd}$$

Fig. 5. T-algebra structure.

with associated functorial isomorphism

$$\mathrm{Mor}_{\mathsf{A}}\big(F(X), Y\big) \cong \mathrm{Mor}_{\mathsf{B}}\big(X, G(Y)\big), \quad X \in \mathsf{B},\ Y \in \mathsf{A},$$

there is a triple in B defined by $T := GF$. The unit of the adjunction $id \to GF$ defines the unit υ of the triple and the counit of the adjunction $FG \to id$ induces a natural transformation $GFGF \to GF$ which defines the multiplication μ. In fact, it is a theorem of Eilenberg and Moore [20] that all triples arise in this way from adjoint pairs. This is exactly the situation with the free operad and free non-unital operad functors that were described in Section 4. We will show how operads and non-unital operads can actually be defined using the concept of an algebra over a triple:

DEFINITION 39. A *T-algebra* or *algebra over the triple T* is an object A of $\mathcal{C}$ together with a structure morphism $\alpha : T(A) \to A$ satisfying

$$\alpha\big(T(\alpha)\big) = \alpha(\mu_A) \text{ and } \alpha\upsilon_A = id_A,$$

see Fig. 5.

The category of T-algebras in $\mathcal{C}$ will be denoted $\mathtt{Alg}_T(\mathcal{C})$. Since the free non-unital operad functor Ψ and the free operad functor Γ described in Section 4 are left adjoints to $\square_\psi : \psi\mathtt{Oper} \to \Sigma\text{-}\mathtt{mod}$ and $\square : \mathtt{Oper} \to \Sigma\text{-}\mathtt{mod}$, respectively, the functors $\square_\psi\Psi$ (denoted simply Ψ) and $\square\Gamma$ (denoted Γ) define triples on $\Sigma\text{-}\mathtt{mod}$.

THEOREM 40. *A Σ-module $\mathcal{S}$ is a Ψ-algebra if and only if it is a non-unital operad and it is a Γ-algebra if and only if it is an operad. In shorthand*:

$$\mathtt{Alg}_\Psi(\Sigma\text{-}\mathtt{mod}) \cong \psi\mathtt{Oper} \quad \textit{and} \quad \mathtt{Alg}_\Gamma(\Sigma\text{-}\mathtt{mod}) \cong \mathtt{Oper}.$$

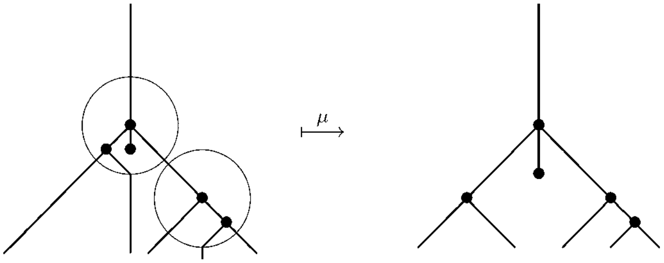
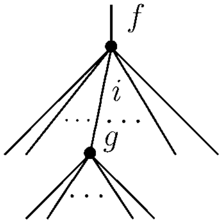


Fig. 6. Bracketed trees. The left picture shows an element of $\Psi\Psi(E)(5)$ while the right picture shows the same element interpreted, after erasing the braces indicated by thin cycles, as an element of $\Psi(E)(5)$. For simplicity, we did not show the decoration of vertices by elements of E.

PROOF. We outline first the proof of the implication in the direction from algebra to non-unital operad. Let $\mathcal{S}$ be a Ψ-algebra. The restriction of the structure morphism $\alpha : \Psi(\mathcal{S}) \to \mathcal{S}$ to the components of $\Psi(\mathcal{S})$ supported on trees with one internal edge defines the non-unital operad composition maps $\circ_i$, as indicated by



$$\overset{\alpha}{\longmapsto} \quad f \circ_i g.$$

In the opposite direction, for a non-unital operad $\mathcal{S}$, the Ψ-algebra structure $\alpha : \Psi(\mathcal{S}) \to \mathcal{S}$ is the contraction along the edges of underlying trees, using the $\circ_i$-operations. The proof that Γ-algebras are operads is similar. □

Let us change our perspective and consider formula (25) as defining an endofunctor $\Psi : \Sigma\text{-}\mathtt{mod} \to \Sigma\text{-}\mathtt{mod}$, ignoring that we already know that it represents free non-unital operads. We are going to construct maps

$$\mu : \Psi\Psi \to \Psi \quad \text{and} \quad \upsilon : id \to \Psi$$

making Ψ a triple on the category $\Sigma\text{-}\mathtt{mod}$. Let us start with the triple multiplication μ. It follows from (25) that, for each Σ-module E,

$$\Psi\Psi(E)(n) := \operatorname*{colim}_{T \in \mathtt{Tree}_n} \Psi(E)(T), \quad n \geqslant 0.$$

The elements in the right-hand side are represented by rooted trees T with vertices decorated by elements of $\Psi(E)$, while elements of $\Psi(E)$ are represented by rooted trees with vertices decorated by E. We may therefore imagine elements of $\Psi\Psi(E)$ as 'bracketed' rooted trees, in the sense indicated in Fig. 6. The triple multiplication $\mu_E : \Psi\Psi(E) \to \Psi(E)$ then simply erases the braces. The triple unit $\upsilon_E : E \to \Psi(E)$ identifies elements of

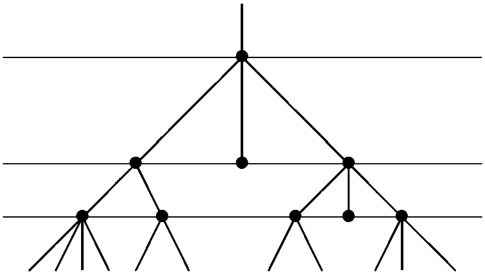
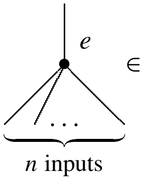


Fig. 7. A May tree.

E with decorated corollas:

$E(n) \ni e \quad \longleftrightarrow$  $\in \Psi(E)(n), \quad n \geqslant 0.$

It is not difficult to verify that the above constructions indeed make Ψ a triple, compare [83, §II.1.12]. Now we can *define* non-unital operads as algebras over the triple (Ψ, μ, υ). The advantage of this approach is that, by replacing $\mathtt{Tree}_n$ in (25) by another category of trees or graphs, one may obtain triples defining other types of operads and their generalizations.

We have already seen in (27) that enlarging $\mathtt{Tree}_n$ into $\mathtt{UTree}_n$ by adding the exceptional tree, one gets the triple Γ describing (unital) operads. It is not difficult to see that non-unital May operads are related to the category $\mathtt{MTree}_n$ of *May trees* which are, by definition, rooted trees whose vertices can be arranged into levels as in Fig. 7. Non-unital May operads are then algebras over the triple $\mathrm{M}\colon \Sigma\text{-}\mathtt{mod} \to \Sigma\text{-}\mathtt{mod}$ defined by

$$\mathrm{M}(E)(n) := \operatorname*{colim}_{T \in \mathtt{MTree}_n} E(T), \quad n \geqslant 0.$$

These observations are summarized in the first three lines of the table in Fig. 14 on p. 136.

6. Cyclic operads

In the following two sections we use the approach developed in Section 5 to introduce cyclic and modular operads. We recalled, in Example 14, the operad $\widehat{\mathfrak{M}}_0 = \{\widehat{\mathfrak{M}}_0(n)\}_{n \geqslant 0}$ of Riemann spheres with parametrized labeled holes. Each $\widehat{\mathfrak{M}}_0(n)$ was a right Σ_n-space, with the operadic right Σ_n-action permuting the labels $1, \dots, n$ of the holes $u_1, \dots, u_n$. But each $\widehat{\mathfrak{M}}_0(n)$ obviously admits a higher type of symmetry which interchanges the labels $0, \dots, n$ of *all* holes, including the label of the 'output' hole u_0. Another example admitting a similar higher symmetry is the configuration (non-unital) operad $\overline{\mathcal{M}}_0 = \{\overline{\mathcal{M}}_0(n)\}_{n \geqslant 2}$.

These examples indicate that, for some operads, there is no clear distinction between 'inputs' and the 'output'. Cyclic operads, introduced by E. Getzler and M.M. Kapranov in [32], formalize this phenomenon. They are, roughly speaking, operads with an extra symmetry that interchanges the output with one of the inputs. Let us recall some notions necessary to give a precise definition.

We remind the reader that in this section, as well as everywhere in this chapter, the main definitions are formulated over the underlying category of **k**-modules, where **k** is a commutative associative unital ring. However, for some constructions, we will require **k** to be a *field*; we will indicate this, as usual, by speaking about *vector spaces* instead of **k**-modules.

Let Σ_n^+ be the permutation group of the set $\{0, \dots, n\}$. The group Σ_n^+ is, of course, non-canonically isomorphic to the symmetric group Σ_{n+1}. We identify Σ_n with the subgroup of Σ_n^+ consisting of permutations $\sigma \in \Sigma_n^+$ such that $\sigma(0) = 0$. If $\tau_n \in \Sigma_n^+$ denotes the cycle $(0, \dots, n)$, that is, the permutation with $\tau_n(0) = 1,\ \tau_n(1) = 2,\ \dots, \tau_n(n) = 0$, then τ_n and Σ_n generate Σ_n^+.

Recall that a *cyclic Σ-module* or a *Σ^+-module* is a sequence $W = \{W(n)\}_{n \geqslant 0}$ such that each $W(n)$ is a (right) $\mathbf{k}[\Sigma_n^+]$-module. Let $\Sigma^+\text{-}\mathtt{mod}$ denote the category of cyclic Σ-modules. As (ordinary) operads were Σ-modules with an additional structure, cyclic operads are Σ^+-modules with an additional structure.

We will also need the following 'cyclic' analog of (23): if X is a set with $n + 1$ elements and $W \in \Sigma^+\text{-}\mathtt{mod}$, then

$$W(\!(X)\!) := W(n) \times_{\Sigma_n^+} \mathrm{Bij}\big([n]^+, X\big), \tag{30}$$

where $[n]^+ := \{0, \dots, n\}$, $n \geqslant 0$. The double brackets in $W(\!(X)\!)$ remind us that the n-th piece of the cyclic Σ-module $W = \{W(n)\}_{n \geqslant 0}$ is applied to a set with $n + 1$ elements, using the extended Σ_n^+-symmetry. Therefore

$$W\big(\!\big(\{0, \dots, n\}\big)\!\big) \cong W(n) \quad \text{while } W\big(\{0, \dots, n\}\big) \cong W(n+1), \quad n \geqslant 0.$$

Pasting schemes for cyclic operads are *cyclic* (*leg-*) *labeled n-trees*, by which we mean *unrooted* trees as on p. 109, with legs labeled by the set $\{0, \dots, n\}$. An example of such a tree is given in Fig. 8. Since we do not assume a choice of the root, the edges of a cyclic tree C are not directed and it does not make sense to speak about inputs and the output of a vertex $v \in \mathrm{vert}(C)$. Let $\mathtt{Tree}_n^+$ be the category of cyclic labeled n-trees and their bijections.

For a cyclic Σ-module W and a cyclic labeled tree T we have the following cyclic version of the product (24)

$$W(\!(T)\!) := \bigotimes_{v \in \mathrm{vert}(T)} W\big(\!\big(\mathrm{edge}(v)\big)\!\big).$$

The conceptual difference between (24) and the above formula is that instead of the set $\mathrm{in}(v)$ of incoming edges of a vertex v of a rooted tree, here we use the set $\mathrm{edge}(v)$ of *all* edges incident with v. Let, finally, $\Psi_+ : \Sigma^+\text{-}\mathtt{mod} \to \Sigma^+\text{-}\mathtt{mod}$ be the functor

$$\Psi_+(W)(n) := \operatorname*{colim}_{T \in \mathtt{Tree}_n^+} W(\!(T)\!), \quad n \geqslant 0, \tag{31}$$

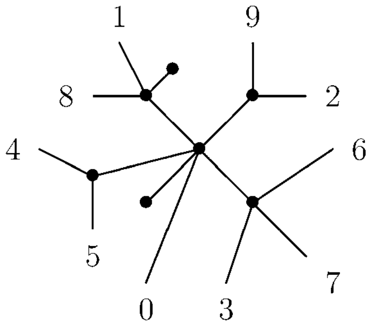


Fig. 8. A cyclic labeled tree from $\mathtt{Tree}_9^+$.

equipped with the triple structure of 'forgetting the braces' similar to that reviewed on p. 115. We will use also the 'extended' triple $\Gamma_+ : \Sigma^+\text{-}\mathtt{mod} \to \Sigma^+\text{-}\mathtt{mod}$,

$$\Gamma_+(W)(n) := \operatorname*{colim}_{T \in \mathtt{UTree}_n^+} W(\!(T)\!), \quad n \geqslant 0,$$

where $\mathtt{UTree}_n^+$ is the obvious extension of the category $\mathtt{Tree}_n^+$ by the exceptional tree $|$.

Definition 41. A *cyclic* (respectively *non-unital cyclic*) *operad* is an algebra over the triple Γ_+ (resppectively the triple Ψ_+) introduced above.

In the following proposition, which slightly improves [32, Theorem 2.2], $\tau_n \in \Sigma_n^+$ denotes the cycle $(0, \dots, n)$.

Proposition 42. *A non-unital cyclic operad is the same as a non-unital operad* $\mathcal{C} = \{\mathcal{C}(n)\}_{n \geqslant 0}$ (*Definition* 11) *such that the right* Σ_n*-action on* $\mathcal{C}(n)$ *extends, for each* $n \geqslant 0$, *to an action of* Σ_n^+ *with the property that for* $p \in \mathcal{C}(m)$ *and* $q \in \mathcal{C}(n)$, $1 \leqslant i \leqslant m$, $n \geqslant 0$, *the composition maps satisfy*

$$(p \circ_i q)\tau_{m+n-1} = \begin{cases} (q\tau_n) \circ_n (p\tau_m) & \text{if } i = 1, \\ (p\tau_m) \circ_{i-1} q, & \text{for } 2 \leqslant i \leqslant m. \end{cases}$$

The above structure is a (*unital*) *cyclic operad if moreover there exists a* Σ_1^+*-invariant operadic unit* $e \in \mathcal{C}(1)$.

Proposition 42 gives a biased definition of cyclic operads whose obvious modification (see [83, Definition II.5.2]) makes sense in an arbitrary symmetric monoidal category. We can therefore speak about topological cyclic operads, differential graded cyclic operads, simplicial cyclic operads &c. Observe that there are no *non-unital cyclic May operads* because it does not make sense to speak about levels in trees without a choice of the root.

Example 43. Let V be a finite-dimensional vector space and $B : V \otimes V \to \mathbf{k}$ a nondegenerate symmetric bilinear form. The form B induces an identification

$$\operatorname{Lin}(V^{\otimes n}, V) \ni f \mapsto \widehat{B}(f) := B\big(-, f(-)\big) \in \operatorname{Lin}(V^{\otimes(n+1)}, \mathbf{k})$$

of the spaces of linear maps. The standard right Σ_n^+-action

$$\widehat{B}(f)\sigma(v_0,\dots,v_n)=\widehat{B}(f)(v_{\sigma^{-1}(0)},\dots,v_{\sigma^{-1}(n)}),\quad \sigma\in\Sigma_n^+,\ v_0,\dots,v_n\in V,$$

defines, via this identification, a right Σ_n^+-action on $\mathrm{Lin}(V^{\otimes n},V)$, that is, on the n-th piece of the endomorphism operad $\mathcal{E}\mathrm{nd}_V=\{\mathcal{E}\mathrm{nd}_V(n)\}_{n\geqslant 0}$ recalled in Example 2. It is easy to show that, with the above action, $\mathcal{E}\mathrm{nd}_V$ is a cyclic operad in the monoidal category of vector spaces, called the *cyclic endomorphism operad* of the pair $V=(V,B)$. The biased definition of cyclic operads given in Proposition 42 can be read off from this example.

EXAMPLE 44. We saw in Example 7 that a unital operad $\mathcal{A}=\{\mathcal{A}(n)\}_{n\geqslant 0}$ such that $\mathcal{A}(n)=0$ for $n\neq 1$ is the same as a unital associative algebra. Similarly, it can be easily shown that a cyclic operad $\mathcal{C}=\{\mathcal{C}(n)\}_{n\geqslant 0}$ satisfying $\mathcal{C}(n)=0$ for $n\neq 1$ is the same as a unital associative algebra A with a linear involutive antiautomorphism, by which we mean a $\mathbf{k}$-linear map $^*\ :A\to A$ such that

$$(ab)^*=b^*a^*,\qquad (a^*)^*=a\quad\text{and}\quad 1^*=1,$$

for arbitrary $a,b\in A$.

Let $\mathcal{P}=\Gamma(E)/(R)$ be a quadratic operad as in Definition 37. The action of Σ_2 on E extends to an action of Σ_2^+, via the sign representation $\mathrm{sgn}:\Sigma_2^+\to\{\pm 1\}=\Sigma_2$. It can be easily verified that this action induces a cyclic operad structure on the free operad $\Gamma(E)$. In particular, $\Gamma(E)(3)$ is a right Σ_3^+-module.

DEFINITION 45. We say that the operad $\mathcal{P}$ is a *cyclic quadratic operad* if, in the above presentation, R is a Σ_3^+-invariant subspace of $\Gamma(E)(3)$.

If the condition of the above definition is satisfied, $\mathcal{P}$ has a natural induced cyclic operad structure.

EXAMPLE 46. By [32, Proposition 3.6], all quadratic operads generated by a one-dimensional space are cyclic quadratic, therefore the operads $\mathcal{L}\mathrm{ie}$ and $\mathcal{C}\mathrm{om}$ are cyclic quadratic. Also the operads $\mathcal{A}\mathrm{ss}$ and the operad $\mathcal{P}\mathrm{oiss}$ for Poisson algebras are cyclic quadratic [32, Proposition 3.11]. A surprisingly simple operad which is cyclic and quadratic, but not cyclic quadratic, is constructed in [82, Remark 15].

The operad $\widehat{\mathfrak{M}}_0$ of Riemann spheres with labeled punctures reviewed in Example 14 is a topological cyclic operad. The configuration operad $\overline{\mathcal{M}}_0$ recalled on p. 100 is a non-unital topological cyclic operad. Important examples of non-cyclic operads are the operad pre-Lie for pre-Lie algebras [82, Section 3] and the operad $\mathcal{L}\mathrm{eib}$ for Leibniz algebras [32, §3.15].

Let $\mathcal{C}$ be an operad, $\alpha:\mathcal{C}(n)\otimes V^{\otimes n}\to V$, $n\geqslant 0$, a $\mathcal{C}$-algebra with the underlying vector space V as in Proposition 24 and $B:V\otimes V\to U$ a bilinear form on V with values in a vector space U. We can form a map

$$\widetilde{B}(\alpha):\mathcal{C}(n)\otimes V^{\otimes(n+1)}\to U,\quad n\geqslant 0,\tag{32}$$

by the formula

$$\widetilde{B}(\alpha)(c \otimes v_0 \otimes \cdots \otimes v_n) := B\big(v_0, \alpha(c \otimes v_1 \otimes \cdots \otimes v_n)\big),$$

$c \in \mathcal{C}(n)$, $v_0, \ldots, v_n \in V$. Suppose now that the operad $\mathcal{C}$ is cyclic, in particular, that each $\mathcal{C}(n)$ is a right Σ_n^+-module. We say that the bilinear form $B : V \otimes V \to U$ is *invariant* [32, Definition 4.1], if the maps $\widetilde{B}(\alpha)$ in (32) are, for each $n \geqslant 0$, invariant under the diagonal action of Σ_n^+ on $\mathcal{C}(n) \otimes V^{\otimes(n+1)}$. We leave as an exercise to verify that the invariance of $\widetilde{B}(\alpha)$ for $n = 1$ together with the existence of the operadic unit implies that B is symmetric,

$$B(v_0, v_1) = B(v_1, v_0), \quad v_0, v_1 \in V.$$

Definition 47. A *cyclic algebra* over a cyclic operad $\mathcal{C}$ is a $\mathcal{C}$-algebra structure on a vector space V together with a nondegenerate invariant bilinear form $B : V \otimes V \to \mathbf{k}$.

By [83, Proposition II.5.14], a cyclic algebra is the same as a cyclic operad homomorphism $\mathcal{C} \to \mathcal{E}nd_V$, where $\mathcal{E}nd_V$ is the cyclic endomorphism operad of the pair (V, B) recalled in Example 43.

Example 48. A cyclic algebra over the cyclic operad $\mathcal{C}om$ is a *Frobenius algebra*, that is, a structure consisting of a commutative associative multiplication $\cdot : V \otimes V \to V$ as in Example 25 together with a non-degenerate symmetric bilinear form $B : V \otimes V \to \mathbf{k}$, invariant in the sense that

$$B(a \cdot b, c) = B(a, b \cdot c), \quad \text{for all } a, b, c \in V.$$

Similarly, a cyclic Lie algebra is given by a Lie bracket $[-,-] : V \otimes V \to V$ and a non-degenerate symmetric bilinear form $B : V \otimes V \to \mathbf{k}$ satisfying

$$B\big([a, b], c\big) = B\big(a, [b, c]\big), \quad \text{for } a, b, c \in V.$$

For algebras over cyclic operads, one may introduce cyclic cohomology that generalizes the classical cyclic cohomology of associative algebras [12,66,109] as the non-Abelian derived functor of the universal bilinear form [32], [83, Proposition II.5.26]. Let us close this section by mentioning two examples of operads with other types of higher symmetries. The symmetry required for *anticyclic operads* differs from the symmetry of cyclic operads by the sign [83, Definition II.5.20]. *Dihedral operads* exhibit a symmetry governed by the dihedral groups [82, Definition 16].

7. Modular operads

Let us consider again the Σ^+-module $\widehat{\mathfrak{M}}_0 = \{\widehat{\mathfrak{M}}_0(n)\}_{n \geqslant 0}$ of Riemann spheres with punctures. We saw that the operation $M, N \mapsto M \circ_i N$ of sewing the 0-th hole of the surface N to the i-th hole of the surface M defined on $\widehat{\mathfrak{M}}_0$ a cyclic operad structure. One may generalize this operation by defining, for $M \in \widehat{\mathfrak{M}}_0(m)$, $N \in \widehat{\mathfrak{M}}_0(n)$, $0 \leqslant i \leqslant m$, $0 \leqslant j \leqslant n$, the element $M_i \circ_j N \in \widehat{\mathfrak{M}}_0(m+n-1)$ by sewing the j-th hole of M to the i-th hole of N. Under this notation, $\circ_i = {}_i\circ_0$. In the same manner, one may consider a single surface $M \in \widehat{\mathfrak{M}}_0(n)$, choose labels i, j, $0 \leqslant i \neq j \leqslant n$, and sew the i-th hole of M along the j-th hole of the *same* surface. The result is a new surface $\xi_{\{i,j\}}(M)$, with $n-2$ holes and genus 1.

This leads us to the system $\widehat{\mathfrak{M}} = \{\widehat{\mathfrak{M}}(g,n)\}_{g\geqslant 0, n\geqslant -1}$, where $\widehat{\mathfrak{M}}(g,n)$ denotes now the moduli space of genus g Riemann surfaces with $n+1$ holes. Observe that we include $\widehat{\mathfrak{M}}(g,n)$ also for $n=-1$; $\widehat{\mathfrak{M}}(g,-1)$ is the moduli space of Riemann surfaces of genus g. The operations ${}_i\circ_j$ and $\xi_{\{i,j\}}$ act on $\widehat{\mathfrak{M}}$. Clearly, for $M \in \widehat{\mathfrak{M}}(g,m)$ and $N \in \widehat{\mathfrak{M}}(h,n)$, $0 \leqslant i \leqslant m$, $0 \leqslant j \leqslant n$ and $g, h \geqslant 0$,

$$M_i \circ_j N \in \widehat{\mathfrak{M}}(g+h, m+n-1) \tag{33}$$

and, for $m \geqslant 1$ and $g \geqslant 0$,

$$\xi_{\{i,j\}}(M) \in \widehat{\mathfrak{M}}(g+1, m-2). \tag{34}$$

A particular case of (33) is the non-operadic composition

$${}_0\circ_0 : \widehat{\mathfrak{M}}(g,0) \times \widehat{\mathfrak{M}}(h,0) \to \widehat{\mathfrak{M}}(g+h,-1), \quad g, h \geqslant 0. \tag{35}$$

Modular operads are abstractions of the above structure satisfying a certain additional stability condition. The following definitions, taken from [33], are made for the category of $\mathbf{k}$-modules, but they can be easily generalized to an arbitrary symmetric monoidal category with finite colimits, whose monoidal product $\odot$ is distributive over colimits. Let us introduce the underlying category for modular operads.

A *modular Σ-module* is a sequence $\mathcal{E} = \{\mathcal{E}(g,n)\}_{g\geqslant 0, n\geqslant -1}$ of $\mathbf{k}$-modules such that each $\mathcal{E}(g,n)$ has a right $\mathbf{k}[\Sigma_n^+]$-action. We say that $\mathcal{E}$ is *stable* if

$$\mathcal{E}(g,n) = 0 \quad \text{for } 2g+n-1 \leqslant 0 \tag{36}$$

and denote $\mathtt{MMod}$ the category of stable modular Σ-modules.

Stability (36) says that $\mathcal{E}(g,n)$ is trivial for $(g,n) = (0,-1),\ (1,-1),\ (0,0)$ and $(0,1)$. We will sometimes express the stability of $\mathcal{E}$ by writing $\mathcal{E} = \{\mathcal{E}(g,n)\}_{(g,n)\in\mathfrak{S}}$, where

$$\mathfrak{S} := \big\{(g,n) \mid g \geqslant 0,\ n \geqslant -1 \text{ and } 2g+n-1 > 0\big\}.$$

Recall that a genus g Riemann surface with k marked points is stable if it does not admit infinitesimal automorphisms. This happens if and only if $2(g-1)+k > 0$, that is, excluded is the torus with no marked points and the sphere with less than three marked points. Thus the stability property of modular Σ-modules is analogous to the stability of Riemann surfaces.

Now we introduce graphs that serve as pasting schemes for modular operads. The naive notion of a graph as we have used it up to this point is not subtle enough; we need to replace it by a more sophisticated notion:

DEFINITION 49. A *graph* Γ is a finite set $\mathrm{Flag}(\Gamma)$ (whose elements are called *flags* or *half-edges*) together with an involution σ and a partition λ. The *vertices* $\mathrm{vert}(\Gamma)$ of a graph Γ are the blocks of the partition λ, we assume also that the number of these blocks is finite.

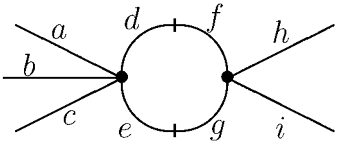


Fig. 9. The sputnik.

The *edges* $\mathrm{Edg}(\Gamma)$ are pairs of flags forming a two-cycle of σ. The *legs* $\mathrm{Leg}(\Gamma)$ are the fixed points of σ.

We also denote by $\mathrm{edge}(v)$ the flags belonging to the block v or, in common speech, half-edges adjacent to the vertex v. We say that graphs Γ_1 and Γ_2 are *isomorphic* if there exists a set isomorphism $\varphi : \mathrm{Flag}(\Gamma_1) \to \mathrm{Flag}(\Gamma_2)$ that preserves the partitions and commutes with the involutions. We may associate to a graph Γ a finite one-dimensional cell complex $|\Gamma|$, obtained by taking one copy of $[0, \frac{1}{2}]$ for each flag, a point for each block of the partition, and imposing the following equivalence relation: the points $0 \in [0, \frac{1}{2}]$ are identified for all flags in a block of the partition λ with the point corresponding to the block, and the points $\frac{1}{2} \in [0, \frac{1}{2}]$ are identified for pairs of flags exchanged by the involution σ.

We call $|\Gamma|$ the *geometric realization* of Γ. Observe that empty blocks of the partition generate isolated vertices in the geometric realization. We will usually make no distinction between a graph and its geometric realization. As an example (taken from [33]), consider the graph with $\{a, b, \dots, i\}$ as the set of flags, the involution $\sigma = (df)(eg)$ and the partition $\{a, b, c, d, e\} \cup \{f, g, h, i\}$. The geometric realization of this graph is the 'sputnik' in Fig. 9.

Let us introduce labeled versions of the above notions. A (*vertex-*) *labeled graph* is a connected graph Γ together with a map g (the *genus map*) from $\mathrm{vert}(\Gamma)$ to the set $\{0, 1, 2, \dots\}$. The labeled graphs Γ_1 and Γ_2 are isomorphic if there exists an isomorphism preserving the labels of the vertices. The *genus* $g(\Gamma)$ of a labeled graph Γ is defined by

$$g(\Gamma) := b_1(\Gamma) + \sum_{v \in \mathrm{vert}(\Gamma)} g(v), \tag{37}$$

where $b_1(\Gamma) := \dim H_1(|\Gamma|)$ is the first Betti number of the graph $|\Gamma|$, i.e. the number of independent circuits of Γ. A graph Γ is *stable* if

$$2\big(g(v) - 1\big) + \big|\mathrm{edge}(v)\big| > 0,$$

at each vertex $v \in \mathrm{vert}(\Gamma)$.

For $g \geqslant 0$ and $n \geqslant -1$, let $\mathtt{MGr}(g, n)$ be the groupoid whose objects are pairs (Γ, ℓ) consisting of a stable (vertex-) labeled graph Γ of genus g and an isomorphism $\ell : \mathrm{Leg}(\Gamma) \to \{0, \dots, n\}$ labeling the legs of Γ by elements of $\{0, \dots, n\}$. The morphisms of $\mathtt{MGr}(g, S)$ are isomorphisms of vertex-labeled graphs preserving the labeling of the legs. The stability implies, via elementary combinatorial topology that, for each fixed $g \geqslant 0$ and $n \geqslant -1$, there is only a finite number of isomorphism classes of stable graphs $\Gamma \in \mathtt{MGr}(g, n)$, see [33, Lemma 2.16].

We will also need the following obvious generalization of (30): if $\mathcal{E} = \{\mathcal{E}(g,n)\}_{g \geqslant 0, n \geqslant -1}$ is a modular Σ-module and X a set with $n+1$ elements, then

$$\mathcal{E}(\!(g, X)\!) := \mathcal{E}(g,n) \times_{\Sigma_n^+} \mathrm{Bij}\big([n]^+, X\big), \quad g \geqslant 0,\ n \geqslant -1. \tag{38}$$

For a modular Σ-module $\mathcal{E} = \{\mathcal{E}(g,n)\}_{g \geqslant 0, n \geqslant -1}$ and a labeled graph Γ, let $\mathcal{E}(\!(\Gamma)\!)$ be the product

$$\mathcal{E}(\!(\Gamma)\!) := \bigotimes_{v \in \mathrm{vert}(\Gamma)} \mathcal{E}\big(\!\big(g(v), \mathrm{edge}(v)\big)\!\big). \tag{39}$$

Evidently, the correspondence $\Gamma \mapsto \mathcal{E}(\!(\Gamma)\!)$ defines a functor from the category $\mathtt{MGr}(g,n)$ to the category of $\mathbf{k}$-modules and their isomorphisms. We may thus define an endofunctor $\mathbb{M}$ on the category $\mathtt{MMod}$ of stable modular Σ-modules by the formula

$$\mathbb{M}\mathcal{E}(g,n) := \operatorname*{colim}_{\Gamma \in \mathtt{MGr}(g,n)} \mathcal{E}(\!(\Gamma)\!), \quad g \geqslant 0,\ n \geqslant -1.$$

Choosing a representative for each isomorphism class in $\mathtt{MGr}(g,n)$, one obtains the identification

$$\mathbb{M}\mathcal{E}(g,n) \cong \bigoplus_{[\Gamma] \in \{\mathtt{MGr}(g,n)\}} \mathcal{E}(\!(\Gamma)\!)_{\mathrm{Aut}(\Gamma)}, \quad g \geqslant 0,\ n \geqslant -1, \tag{40}$$

where $\{\mathtt{MGr}(g,n)\}$ is the set of isomorphism classes of objects of the groupoid $\mathtt{MGr}(g,n)$ and the subscript $\mathrm{Aut}(\Gamma)$ denotes the space of coinvariants. Stability (36) implies that the summation in the right-hand side of (40) is finite. Formula (40) generalizes (26) which does not contain coinvariants because there are no nontrivial automorphisms of leaf-labeled trees. On the other hand, stable labeled graphs with nontrivial automorphisms are abundant, an example can be easily constructed from the graph in Fig. 9. The functor $\mathbb{M}$ carries a triple structure of 'erasing the braces' similar to the one used on pp. 115 and 118.

Definition 50. A *modular operad* is an algebra over the triple $\mathbb{M} : \mathtt{MMod} \to \mathtt{MMod}$.

Therefore a modular operad is a stable modular Σ-module $\mathcal{A} = \{\mathcal{A}(g,n)\}_{(g,n) \in \mathfrak{S}}$ equipped with operations that determine coherent contractions along stable modular graphs. Observe that the stability condition is built firmly into the very definition. Very crucially, modular operads *do not have units*, because such a unit ought to be an element of the space $\mathcal{A}(0,1)$ which is empty, by (36).

One can easily introduce un-stable modular operads and their unital versions, but the main motivating example reviewed below is stable. We will consider an extension of the Grothendieck–Knudsen configuration operad $\overline{\mathcal{M}}_0 = \{\overline{\mathcal{M}}_0(n)\}_{n \geqslant 2}$ consisting of moduli spaces of stable curves of arbitrary genera in the sense of the following generalization of Definition 15.

Definition 51. A *stable* $(n+1)$*-pointed curve*, $n \geqslant 0$, is a connected complex projective curve C with at most nodal singularities, together with a 'marking' given by a choice $x_0, \dots, x_n \in C$ of smooth points. Stability means, as usual, that there are no infinitesimal automorphisms of C fixing the marked points and double points.

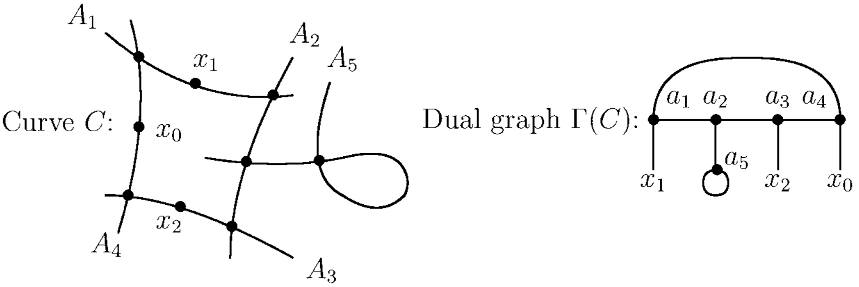


Fig. 10. A stable curve and its dual graph. The curve C on the left has five components A_i, $1 \leqslant i \leqslant 5$, and three marked points x_0, x_1 and x_2. The dual graph $\Gamma(C)$ on the right has five vertices a_i, $1 \leqslant i \leqslant 5$, corresponding to the components of the curve and three legs labeled by the marked points.

The stability in Definition 51 is equivalent to saying that each smooth component of C isomorphic to complex projective space $\mathbb{CP}^1$ has at least three special points and that each smooth component isomorphic to the torus has at least one special point, where by a special point we mean either a double point or a node.

The *dual graph* $\Gamma = \Gamma(C)$ of a stable $(n+1)$-pointed curve $C = (C, x_0, \dots, x_n)$ is a labeled graph whose vertices are the components of C, edges are the nodes and its legs are the points $\{x_i\}_{0 \leqslant i \leqslant n}$. An edge e_y corresponding to a nodal point y joins the vertices corresponding to the components intersecting at y. The vertex v_K corresponding to a branch K is labeled by the genus of the normalization of K. See [37, p. 23] for the normalization and recall that a curve is normal if and only if it is nonsingular. The construction of $\Gamma(C)$ from a curve C is visualized in Fig. 10.

Let us denote by $\overline{\mathcal{M}}_{g,n+1}$ the coarse moduli space [37, p. 347] of stable $(n+1)$-pointed curves C such that the dual graph $\Gamma(C)$ has genus g, in the sense of (37). The genus of $\Gamma(C)$ in fact equals the arithmetic genus of the curve C, thus $\overline{\mathcal{M}}_{g,n+1}$ is the coarse moduli space of stable curves of arithmetic genus g with $n+1$ marked points. By a result of P. Deligne, F.F. Knudsen and D. Mumford [18,51,50], $\overline{\mathcal{M}}_{g,n+1}$ is a projective variety.

Observe that, for a curve $C \in \overline{\mathcal{M}}_{0,n+1}$, the graph $\Gamma(C)$ must necessarily be a tree and all components of C must be smooth of genus 0, therefore $\overline{\mathcal{M}}_{0,n+1}$ coincides with the moduli space $\overline{\mathcal{M}}_0(n)$ of genus 0 stable curves with $n+1$ marked points that we discussed in Section 2. Dual graphs of curves $C \in \overline{\mathcal{M}}_{g,n+1}$ are stable labeled graphs belonging to $\mathtt{MGr}(g, n+1)$.

The symmetric group Σ_n^+ acts on $\overline{\mathcal{M}}_{g,n+1}$ by renumbering the marked points, therefore

$$\overline{\mathcal{M}} := \left\{\overline{\mathcal{M}}(g,n)\right\}_{g \geqslant 0, n \geqslant -1},$$

with $\overline{\mathcal{M}}(g,n) := \overline{\mathcal{M}}_{g,n+1}$, is a modular Σ-module in the category of projective varieties. Since there are no stable curves of genus g with $n+1$ punctures if $2g+n-1 \leqslant 0$, $\overline{\mathcal{M}}$ is a *stable* modular Σ-module. Let us define the contraction along a stable graph $\Gamma \in \mathtt{MGr}(g,n)$

$$\alpha_\Gamma : \overline{\mathcal{M}}(\!(\Gamma)\!) = \prod_{v \in \mathrm{vert}(\Gamma)} \overline{\mathcal{M}}\big(\!\big(g(v), \mathrm{edge}(v)\big)\!\big) \to \overline{\mathcal{M}}(g,n) \tag{41}$$

by gluing the marked points of curves from $\overline{\mathcal{M}}(\!(g(v), \mathrm{edge}(v))\!)$, $v \in \mathrm{vert}(\Gamma)$, according to the graph Γ. To be more precise, let

$$\prod_{v \in \mathrm{vert}(\Gamma)} C_v, \quad \text{where } C_v \in \overline{\mathcal{M}}\big(\!\big(g(v), \mathrm{edge}(v)\big)\!\big),$$

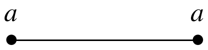
be an element of $\overline{\mathcal{M}}(\!(\Gamma)\!)$. Let e be an edge of the graph Γ connecting vertices v_1 and v_2, $e = \{y^e_{v_1}, y^e_{v_2}\}$, where $y^e_{v_i}$ is a marked point of the component C_{v_i}, $i = 1, 2$, which is also the name of the corresponding flag of the graph Γ. The curve $\alpha_\Gamma(C)$ is then obtained by the identifications $y^e_{v_1} = y^e_{v_2}$, introducing a nodal singularity, for all $e \in \mathrm{Edg}(\Gamma)$. The procedure is the same as that described for the tree level in Section 2. As proved in [33, §6.2], the contraction maps (41) define on the stable modular Σ-module of coarse moduli spaces $\overline{\mathcal{M}} = \{\overline{\mathcal{M}}(g,n)\}_{(g,n)\in\mathfrak{S}}$ a modular operad structure in the category of complex projective varieties.

Let us look more closely at the structure of the modular triple $\mathbb{M}$. Given a (stable or unstable) modular Σ-module $\mathcal{E}$, there is, for each $g \geqslant 0$ and $n \geqslant -1$, a natural decomposition

$$\mathbb{M}(\mathcal{E})(g,n) = \mathbb{M}_0(\mathcal{E})(g,n) \oplus \mathbb{M}_1(\mathcal{E})(g,n) \oplus \mathbb{M}_2(\mathcal{E})(g,n) \oplus \cdots,$$

with $\mathbb{M}_k(\mathcal{E})(g,n)$ the subspace obtained by summing over graphs Γ with $\dim H_1(|\Gamma|) = k$, $k \geqslant 0$. In particular, $\mathbb{M}_0(\mathcal{E})(g,n)$ is a summation over simply connected graphs. It is not difficult to see that $\mathbb{M}_0(\mathcal{E})$ is a *subtriple* of $\mathbb{M}(\mathcal{E})$. This shows that modular operads are $\mathbb{M}_0$-algebras with some additional operations (the 'contractions') that raise the genus and generate the higher components $\mathbb{M}_k$, $k \geqslant 1$, of the modular triple $\mathbb{M}$.

There seems to be a belief expressed in the proof of [33, Lemma 3.4] and also in [33, Theorem 3.7] that, in the stable case, the triple $\mathbb{M}_0$ is equivalent to the non-unital cyclic operad triple Ψ_+, but it is not so. The triple $\mathbb{M}_0$ is *much bigger*, for example, if $a \in \mathcal{E}(1,0)$, then $\mathbb{M}_0(\mathcal{E})(2,-1)$ contains a non-operadic element



which can be also written, using (35), as $a \;{}_0\!\circ_0\; a$. The corresponding part $\Psi_+(\mathcal{E})(-1)$ of the cyclic triple is empty. In the Grothendieck–Knudsen modular operad $\overline{\mathcal{M}}$, an element of the above type is realized by two tori meeting at a nodal point.

On the other hand, the triple $\mathbb{M}_0$ restricted to the subcategory of stable modular Σ-modules $\mathcal{E}$ such that $\mathcal{E}(g,n) = 0$ for $g > 0$ indeed coincides with the non-unital cyclic operad triple Ψ_+, as was in fact proved in [33, p. 81]. Therefore, given a modular operad $\mathcal{A} = \{\mathcal{A}(g,n)\}_{(g,n)\in\mathfrak{S}}$, there is an induced non-unital cyclic operad structure on the cyclic collection $\mathcal{A}^\flat := \{\mathcal{A}(0,n)\}_{n\geqslant 2}$. We will call $\mathcal{A}^\flat$ the *associated cyclic operad*. For example, the cyclic operad associated to the Grothendieck–Knudsen modular operad $\overline{\mathcal{M}}$ equals its genus zero part $\overline{\mathcal{M}}_0$.

A *biased* definition of modular operads can be found in [83, Definition II.5.35]. It is formulated in terms of operations

$$\big\{{}_i\circ_j : \mathcal{A}(g,m) \otimes \mathcal{A}(h,n) \to \mathcal{A}(g+h, m+n);\ 0 \leqslant i \leqslant m,\ 0 \leqslant j \leqslant n,\ g,h \geqslant 0\big\}$$

together with contractions

$$\left\{\xi_{\{i,j\}} : \mathcal{A}(g,m) \to \mathcal{A}(g+1, m-2);\ m \geqslant 1,\ g \geqslant 0\right\}$$

that generalize (33) and (34).

EXAMPLE 52. Let $V = (V, B)$ be a vector space with a symmetric inner product $B : V \otimes V \to \mathbf{k}$. Denote, for each $g \geqslant 0$ and $n \geqslant -1$,

$$\mathcal{E}\mathrm{nd}_V(g,n) := V^{\otimes(n+1)}.$$

It is clear from definition (39) that, for any labeled graph $\Gamma \in \mathtt{MGr}(g,n)$, $\mathcal{E}\mathrm{nd}_V\,(\!(\Gamma)\!) = V^{\otimes \mathrm{Flag}(\Gamma)}$.

Let $B^{\otimes \mathrm{Edg}(\Gamma)} : V^{\otimes \mathrm{Flag}(\Gamma)} \to V^{\otimes \mathrm{Leg}(\Gamma)}$ be the multilinear form which contracts the factors of $V^{\otimes \mathrm{Flag}(\Gamma)}$ corresponding to the flags which are paired up as edges of Γ. Then we define $\alpha_\Gamma : \mathcal{E}\mathrm{nd}_V\,(\!(\Gamma)\!) \to \mathcal{E}\mathrm{nd}_V(g,n)$ to be the map

$$\alpha_\Gamma : \mathcal{E}\mathrm{nd}_V\,(\!(\Gamma)\!) = V^{\otimes \mathrm{Flag}(\Gamma)} \xrightarrow{B^{\otimes \mathrm{Edg}(\Gamma)}} V^{\otimes \mathrm{Leg}(\Gamma)} \xrightarrow{V^{\otimes \ell}} V^{\otimes(n+1)} = \mathcal{E}\mathrm{nd}_V(g,n),$$

where $\ell : \mathrm{Leg}(\Gamma) \to \{0, \dots, n\}$ is the labeling of the legs of Γ. It is easy to show that the compositions $\{\alpha_\Gamma;\ \Gamma \in \mathtt{MGr}(g,n)\}$ define on $\mathcal{E}\mathrm{nd}_V$ the structure of an un-stable unital modular operad, see [33, §2.25].

An *algebra* over a modular operad $\mathcal{A}$ is a vector space V with an inner product B, together with a morphism $\rho : \mathcal{A} \to \mathcal{E}\mathrm{nd}_V$ of modular operads. Several important structures are algebras over modular operads. For example, an algebra over the homology $H_*(\overline{\mathcal{M}})$ of the Grothendieck–Knudsen modular operad is the same as a cohomological field theory in the sense of [55]. Other physically relevant algebras over modular operads can be found in [33,78,83]. Relations between modular operads, chord diagrams and Vassiliev invariants are studied in [42].

8. PROPs

Operads are devices invented to describe structures consisting of operations with several inputs and *one* output. There are, however, important structures with operations having several inputs and *several* outputs. Let us recall the most prominent one.

EXAMPLE 53. A (associative) *bialgebra* is a $\mathbf{k}$-module V with a *multiplication* $\mu : V \otimes V \to V$ and a *comultiplication* (also called a *diagonal*) $\Delta : V \to V \otimes V$. The multiplication is associative:

$$\mu(\mu \otimes id_V) = \mu(id_V \otimes \mu),$$

the comultiplication is coassociative:

$$(\Delta \otimes id_V)\Delta = (id_V \otimes \Delta)\Delta$$

and the usual compatibility between μ and Δ is assumed:

$$\Delta(u \cdot v) = \Delta(u) \cdot \Delta(v) \quad \text{for } u, v \in V, \tag{42}$$

where $u \cdot v := \mu(u, v)$ and the dot $\cdot$ in the right-hand side denotes the multiplication induced on $V \otimes V$ by μ. Loosely speaking, bialgebras are Hopf algebras without unit, counit and antipode.

PROPs (an abbreviation of **pro**duct and **p**ermutation category) describe structures as in Example 53. Although PROPs are more general than operads, they appeared much earlier, in a 1965 paper of Mac Lane [68]. This might be explained by the fact that the definition of PROPs is more compact than that of operads – compare Definition 54 below with Definition 1 in Section 1. PROPs then entered the 'renaissance of operads' in 1996 via [73].

Definition 54 uses the notion of a symmetric strict monoidal category which we consider so basic and commonly known that we will not recall it, standard citations are [68,67], see also [83, §II.1.1]. An example is the category $\mathtt{Mod}_{\mathbf{k}}$ of $\mathbf{k}$-modules, with the monoidal product $\odot$ given by the tensor product $\otimes = \otimes_{\mathbf{k}}$, the symmetry $S_{U,V} : U \otimes V \to V \otimes U$ defined as $S_{U,V}(u, v) := v \otimes u$ for $u \in U$ and $v \in V$, and the unit $\mathtt{1}$ the ground ring $\mathbf{k}$.

Definition 54. A ($\mathbf{k}$-linear) *PROP* (called a *theory* in [73]) is a symmetric strict monoidal category $\mathsf{P} = (\mathsf{P}, \odot, S, \mathtt{1})$ enriched over $\mathtt{Mod}_{\mathbf{k}}$ such that:

(i) the objects are indexed by (or identified with) the set $\mathbb{N} = \{0, 1, 2, \dots\}$ of natural numbers, and

(ii) the product satisfies $m \odot n = m + n$, for any $m, n \in \mathbb{N} = \mathrm{Ob}(\mathsf{P})$ (hence the unit $\mathtt{1}$ equals 0).

Recall that the $\mathtt{Mod}_{\mathbf{k}}$-enrichment in the above definition means that each hom-set $\mathrm{Mor}_{\mathsf{P}}(m, n)$ is a $\mathbf{k}$-module and the operations of the monoidal category P (the composition $\circ$, the product $\odot$ and the symmetry S) are compatible with this $\mathbf{k}$-linear structure.

For a PROP P denote $\mathsf{P}(m, n) := \mathrm{Mor}_{\mathsf{P}}(m, n)$. The symmetry S induces, via the canonical identifications $m \cong \mathtt{1}^{\odot m}$ and $n \cong \mathtt{1}^{\odot n}$, on each $\mathsf{P}(m, n)$ a structure of (Σ_m, Σ_n)-bimodule (left Σ_m–right Σ_n-module such that the left action commutes with the right one). Therefore a PROP is a collection $\mathsf{P} = \{\mathsf{P}(m, n)\}_{m,n \geqslant 0}$ of (Σ_m, Σ_n)-bimodules, together with two types of compositions, *horizontal*

$$\otimes : \mathsf{P}(m_1, n_1) \otimes \cdots \otimes \mathsf{P}(m_s, n_s) \to \mathsf{P}(m_1 + \cdots + m_s, n_1 + \cdots + n_s),$$

induced, for all $m_1, \dots, m_s, n_1, \dots, n_s \geqslant 0$, by the monoidal product $\odot$ of P, and *vertical*

$$\circ : \mathsf{P}(m, n) \otimes \mathsf{P}(n, k) \to \mathsf{P}(m, k),$$

given, for all $m, n, k \geqslant 0$, by the categorial composition. The monoidal unit is an element $e := \mathtt{1} \in \mathsf{P}(1, 1)$. In Definition 54, $\mathtt{Mod}_{\mathbf{k}}$ can be replaced by an arbitrary symmetric strict monoidal category.

Let $\mathsf{P} = \{\mathsf{P}(m, n)\}_{m,n \geqslant 0}$ and $\mathsf{Q} = \{\mathsf{Q}(m, n)\}_{m,n \geqslant 0}$ be two PROPs. A *homomorphism* $f : \mathsf{P} \to \mathsf{Q}$ is a sequence $f = \{f(m, n) : \mathsf{P}(m, n) \to \mathsf{Q}(m, n)\}_{m,n \geqslant 0}$ of bi-equivariant maps which commute with both the vertical and horizontal compositions. An *ideal* in a PROP P is a system $\mathsf{I} = \{\mathsf{I}(m,n)\}_{m,n \geqslant 0}$ of left Σ_m–right Σ_n-invariant subspaces $\mathsf{I}(m,n) \subset \mathsf{P}(m,n)$ which is closed, in the obvious sense, under both the vertical and horizontal compositions. Kernels, images, &c., of homomorphisms between PROPs, as well as quotients of PROPs by PROPic ideals, are defined componentwise, see [73,111–113] for details.

EXAMPLE 55. The *endomorphism* PROP of a $\mathbf{k}$-module V is the system

$$\mathcal{E}\mathrm{nd}_V = \left\{\mathcal{E}\mathrm{nd}_V(m,n)\right\}_{m,n \geqslant 0}$$

with $\mathcal{E}\mathrm{nd}_V(m,n)$ the space of linear maps $\mathrm{Lin}(V^{\otimes n}, V^{\otimes m})$ with n ‘inputs’ and m ‘outputs’, $e \in \mathcal{E}\mathrm{nd}_V(1,1)$ the identity map, horizontal composition given by the tensor product of linear maps, and vertical composition by the ordinary composition of linear maps.

Also algebras over PROPs can be introduced in a very concise way.

DEFINITION 56. A P-*algebra* is a strict symmetric monoidal functor $\lambda : \mathsf{P} \to \mathtt{Mod}_{\mathbf{k}}$ of enriched monoidal categories. The value $\lambda(1)$ is the *underlying space* of the algebra ρ.

It is easy to see that a P-algebra is the same as a PROP homomorphism $\rho : \mathsf{P} \to \mathcal{E}\mathrm{nd}_V$. As in Proposition 24, a P-algebra is determined by a system

$$\alpha : \mathsf{P}(m,n) \otimes V^{\otimes n} \to V^{\otimes m}, \quad m, n, \geqslant 0,$$

of linear maps satisfying appropriate axioms.

As before, the first step in formulating an unbiased definition of PROPs is to specify their underlying category. A Σ-*bimodule* is a system $E = \{E(m,n)\}_{m,n \geqslant 0}$ such that each $E(m,n)$ is a left $\mathbf{k}[\Sigma_m]$- right $\mathbf{k}[\Sigma_n]$-bimodule. Let $\Sigma\text{-}\mathtt{bimod}$ denote the category of Σ-bimodules. For $E \in \Sigma\text{-}\mathtt{bimod}$ and finite sets Y, X with m, respectively n, elements put

$$E(Y,X) := \mathrm{Bij}\big(Y,[m]\big) \times_{\Sigma_m} E(m,n) \times_{\Sigma_n} \mathrm{Bij}\big([n],X\big), \quad m,n \geqslant 0,$$

where $\mathrm{Bij}(-,-)$ is the same as in (22). Pasting schemes for PROPs are *directed* (m,n)-*graphs*, by which we mean finite, not necessary connected, graphs in the sense of Definition 49 such that

(i) each edge is equipped with a direction;
(ii) there are no directed cycles and;
(iii) the set of legs is divided into the set of inputs labeled by $\{1,\dots,n\}$ and the set of outputs labeled by $\{1,\dots,m\}$.

An example of a directed graph is given in Fig. 11. We denote by $\mathtt{Gr}(m,n)$ the category of directed (m,n)-graphs and their isomorphisms. The direction of edges determines at each vertex $v \in \mathrm{vert}(G)$ of a directed graph G a disjoint decomposition

$$\mathrm{edge}(v) = \mathrm{in}(v) \sqcup \mathrm{out}(v)$$

of the set of edges adjacent to v into the set $\mathrm{in}(v)$ of incoming edges and the set $\mathrm{out}(v)$ of outgoing edges. The pair $(\#(\mathrm{out}(v)), \#(\mathrm{in}(v))) \in \mathbb{N} \times \mathbb{N}$ is called the *biarity* of v. To incorporate the unit, we need to extend the category $\mathtt{Gr}(m,n)$, for $m = n$, into the category

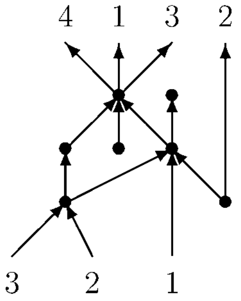


Fig. 11. A directed graph from $\mathtt{Gr}(4,3)$.

$\mathtt{UGr}(m,n)$ by allowing the exceptional graph

$$\uparrow\uparrow\uparrow\ \dots\uparrow\ \in \mathtt{UGr}(n,n), \quad n \geqslant 1,$$

with n inputs, n outputs and no vertices. For a graph $G \in \mathtt{UGr}(m,n)$ and a Σ-bimodule E, let

$$E(G) := \bigotimes_{v \in \mathrm{vert}(G)} E\big(\mathrm{out}(v), \mathrm{in}(v)\big)$$

and

$$\Gamma_{\mathsf{P}}(E)(m,n) := \operatorname*{colim}_{G \in \mathtt{UGr}(m,n)} E(G), \quad m,n \geqslant 0. \tag{43}$$

The Σ-bimodule $\Gamma_{\mathsf{P}}(E)$ is a PROP, with the vertical composition given by the disjoint union of graphs, the horizontal composition by grafting the legs, and the unit the exceptional graph $\uparrow \in \Gamma_{\mathsf{P}}(E)(1,1)$. The following proposition follows from [84] and [111–113].

Proposition 57. *The PROP $\Gamma_{\mathsf{P}}(E)$ is the free PROP generated by the Σ-bimodule E.*

As in the previous sections, (43) defines a triple $\Gamma_{\mathsf{P}} : \Sigma\text{-}\mathtt{bimod} \to \Sigma\text{-}\mathtt{bimod}$ with the triple multiplication of erasing the braces. According to general principles [20], Proposition 57 is almost equivalent to

Proposition 58. *PROPs are algebras over the triple Γ_{P}.*

One may obviously consider *non-unital PROPs* defined as algebras over the triple

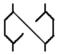
$$\Psi_{\mathsf{P}}(E)(m,n) := \operatorname*{colim}_{G \in \mathtt{Gr}(m,n)} E(G), \quad m,n \geqslant 0,$$

and develop a theory parallel to the theory of non-unital operads reviewed in Section 2.

Example 59. We will use the graphical language explained in Example 36. Let $\Gamma(\curlywedge, \curlyvee)$ be the free PROP generated by one operation $\curlywedge$ of biarity $(1,2)$ and one operation $\curlyvee$ of biarity $(2,1)$. As we noted already in [72,73], the PROP B describing bialgebras equals

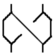
$$\mathsf{B} = \Gamma(\text{⅄}, \text{Y})/\mathsf{I}_{\mathsf{B}},$$

where I_{B} is the PROPic ideal generated by

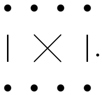
$-$, $-$ and $-$ . (44)

In the above display we denoted

$:= \text{⅄} \circ (\text{⅄} \otimes e)$, $:= \text{⅄} \circ (e \otimes \text{⅄})$, $:= (\text{Y} \otimes e) \circ \text{Y}$,
$:= (e \otimes \text{Y}) \circ \text{Y}$, $:= \text{Y} \circ \text{⅄}$ and

 $:= (\text{⅄} \otimes \text{⅄}) \circ \kappa \circ (\text{Y} \otimes \text{Y})$,

where $\kappa \in \Sigma_4$ is the permutation

$$\kappa := \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix} =$$ . (45)

The above description of B is 'tautological', but B. Enriquez and P. Etingof found in [24, Proposition 6.2] the following basis of the $\mathbf{k}$-linear space $\mathsf{B}(m, n)$ for arbitrary $m, n \geqslant 1$. Let $\text{⅄} \in \mathsf{B}(1, 2)$ be the equivalence class, in $\mathsf{B} = \Gamma(\text{⅄}, \text{Y})/\mathsf{I}_{\mathsf{B}}$, of the generator $\text{⅄} \in \Gamma(\text{Y}, \text{⅄})(1, 2)$ (we use the same symbol both for a generator and its equivalence class). Define $\text{⅄}^{[1]} := e \in \mathsf{B}(1, 1)$ and, for $a \geqslant 2$, let

$$\text{⅄}^{[a]} := \text{⅄} \circ (\text{⅄} \otimes e) \circ (\text{⅄} \otimes e^{\otimes 2}) \circ \cdots \circ \big(\text{⅄} \otimes e^{\otimes(a-2)}\big) \in \mathsf{B}(1, a).$$

Let $\text{Y}_{[b]} \in \mathsf{B}(b, 1)$ has the obvious similar meaning. The elements

$$\big(\text{⅄}^{[a_1]} \otimes \cdots \otimes \text{⅄}^{[a_m]}\big) \circ \sigma \circ (\text{Y}_{[b^1]} \otimes \cdots \otimes \text{Y}_{[b^n]}), \tag{46}$$

where $\sigma \in \Sigma_N$ for some $N \geqslant 1$, and $a_1 + \cdots + a_m = b^1 + \cdots + b^m = N$, form a $\mathbf{k}$-linear basis of $\mathsf{B}(m, n)$. This result can also be found in [57]. See also [59,94] for the bialgebra PROP viewed from a different perspective.

EXAMPLE 60. Each operad $\mathcal{P}$ generates a unique PROP P such that $\mathsf{P}(1, n) = \mathcal{P}(n)$ for each $n \geqslant 0$. The components of such a PROP are given by

$$\mathsf{P}(m, n) = \bigoplus_{r_1+\cdots+r_m=n} \big[\mathcal{P}(1, r_1) \otimes \cdots \otimes \mathcal{P}(1, r_m)\big] \times_{\Sigma_{r_1} \times \cdots \times \Sigma_{r_m}} \Sigma_n,$$

for each $m, n \geqslant 0$. The (topological) PROPs considered in [10] are all of this type. On the other hand, Example 59 shows that not each PROP is of this form. A PROP P is generated by an operad if and only if it has a presentation $\mathsf{P} = \Gamma_{\mathsf{P}}(E)/(R)$, where E is a Σ-bimodule such that $E(m, n) = 0$ for $m \neq 1$ and R is generated by elements in $\Gamma_{\mathsf{P}}(E)(1, n)$, $n \geqslant 0$.

9. Properads, dioperads and $\frac{1}{2}$PROPs

As we saw in Proposition 33, under some mild assumptions, the components of free operads are finite-dimensional. In contrast, PROPs are huge objects. For example, the component $\Gamma_{\mathsf{P}}(⅄, Y)(m,n)$ of the free PROP $\Gamma_{\mathsf{P}}(⅄, Y)$ used in the definition of the bialgebra PROP B in Example 59 is infinite-dimensional for each $m, n \geqslant 1$, and also the components of the bialgebra PROP B itself are infinite-dimensional, as follows from the fact that the Enriquez–Etingof basis (46) of $\mathsf{B}(m,n)$ has, for $m, n \geqslant 1$, infinitely many elements.

To handle this combinatorial explosion of PROPs combined with lack of suitable filtrations, smaller versions of PROPs were invented. Let us begin with the simplest modification which we use as an example which explains the general scheme of modifying PROPs. Denote $\mathtt{UGr}_c(m,n)$ the full subcategory of $\mathtt{UGr}(m,n)$ consisting of *connected* graphs and consider the triple defined by

$$\Gamma_{\mathsf{C}}(E)(m,n) := \operatorname*{colim}_{G \in \mathtt{UGr}_c(m,n)} E(G), \quad m, n \geqslant 0, \tag{47}$$

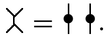
for $E \in \Sigma\text{-}\mathtt{bimod}$. The following notion was introduced by B. Vallette [111–113].

DEFINITION 61. *Properads* are algebras over the triple $\Gamma_{\mathsf{C}} : \Sigma\text{-}\mathtt{bimod} \to \Sigma\text{-}\mathtt{bimod}$.

A properad is therefore a Σ-bimodule with operations that determine coherent contractions along connected graphs. A biased definition of properads is given in [111–113]. Since Γ_c is a subtriple of Γ_{P}, each PROP is automatically also a properad. Therefore one may speak about the *endomorphism properad* $\mathcal{E}\mathrm{nd}_V$ and define *algebras* over a properad P as properad homomorphisms $\rho : P \to \mathcal{E}\mathrm{nd}_V$. Algebras over other versions of PROPs recalled below can be defined in a similar way.

EXAMPLE 62. Associative bialgebras reviewed in Example 59 are algebras over the properad B defined (tautologically) as the quotient of the free properad $\Gamma_{\mathsf{C}}(⅄, Y)$ by the properadic ideal generated by the elements listed in (44). We leave it as an exercise to describe the sub-basis of (46) that spans $B(m,n)$, $m, n \geqslant 1$.

The following slightly artificial structure exists over PROPs but not over properads. It consists of a 'multiplication' $\mu =$ ⅄$: V \otimes V \to V$, a 'comultiplication' $\Delta =$ Y$: V \to V \otimes V$ and a linear map $f =$ ⫯ $: V \to V$ satisfying $\Delta \circ \mu = f \otimes f$ or, diagrammatically



This structure cannot be a properad algebra because the graph on the right-hand side of the above display is not connected.

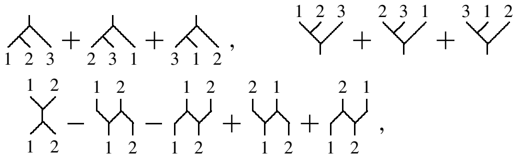
Properads are still huge objects. The first really small version of PROPs were dioperads introduced in 2003 by W.L. Gan [28]. As a motivation for his definition, consider the following

EXAMPLE 63. A *Lie bialgebra* is a vector space V with a Lie algebra structure $[-,-] =$ ⅄$: V \otimes V \to V$ and a Lie diagonal $\delta =$ Y$: V \to V \otimes V$. We assume that $[-,-]$ and δ are related by

$$\delta[a,b] = \sum\bigl([a_{(1)},b]\otimes a_{(2)} + [a,b_{(1)}]\otimes b_{(2)} + a_{(1)}\otimes[a_{(2)},b] + b_{(1)}\otimes[a,b_{(2)}]\bigr) \qquad (48)$$

for any $a, b \in V$, with the Sweedler notation $\delta a = \sum a_{(1)} \otimes a_{(2)}$ and $\delta b = \sum b_{(1)} \otimes b_{(2)}$.

Lie bialgebras are governed by the PROP $\mathsf{LieB} = \Gamma(⅄, Y)/\mathsf{I}_{\mathsf{LieB}}$, where ⅄ and Y are now *antisymmetric* and $\mathsf{I}_{\mathsf{LieB}}$ denotes the ideal generated by



and (49)

with labels indicating the corresponding permutations of the inputs and outputs.

We observe that all graphs in (49) are not only connected as demanded for properads, but also simply-connected. This suggests considering the full subcategory $\mathtt{UGr}_{\mathsf{D}}(m,n)$ of $\mathtt{UGr}(m,n)$ consisting of *connected simply-connected* graphs and the related triple

$$\Gamma_{\mathsf{D}}(E)(m,n) := \operatorname*{colim}_{G\in \mathtt{UGr}_{\mathsf{D}}(m,n)} E(G), \quad m,n \geqslant 0. \qquad (50)$$

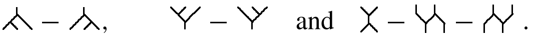
DEFINITION 64. *Dioperads are algebras over the triple* $\Gamma_{\mathsf{D}} : \Sigma\text{-}\mathtt{bimod} \to \Sigma\text{-}\mathtt{bimod}$.

A biased definition of dioperads can be found in [28]. As observed by T. Leinster, dioperads are more or less equivalent to polycategories, in the sense of [108], with one object. Lie bialgebras reviewed in Example 63 are algebras over a dioperad. Another important class of dioperad algebras is recalled in:

EXAMPLE 65. An *infinitesimal bialgebra* [48] (called in [26, Example 11.7] a *mock bialgebra*) is a vector space V with an associative multiplication $\cdot : V \otimes V \to V$ and a coassociative comultiplication $\Delta : V \to V \otimes V$ such that

$$\Delta(a\cdot b) = \sum\bigl(a_{(1)}\otimes a_{(2)}\cdot b + a\cdot b_{(1)}\otimes b_{(2)}\bigr)$$

for any $a, b \in V$. It is easy to see that the axioms of infinitesimal bialgebras are encoded by the following simply connected graphs:



Observe that associative bialgebras, as recalled in Example 53, cannot be defined over dioperads, because the rightmost graph in (44) is not simply connected. The following proposition, which should be compared to Proposition 33, shows that dioperads are of the same size as operads.

PROPOSITION 66. *Let* $E = \{E(m,n)\}_{m,n\geqslant 0}$ *be a* Σ*-bimodule such that*

$$E(m,n) = 0, \quad \textit{for } m+n \leqslant 2 \qquad (51)$$

and that $E(m, n)$ is finite-dimensional for all remaining m, n. Then the components $\Gamma_{\mathsf{D}}(E)(m, n)$ of the free dioperad $\Gamma_{\mathsf{D}}(E)$ are finite-dimensional, for all $m, n \geqslant 0$.

The proof, similar to the proof of Proposition 33, is based on the observation that the assumption (51) reduces the colimit (50) to a summation over reduced trees (trees whose all vertices have at least three adjacent edges).

An important problem arising in connection with deformation quantization is to find a reasonably small, explicit cofibrant resolution of the bialgebra PROP B. Here by a resolution we mean a differential graded PROP R together with a homomorphism $\beta : \mathsf{R} \to \mathsf{B}$ inducing a homology isomorphism. Cofibrant in this context means that R is of the form $(\Gamma_{\mathsf{P}}(E), \partial)$, where the generating Σ-bimodule E decomposes as $E = \bigoplus_{n \geqslant 0} E_n$ and the differential decreases the filtration, that is

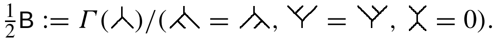
$$\partial(E_n) \subset \Gamma_{\mathsf{P}}(E)_{<n}, \quad \text{for each } n \geqslant 0,$$

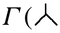
where $\Gamma_{\mathsf{P}}(E)_{<n}$ denotes the sub-PROP of $\Gamma(E)$ generated by $\bigoplus_{j<n} E_j$. This notion is an PROPic analog of the Koszul–Sullivan algebra in rational homotopy theory [36]. Several papers devoted to finding R appeared recently [57,99,97,98,101,103,102]. The approach of [79] is based on the observation that B is a deformation, in the sense explained below, of the PROP describing structures recalled in the following.

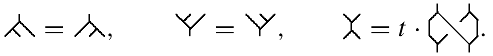
DEFINITION 67. A *half-bialgebra* or simply a $\frac{1}{2}$*bialgebra* is a vector space V with an associative multiplication $\mu : V \otimes V \to V$ and a coassociative comultiplication $\Delta : V \to V \otimes V$ that satisfy

$$\Delta(u \cdot v) = 0, \quad \text{for each } u, v \in V. \tag{52}$$

We chose this strange name because (52) is indeed one half of the compatibility relation (42) of associative bialgebras. $\frac{1}{2}$ bialgebras are algebras over the PROP



Now define, for a formal variable t, B_t to be the quotient of the free PROP $\Gamma($, $)$ by the ideal generated by



Thus B_t is a one-parameter family of PROPs with the property that $\mathsf{B}_0 = \frac{1}{2}\mathsf{B}$. At a generic t, B_t is isomorphic to the bialgebra PROP B. In other words, the PROP for bialgebras is a deformation of the PROP for $\frac{1}{2}$bialgebras. According to general principles of homological perturbation theory [35], one may try to construct the resolution R as a perturbation of a cofibrant resolution $\frac{1}{2}\mathsf{R}$ of the PROP $\frac{1}{2}\mathsf{B}$. Since $\frac{1}{2}\mathsf{B}$ is simpler that B, one may expect that resolving $\frac{1}{2}\mathsf{B}$ would be a simpler task than resolving B.

For instance, one may realize that $\frac{1}{2}$ bialgebras are algebras over a dioperad $\frac{1}{2}\mathsf{B}$, use [28] to construct a resolution $\frac{1}{2}\mathsf{R}$ of the dioperad $\frac{1}{2}\mathsf{B}$, and then take $\frac{1}{2}\mathsf{R}$ to be the PROP generated by $\frac{1}{2}\mathsf{R}$. More precisely, one denotes

$$F_1 : \mathtt{diOp} \to \mathtt{PROP} \tag{53}$$

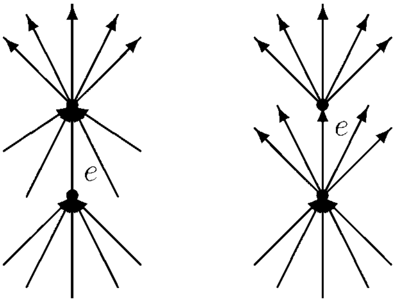


Fig. 12. Edges allowed in a $\frac{1}{2}$ graph.

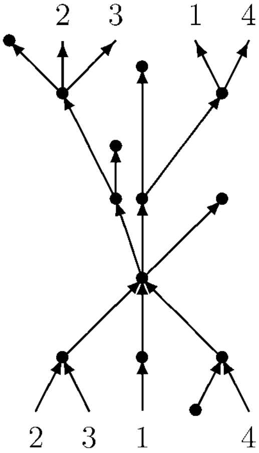


Fig. 13. A graph from $\mathsf{Gr}_{\frac{1}{2}}(4,4)$.

the left adjoint to the forgetful functor $\mathsf{PROP} \xrightarrow{\square_1} \mathtt{diOp}$ and defines $\frac{1}{2}\mathsf{R} := F_1(\frac{1}{2}\mathsf{R})$.

The problem is that we do not know whether the functor F_1 is exact, so it is not clear if $\frac{1}{2}\mathsf{R}$ constructed in this way is really a resolution of $\frac{1}{2}\mathsf{B}$. To get around this subtlety, M. Kontsevich observed that $\frac{1}{2}$ bialgebras live over a version of PROPs which is smaller than dioperads. It can be defined as follows.

Let an (m,n)-$\frac{1}{2}$ *graph* be a connected simply-connected directed (m,n)-graph each of whose edges e has the following property: either e is the unique outgoing edge of its initial vertex or e is the unique incoming edge of its terminal vertex, see Fig. 12. An example of an (m,n)-$\frac{1}{2}$ graph is given in Fig. 13. Let $\mathsf{Gr}_{\frac{1}{2}}(m,n)$ be the category of (m,n)-$\frac{1}{2}$ graphs and their isomorphisms. Define a triple $\Gamma_{\frac{1}{2}} : \Sigma\text{-}\mathtt{bimod} \to \Sigma\text{-}\mathtt{bimod}$ by

$$\Gamma_{\frac{1}{2}}(E)(m,n) := \operatorname*{colim}_{G \in \mathsf{Gr}_{\frac{1}{2}}(m,n)} E(G), \quad m,n \geqslant 0. \tag{54}$$

DEFINITION 68. A $\frac{1}{2}$PROP (called a *meager PROP* in [53]) is an algebra over the triple $\Gamma_{\frac{1}{2}} : \Sigma\text{-}\mathtt{bimod} \to \Sigma\text{-}\mathtt{bimod}$.

A biased definition of $\frac{1}{2}$PROPs can be found in [53,79,84]. We followed the original convention of [53] that $\frac{1}{2}$PROPs do not have units; the unital version of $\frac{1}{2}$PROPs can be defined in an obvious way, compare also the remarks in [79].

EXAMPLE 69. $\frac{1}{2}$ Bialgebras are algebras over a $\frac{1}{2}$PROP which we denote $\frac{1}{2}\mathsf{b}$. Another example of structures that can be defined over $\frac{1}{2}$PROPs are *Lie* $\frac{1}{2}$*bialgebras* consisting of a Lie algebra bracket $[-,-] : V \otimes V \to V$ and a Lie diagonal $\delta : V \to V \otimes V$ satisfying one-half of (48):

$$\delta[a,b] = 0.$$

Let us denote by

$$F : \frac{1}{2}\mathtt{PROP} \to \mathtt{PROP}$$

the left adjoint to the forgetful functor $\mathtt{PROP} \xrightarrow{\square} \frac{1}{2}\mathtt{PROP}$ from the category of PROPs to the category of $\frac{1}{2}$PROPs. M. Kontsevich observed that, in contrast to $F_1 : \mathtt{diOp} \to \mathtt{PROP}$ in (53), F is a *polynomial* functor, which immediately implies the following important theorem [53,84].

THEOREM 70. *The functor* $F : \frac{1}{2}\mathtt{PROP} \to \mathtt{PROP}$ *is exact.*

Now one may take a resolution $\frac{1}{2}\mathsf{r}$ of the $\frac{1}{2}$PROP $\frac{1}{2}\mathsf{b}$ and put $\frac{1}{2}\mathsf{R} := F(\frac{1}{2}\mathsf{r})$. Theorem 70 guarantees that $\frac{1}{2}\mathsf{R}$ defined in this way is indeed a resolution of the PROP $\frac{1}{2}\mathsf{B}$. Let us mention that there are also other structures invented to study resolutions of the PROP B, as the $\frac{2}{3}$PROPs of Shoikhet [101], matrons of Saneblidze and Umble [99], or special PROPs as considered in [79].

The constructions reviewed in this section can be organized into the following chain of inclusions of full subcategories:

$$\mathtt{Oper} \subset \frac{1}{2}\mathtt{PROP} \subset \mathtt{diOp} \subset \mathtt{Proper} \subset \mathtt{PROP}.$$

The general scheme behind all these constructions is the following. We start by choosing a sub-groupoid $\mathtt{SGr} = \bigsqcup_{m,n \geqslant 0} \mathtt{SGr}(m,n)$ of $\mathtt{Gr} := \bigsqcup_{m,n \geqslant 0} \mathtt{Gr}(m,n)$ (or a sub-groupoid of $\mathtt{UGr} := \bigsqcup_{m,n \geqslant 0} \mathtt{UGr}(m,n)$ if we want units). Then we define a functor $\Gamma_{\mathtt{S}} : \Sigma\text{-}\mathtt{bimod} \to \Sigma\text{-}\mathtt{bimod}$ by

$$\Gamma_{\mathtt{S}}(E)(m,n) := \operatorname*{colim}_{G \in \mathtt{SGr}(m,n)} E(G), \qquad m,n \geqslant 0.$$

It is easy to see that $\Gamma_{\mathtt{S}}$ is a subtriple of the PROP triple $\Gamma_{\mathtt{P}}$ if and only if the following two conditions are satisfied:

Pasting schemes	Corresponding structures
Rooted trees	Non-unital operads
May trees	Non-unital May operads
Extended rooted trees	Operads
Cyclic trees	Non-unital cyclic operads
Extended cyclic trees	Cyclic operads
Stable labeled graphs	Modular operads
Extended directed graphs	PROPs
Extended connected directed graphs	Properads
Extended connected 1-connected dir. graphs	Dioperads
$\frac{1}{2}$Graphs	$\frac{1}{2}$PROPs

Fig. 14. Pasting schemes and the structures they define.

(i) the groupoid $\mathtt{SGr}$ is *hereditary* in the sense that, given a graph from $\mathtt{SGr}$ with vertices decorated by graphs from $\mathtt{SGr}$, then the graph obtained by 'forgetting the braces' again belongs to $\mathtt{SGr}$, and

(ii) $\mathtt{SGr}$ contains all directed corollas.

Hereditarity (i) is necessary for Γ_{S} to be closed under the triple multiplication of Γ_{P} while (ii) guarantees that Γ_{S} has an unit. Plainly, all the three choices used above – $\mathtt{UGr}_c$, $\mathtt{UGr}_{\mathsf{D}}$ and $\mathtt{Gr}_{\frac{1}{2}}$ – satisfy the above assumptions. Let us mention that one may modify the definition of PROPs also by *enlarging* the category $\mathtt{Gr}(m,n)$, as was done for *wheeled PROPs* in [93]. The pasting schemes and the corresponding structures reviewed in this article are listed in Fig. 14.

Acknowledgements

I would like to express my gratitude to Ezra Getzler, Michiel Hazewinkel, Tom Leinster, Peter May, Jim Stasheff, Bruno Vallette and Rainer Vogt for careful reading the manuscript and useful suggestions. I am also indebted to the Institut des Hautes Études Scientifiques, Bures-sur-Yvette, for hospitality during the period when this article was completed.

References

[1] D. Balavoine, Homology and cohomology with coefficients of an algebra over a quadratic operad, J. Pure Appl. Algebra 132 (1998) 221–258.

[2] M.A. Batanin, Monoidal globular categories as a natural environment for the theory of weak n-categories, Adv. Math. 136 (1) (1998) 39–103.

[3] M.A. Batanin, Homotopy coherent category theory and A_∞-structures in monoidal categories, J. Pure Appl. Algebra 123 (1–3) (1998) 67–103.

[4] M.A. Batanin, The Eckmann–Hilton argument, higher operads and E_n-spaces, Preprint, math.CT/0207281, July 2002.

[5] M.A. Batanin, The combinatorics of iterated loop spaces, Preprint, math.CT/0301221, January 2003.

[6] A. Beilinson, V. Drinfel'd, Chiral Algebras, Colloq. Publ., vol. 51, Amer. Math. Soc., Providence, RI, 2004.

[7] C. Berger, I. Moerdijk, Axiomatic homotopy theory for operads, Comment. Math. Helv. 78 (4) (2003) 805–831.

[8] C. Berger, I. Moerdijk, Resolution of coloured operads and rectification of homotopy algebras, Preprint, math.AT/0512576, December 2005.

[9] J.M. Boardman, R.M. Vogt, Homotopy-everything H-spaces, Bull. Amer. Math. Soc. 74 (6) (1968) 1117–1122.

[10] J.M. Boardman, R.M. Vogt, Homotopy Invariant Algebraic Structures on Topological Spaces, Springer-Verlag, Berlin, 1973.

[11] A. Burroni, T-catégories (catégories dans un triple), Cah. Topol. Géom. Différ. 12 (1971) 215–321.

[12] P. Cartier, Homologie cycliques: rapport sur des travaux récents de Connes, Karoubi, Loday, Quillen ..., in: Sémin. Bourbaki Exp. no. 621, 1983–1984, Astérisque 121–122 (1985) 123–146.

[13] M. Chas, D. Sullivan, String topology, Preprint, math.GT/9911159, November 1999.

[14] F.R. Cohen, A.A. Voronov, Notes on string topology, Preprint, math.GT/0503625, March 2005.

[15] A. Connes, D. Kreimer, Renormalization in quantum field theory and the Riemann–Hilbert problem I: The Hopf algebra structure of graphs and the main theorem, Comm. Math. Phys. 210 (1) (2000) 249–273.

[16] P. Deligne, Resumé des premiérs exposés de A. Grothendieck, in: Groupes de Monodromie en Géometrie Algébrique, vol. 288, Springer-Verlag, Berlin, 1972, pp. 1–24.

[17] P. Deligne, A letter to Stasheff, Gerstenhaber, May, Schechtman and Drinfel'd, 1993, unpublished.

[18] P. Deligne, D. Mumford, The irreducibility of the space of curves of given genus, Inst. Hautes Études Sci. Publ. Math. 36 (1969) 75–109.

[19] S.L. Devadoss, Tessellations of moduli spaces and the mosaic operad, in: J.P. Meyer, J. Morava, W.S. Wilson (Eds.), Homotopy Invariant Algebraic Structures, in: Contemp. Math., vol. 239, Amer. Math. Soc., Providence, RI, 1999, pp. 91–114.

[20] S. Eilenberg, J.C. Moore, Adjoint functors and triples, Illinois J. Math. 9 (1965) 381–389.

[21] A.D. Elmendorf, The development of structured ring spectra, in: A. Baker, B. Richter (Eds.), Structured Ring Spectra, in: London Math. Soc. Lecture Note Ser., vol. 315, Cambridge Univ. Press, Cambridge, 2004.

[22] A.D. Elmendorf, I. Kříž, M.A. Mandell, J.P. May, Modern foundations of stable homotopy theory, in: I.M. James (Ed.), Handbook of Algebraic Topology, North-Holland, Amsterdam, 1995, pp. 217–257.

[23] A.D. Elmendorf, I. Kříž, M.A. Mandell, J.P. May, Rings, Modules, and Algebras in Stable Homotopy Theory, Math. Surveys Monogr., vol. 47, Amer. Math. Soc., Providence, RI, 1997.

[24] B. Enriquez, P. Etingof, On the invertibility of quantization functors, Preprint, math.QA/0306212, June 2003.

[25] P. Etingof, A. Henriques, J. Kamnitzer, E. Rains, The cohomology ring of the real locus of the moduli space of stable curves of genus 0 with marked points, Preprint, math.AT/0507514, August 2005.

[26] T.F. Fox, M. Markl, Distributive laws, bialgebras, and cohomology, in: J.-L. Loday, J.D. Stasheff, A.A. Voronov (Eds.), Operads: Proceedings of Renaissance Conferences, in: Contemp. Math., vol. 202, Amer. Math. Soc., Providence, RI, 1997, pp. 167–205.

[27] B. Fresse, Koszul duality of operads and homology of partition posets, Preprint, math.AT/0301365, January 2003.

[28] W.L. Gan, Koszul duality for dioperads, Math. Res. Lett. 10 (1) (2003) 109–124.

[29] M. Gerstenhaber, The cohomology structure of an associative ring, Ann. of Math. 78 (2) (1963) 267–288.

[30] E. Getzler, Operads and moduli spaces of genus 0 Riemann surfaces, in: R. Dijkgraaf, C. Faber, G. van der Geer (Eds.), The Moduli Space of Curves, in: Progr. Math., vol. 129, Birkhäuser, Basel, 1995, pp. 99–230.

[31] E. Getzler, J.D.S. Jones, Operads, homotopy algebra, and iterated integrals for double loop spaces, Preprint, hep-th/9403055, March 1994.

[32] E. Getzler, M.M. Kapranov, Cyclic operads and cyclic homology, in: S.-T. Yau (Ed.), Geometry, Topology and Physics for Raoul Bott, in: Conf. Proc. Lecture Notes Geom. Topol., vol. 4, Internat. Press, 1995, pp. 167–201.

[33] E. Getzler, M.M. Kapranov, Modular operads, Compos. Math. 110 (1) (1998) 65–126.

[34] V. Ginzburg, M.M. Kapranov, Koszul duality for operads, Duke Math. J. 76 (1) (1994) 203–272.

[35] V.K.A.M. Gugenheim, J.D. Stasheff, On perturbations and A_∞-structures, Bull. Soc. Math. Belg. 38 (1986) 237–246.

[36] S. Halperin, Lectures on minimal models, Mem. Soc. Math. France (N.S.) 9–10 (1983), 261 pp.

[37] R. Hartshorne, Algebraic Geometry, Grad. Texts in Math., vol. 52, Springer-Verlag, Berlin, 1977.

[38] P.J. Hilton, U. Stammbach, A Course in Homological Algebra, Grad. Texts in Math., vol. 4, Springer-Verlag, Berlin, 1971.

[39] V. Hinich, Homological algebra of homotopy algebras, Comm. Algebra 10 (25) (1997) 3291–3323.

[40] V. Hinich, Tamarkin's proof of Kontsevich formality theorem, Forum Math. 15 (2003) 591–614.

[41] V. Hinich, V.V. Schechtman, Homotopy Lie algebras, Adv. Soviet Math. 16 (2) (1993) 1–28.

[42] V. Hinich, A. Vaintrob, Cyclic operads and algebra of chord diagrams, Selecta Math. (N.S.) 8 (2) (2002) 237–282.

[43] P. Hu, Higher string topology on general spaces, Preprint, math.AT/0401081, January 2004.

[44] P. Hu, I. Kříž, A.A. Voronov, On Kontsevich's Hochschild cohomology conjecture, Preprint, math.AT/0309369, September 2003.

[45] Y.-Z. Huang, Operadic formulation of topological vertex algebras and Gerstenhaber or Batalin–Vilkovisky algebras, Comm. Math. Phys. 164 (1994) 145–161.

[46] Y.-Z. Huang, Two-Dimensional Conformal Geometry and Vertex Operator Algebras, Birkhäuser Boston, Boston, MA, 1997.

[47] Y.-Z. Huang, J. Lepowsky, Vertex operator algebras and operads, in: The Gel'fand Mathematics Seminars, 1990–1992, Birkhäuser, Boston, MA, 1993, pp. 145–161.

[48] S.A. Joni, G.-C. Rota, Coalgebras and bialgebras in combinatorics, Stud. Appl. Math. 61 (2) (1979) 93–139.

[49] T. Kimura, J.D. Stasheff, A.A. Voronov, On operad structures of moduli spaces and string theory, Comm. Math. Phys. 171 (1995) 1–25.

[50] F.F. Knudsen, The projectivity of the moduli space of stable curves, II: The stacks $M_{g,n}$, Math. Scand. 52 (1983) 161–199.

[51] F.F. Knudsen, D. Mumford, The projectivity of the moduli space of stable curves I: Preliminaries on "det" and "Div", Math. Scand. 39 (1976) 19–55.

[52] M. Kontsevich, Formal (non)commutative symplectic geometry, in: The Gel'fand Mathematics Seminars 1990–1992, Birkhäuser, Basel, 1993.

[53] M. Kontsevich, An e-mail message to M. Markl, November 2002.

[54] M. Kontsevich, Deformation quantization of Poisson manifolds, Lett. Math. Phys. 66 (3) (2003) 157–216.

[55] M. Kontsevich, Y. Manin, Gromov–Witten classes, quantum cohomology, and enumerative geometry, Comm. Math. Phys. 164 (1994) 525–562.

[56] M. Kontsevich, Y. Soibelman, Deformations of algebras over operads and Deligne's conjecture, Preprint, math.QA/0001151, January 2000.

[57] M. Kontsevich, Y. Soibelman, Deformation theory of bialgebras, Hopf algebras and tensor categories, Preprint, March 2002.

[58] I. Kříž, J.P. May, Operads, algebras, modules and motives, Astérisque 233 (1995).

[59] S. Lack, Composing PROPS, Theory Appl. Categ. 13 (9) (2004) 147–163 (electronic).

[60] J. Lambek, Deductive systems and categories II, in: P. Hilton (Ed.), Category Theory, Homology Theory and Their Applications I, in: Lecture Notes in Math., vol. 86, Springer-Verlag, Berlin, 1969, pp. 76–122.

[61] T. Leinster, Higher Operads, Higher Categories, London Math. Soc. Lecture Note Ser., vol. 298, Cambridge Univ. Press, Cambridge, 2004.

[62] T. Leinster, Operads in higher-dimensional category theory, Theory Appl. Categ. 12 (2004) 73–194.

[63] J.-L. Loday, La renaissance des opérades, in: Séminaire Bourbaki, Exp. No. 792. 47ème année 1994–1995, Astérisque 237 (1996) 47–74.

[64] J.-L. Loday, Algèbres ayant deux opérations associatives (digèbres), C. R. Acad. Sci. Paris Sér. I Math. 321 (2) (1995) 141–146.

[65] J.-L. Loday, A. Frabetti, F. Chapoton, F. Goichot, Dialgebras and Related Operads, Lecture Notes in Math., vol. 1763, Springer-Verlag, Berlin, 2001.

[66] J.-L. Loday, D. Quillen, Cyclic homology and the Lie algebra homology of matrices, Comment. Math. Helv. 59 (1984) 565–591.

[67] S. Mac Lane, Natural associativity and commutativity, Rice Univ. Stud. 49 (1) (1963) 28–46.

[68] S. Mac Lane, Categorical algebra, Bull. Amer. Math. Soc. 71 (1965) 40–106.

[69] M.A. Mandell, E_∞ algebras and p-adic homotopy theory, Topology 40 (1) (2001) 43–94.

[70] M.A. Mandell, Cochains and homotopy type, Preprint, math.AT/0311016, November 2003.

[71] Y.I. Manin, Frobenius Manifolds, Quantum Cohomology, and Moduli Spaces, Amer. Math. Soc. Colloq. Publ., vol. 47, Amer. Math. Soc., Providence, RI, 1999.

[72] M. Markl, Deformations and the coherence, in: Proc. of the Winter School ‘Geometry and Physics,’ Zdíkov, Bohemia, January 1993, Rend. Circ. Mat. Palermo (2) Suppl. 37 (1994) 121–151.

[73] M. Markl, Cotangent cohomology of a category and deformations, J. Pure Appl. Algebra 113 (1996) 195–218.

[74] M. Markl, Distributive laws and Koszulness, Ann. Inst. Fourier (Grenoble) 46 (4) (1996) 307–323.

[75] M. Markl, Models for operads, Comm. Algebra 24 (4) (1996) 1471–1500.

[76] M. Markl, A compactification of the real configuration space as an operadic completion, J. Algebra 215 (1999) 185–204.

[77] M. Markl, Cyclic operads and homology of graph complexes, in: Proceedings of the 18th Winter School “Geometry and Physics”, Srní, Czech Republic, January 10–17, 1998, Rend. Circ. Mat. Palermo (2) Suppl. 59 (1999) 161–170.

[78] M. Markl, Loop homotopy algebras in closed string field theory, Comm. Math. Phys. 221 (2001) 367–384.

[79] M. Markl, A resolution (minimal model) of the PROP for bialgebras, J. Pure Appl. Algebra 205 (2) (2006) 341–374.

[80] M. Markl, Homotopy algebras are homotopy algebras, Forum Math. 16 (1) (2004) 129–160.

[81] M. Markl, Intrinsic brackets and the L_∞-deformation theory of bialgebras, Preprint, math.AT/0411456, November 2004.

[82] M. Markl, E. Remm, Algebras with one operation including Poisson and other Lie-admissible algebras, J. Algebra 299 (2006) 171–189.

[83] M. Markl, S. Shnider, J.D. Stasheff, Operads in Algebra, Topology and Physics, Math. Surveys Monogr., vol. 96, Amer. Math. Soc., Providence, RI, 2002.

[84] M. Markl, A.A. Voronov, PROPped up graph cohomology, Preprint, math.QA/0307081, July 2003.

[85] L. Mauri, Algebraic theories in monoidal categories, Preprint, available from the author’s web-site http://math.rutgers.edu/~mauri/, 2005.

[86] J.P. May, The Geometry of Iterated Loop Spaces, Lecture Notes in Math., vol. 271, Springer-Verlag, New York, 1972.

[87] J.P. May, Definitions: Operads, algebras and modules, in: J.L. Loday, J.D. Stasheff, A.A. Voronov (Eds.), Operads: Proceedings of Renaissance Conferences, in: Contemp. Math., vol. 202, Amer. Math. Soc., Providence, RI, 1997, pp. 1–7.

[88] J.P. May, Operadic tensor products and smash products, in: J.L. Loday, J.D. Stasheff, A.A. Voronov (Eds.), Operads: Proceedings of Renaissance Conferences, in: Contemp. Math., vol. 202, Amer. Math. Soc., 1997, pp. 287–303.

[89] J.P. May, Operads, algebras, and modules, in: J.L. Loday, J.D. Stasheff, A.A. Voronov (Eds.), Operads: Proceedings of Renaissance Conferences, in: Contemp. Math., vol. 202, Amer. Math. Soc., Providence, RI, 1997, pp. 15–31.

[90] J.P. May, Brave new worlds in stable homotopy theory, in: M. Mahowald (Ed.), Proceedings of a Conference on Homotopy Theory, Evanston, IL, USA, in: Contemp. Math., vol. 220, Amer. Math. Soc., Providence, RI, 1998, pp. 193–212.

[91] J.P. May, Caterads and algebras over caterads, Preprint, available from the author’s home page, December 2004.

[92] S.A. Merkulov, PROP profile of Poisson geometry, Preprint, math.DG/0401034, January 2003.

[93] S.A. Merkulov, PROP profile of deformation quantization, Preprint, math.QA/0412257, December 2004.

[94] T. Pirashvili, On the PROP corresponding to bialgebras, Cah. Topol. Géom. Différ. Catég. 43 (3) (2002) 221–239.

[95] D. Quillen, Rational homotopy theory, Ann. of Math. 90 (1969) 205–295.

[96] P. Salvatore, Configuration spaces with summable labels, in: J. Aguadé (Ed.), Cohomological Methods in Homotopy Theory. Proceedings of the Barcelona Conference on Algebraic Topology (BCAT), Bellaterra, Spain, June 4–10, 1998, in: Progr. Math., vol. 196, Birkhäuser, Basel, 2001, pp. 375–395.

[97] S. Saneblidze, R. Umble, The biderivative and A_∞-bialgebras, Preprint, math.AT/0406270, April 2004.

[98] S. Saneblidze, R. Umble, The biderivative, matrons and A_∞-bialgebras, Preprint, July 2004.

[99] S. Saneblidze, R. Umble, Matrons, A_∞-bialgebras and the polytopes KK, Preprint, math.AT/0508017, August 2005.

[100] R. Schwänzl, R.M. Vogt, The categories of A_∞- and E_∞-monoids and ring spaces as closed simplicial and topological model categories, Arch. Math. (Basel) 56 (4) (1991) 405–411.

[101] B. Shoikhet, A concept of $\frac{2}{3}$PROP and deformation theory of (co)associative coalgebras, Preprint, math.QA/0311337, November 2003.

[102] B. Shoikhet, The CROCs, non-commutative deformations, and (co)associative bialgebras, Preprint, math.QA/0306143, June 2003.

[103] B. Shoikhet, An explicit deformation theory of (co)associative bialgebras, Preprint, math.QA/0310320, October 2003.

[104] M. Spitzweck, Operads, algebras and modules in general model categories, Preprint, math.AT/0101102, January 2001.

[105] J.D. Stasheff, Math. Reviews reports MR0420609 (54#8623a) and MR0420610 (54#8623b) on [10] and [86], available from MathSciNet.

[106] J.D. Stasheff, Homotopy associativity of H-spaces I,II, Trans. Amer. Math. Soc. 108 (1963) 275–312.

[107] D. Sullivan, Infinitesimal computations in topology, Inst. Hautes Études Sci. Publ. Math. 47 (1977) 269–331.

[108] M.E. Szabo, Polycategories, Comm. Algebra 3 (8) (1975) 663–689.

[109] D. Tamarkin, B. Tsygan, Cyclic formality and index theorems, Lett. Math. Phys. 56 (2001) 85–97.

[110] D.E. Tamarkin, Another proof of M. Kontsevich formality theorem, Preprint, math.QA/9803025, March 1998.

[111] B. Vallette, Dualité de Koszul des PROPs. PhD thesis, Université Louis Pasteur, 2003.

[112] B. Vallette, A Koszul duality for props, Preprint, math.AT/0411542, November 2004.

[113] B. Vallette, Koszul duality for PROPs, C. R. Math. Acad. Sci. Paris 338 (12) (2004) 909–914.

[114] R.M. Vogt, My time as Mike Boardman's student and our work on infinite loop spaces, Preprint, 1998.

[115] R.M. Vogt, Introduction to algebra over "brave new rings", in: The 18th Winter School "Geometry and Physics", Srní, 1998, Rend. Circ. Mat. Palermo (2) Suppl. 59 (1999) 49–82.

[116] R.M. Vogt, Cofibrant operads and universal E-infinity operads, Topology Appl. 133 (1) (2003) 69–87.

[117] A.A. Voronov, Notes on universal algebra, Preprint, math.QA/0111009, November 2001.

[118] R.O. Wells, Differential Analysis on Complex Manifolds, Grad. Texts in Math., vol. 65, Springer-Verlag, Berlin, 1980.

Section 2B
Homological Algebra. Cohomology. Cohomological Methods in Algebra. Homotopical algebra

This page intentionally left blank

A_∞-algebras, A_∞-categories and A_∞-functors

Volodymyr Lyubashenko

Institute of Mathematics, *National Academy of Sciences of Ukraine*, 3 *Tereshchenkivska st.*,
Kyiv-4, 01601 *MSP*, *Ukraine*
E-mail: *lub@imath.kiev.ua*

Oleksandr Manzyuk

Fachbereich Mathematik, *Postfach* 3049, 67653 *Kaiserslautern*, *Germany*
E-mail: *manzyuk@mathematik.uni-kl.de*

Contents

HANDBOOK OF ALGEBRA, VOL. 5
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(07)05003-6

1. Introduction

A_∞-algebras were introduced by Stasheff in 1963 as algebraic counterpart of his theory of H-spaces, topological monoids associative up to homotopy which are in turn coherent up to higher homotopies, etc. [34]. Topological applications of A_∞-algebras were developed in a series of papers of Smirnov starting from [31] and culminating in [32] and in papers of Kadeishvili beginning with [14]. Algebraic aspects of A_∞-algebras were studied by Kadeishvili [15] and by Getzler and Jones [11]. A_∞-modules over A_∞-algebras are considered by Keller [17] and by Lefèvre-Hasegawa [23].

A_∞-categories generalize differential graded categories on the one hand and A_∞-algebras on the other. The basic notions of A_∞-categories and of A_∞-functors have been studied in works of Fukaya [9], Keller [17], the first author [24] and Soibelman [33]. The homological mirror symmetry conjecture formulated by Kontsevich [20] states equivalence of two A_∞-categories, one coming from the symplectic structure of a manifold, another from the complex structure of its mirror manifold. It was proven in some cases by Kontsevich and Soibelman [21]. This subject is linked with deformation quantization theory via A_∞-algebras and A_∞-categories, as shown in the works of Fukaya [10] and of Bressler and Soibelman [5].

Polishchuk [30] classifies A_∞-structures on the category of line bundles over an elliptic curve with some additional requirements. He applies A_∞-categories to deformation theory questions in algebraic geometry and to the Fourier–Mukai transform in [29]. Applications of A_∞-algebras and of so called cyclic A_∞-algebras in mathematical physics are described by Kajiura [16]. A_∞-categories with additional properties are important for topological conformal field theories as shown by Costello [6].

In the same year, 1963 (simultaneously with the discovery of A_∞-algebras) Verdier has realized the ideas of Grothendieck about homological algebra in the notion of triangulated category [36]. It was clear, however, that the new concept was but (an offshoot) of something underlying it. An insight came from Bondal and Kapranov [2], who suggested to look for pretriangulated differential graded (dg) categories, whose 0-th homology would give the triangulated category in question. Drinfeld has succeeded in doing this for derived categories [8] by relating quotient constructions for pretriangulated dg-categories and quotients (localizations) of triangulated categories.

Clearly, the class of differential graded functors between pretriangulated dg-categories is too small to provide a sufficient supply of morphisms. We propose to extend this class to unital A_∞-functors. It is advantageous to extend simultaneously the class of objects to pretriangulated A_∞-categories. Taking 0-th homology one can get some known results in the theory of triangulated (derived) categories. However, working with pretriangulated A_∞-categories instead of triangulated categories opens up new possibilities.

1.1. *Notation*

We work within set theory in which all sets are elements of some universe [12]. In particular, a universe is an element of another universe. One of them, $\mathscr{U}$ (containing an element which is an infinite set) is considered basic.

A structure is called *$\mathscr{U}$-small* if it consists of sets which are in bijection with elements of the universe $\mathscr{U}$ [12, Exposé I.1]. For instance, sets, rings, modules, categories, etc. can be $\mathscr{U}$-small. A category $\mathcal{V}$ is a *$\mathscr{U}$-category* if the sets of morphisms $\mathcal{V}(X, Y)$ are $\mathscr{U}$-small for all objects X, Y of $\mathcal{V}$. A $\mathscr{U}$-category $\mathcal{V}$ is $\mathscr{U}$-small if and only if $\operatorname{Ob}\mathcal{V}$ is a $\mathscr{U}$-small set.

Denote by $\mathbf{n}$ the linearly ordered set $\{1 < 2 < \dots < n\}$ for an integer $n \geqslant 0$, where $\mathbf{0} = \varnothing$. Let $\mathcal{O}$ be the category with objects $\mathbf{n}$ for $n \geqslant 0$, whose morphisms are non-decreasing maps. Let $\mathcal{S}$ denote the category, whose objects are the same as in $\mathcal{O}$, but the morphisms are all mappings.

Let $\Bbbk$ be a $\mathscr{U}$-small commutative associative ring with unity. Denote by $\mathbf{gr}$ (resp. $\mathbf{dg}$) the category of graded (resp. differential graded) $\mathscr{U}$-small $\Bbbk$-modules. Let $\mathsf{C}_\Bbbk = \underline{\mathbf{dg}}$ be the differential graded category of differential graded $\Bbbk$-modules (complexes of $\Bbbk$-modules).

A (*differential*) *graded quiver* $\mathcal{A}$ always means for us a $\mathscr{U}$-small set of objects $\operatorname{Ob}\mathcal{A}$ together with $\mathscr{U}$-small (differential) $\mathbb{Z}$-graded $\Bbbk$-modules of morphisms $\mathcal{A}(X, Y)$, given for each pair $X, Y \in \operatorname{Ob}\mathcal{A}$. A *span morphism* $r : \mathcal{A} \to \mathcal{B}$ is a pair of maps $f = \operatorname{Ob}_s r$, $g = \operatorname{Ob}_t r : \operatorname{Ob}\mathcal{A} \to \operatorname{Ob}\mathcal{B}$ and a collection of graded $\Bbbk$-linear maps $r : \mathcal{A}(X, Y) \to \mathcal{B}(Xf, Yg)$, given for each pair X, Y of objects of $\mathcal{A}$. Morphisms of quivers are span morphisms $r : \mathcal{A} \to \mathcal{B}$ of degree 0 such that $\operatorname{Ob}_s r = \operatorname{Ob}_t r : \operatorname{Ob}\mathcal{A} \to \operatorname{Ob}\mathcal{B}$. This map is denoted simply $\operatorname{Ob} r$. The category of quivers is denoted $\mathscr{Q}$. The category $\mathscr{Q}$ has a natural symmetric monoidal structure $\mathscr{Q}_p = (\mathscr{Q}, \boxtimes)$. For given quivers $\mathcal{Q}_i$ the quiver $\boxtimes^{i\in I}\mathcal{Q}_i$ has a set of objects $\prod_{i\in I}\operatorname{Ob}\mathcal{Q}_i$ and as (differential) graded $\Bbbk$-modules of morphisms $(\boxtimes^{i\in I}\mathcal{Q}_i)((X_i)_{i\in I}, (Y_i)_{i\in I}) = \otimes^{i\in I}\mathcal{Q}_i(X_i, Y_i)$. The unit object $\mathbb{1}_p = \boxtimes^{\varnothing}()$ of $\mathscr{Q}_p$ is the graded $\Bbbk$-quiver with a unique object $*$ and module of homomorphisms $\mathbb{1}_p(*, *) = \Bbbk$.

For any graded $\Bbbk$-module M there is another graded $\Bbbk$-module $sM = M[1]$, its *suspension*, with the shifted grading $(sM)^k = M[1]^k = M^{k+1}$. The mapping $s : M \to sM$ given by the identity maps $M^k \to M[1]^{k-1}$ has degree -1. These notions extend to graded $\Bbbk$-quivers.

Let S be a $\mathscr{U}$-small set. Let $\mathscr{Q}/S$ be the category of quivers $\mathcal{C}$ with set of objects S, whose morphisms are morphisms of quivers $f : \mathcal{A} \to \mathcal{B}$ such that $\operatorname{Ob} f = \mathrm{id}_S$. The $\Bbbk$-linear Abelian category $\mathscr{Q}/S$ admits the following structure of a monoidal category $(\mathscr{Q}/S, \otimes)$:

$$\bigl(\otimes^{i\in\mathbf{n}}\mathcal{Q}_i\bigr)(X, Z) = \bigoplus_{Y_i\in S, 0\leqslant i\leqslant n}^{Y_0=X, Y_n=Z} \otimes^{i\in\mathbf{n}}\mathcal{Q}_i(Y_{i-1}, Y_i).$$

In particular, the unit object $\otimes^{\mathbf{0}}() = \Bbbk S$ has set of objects S, and the graded $\Bbbk$-module of morphisms is $\Bbbk S(X, Y) = \Bbbk$ if $X = Y \in S$ and $\Bbbk S(X, Y) = 0$ if $X \neq Y$. Given a quiver $\mathcal{C}$, we abbreviate the quiver $\Bbbk\operatorname{Ob}\mathcal{C}$ to $\Bbbk\mathcal{C}$. We view it as a differential graded quiver with zero differential. A map $f : R \to S$ induces a $\Bbbk$-quiver morphism $\Bbbk f : \Bbbk R \to \Bbbk S$, $\operatorname{Ob}\Bbbk f = f$, $\Bbbk f = \mathrm{id}_\Bbbk : \Bbbk R(X, X) \to \Bbbk S(Xf, Xf)$ for all $X \in R$, and $\Bbbk f = 0 : \Bbbk R(X, Y) \to \Bbbk S(Xf, Yf)$ if $X \neq Y \in R$. For a quiver morphism $f : \mathcal{A} \to \mathcal{B}$ we denote by $\Bbbk f$ the quiver map $\Bbbk\operatorname{Ob} f : \Bbbk\mathcal{A} \to \Bbbk\mathcal{B}$.

Let $\mathcal{A}_i, \mathcal{B}_i$, $i \in \mathbf{n}$, be quivers with $\operatorname{Ob}\mathcal{A}_i = S$, $\operatorname{Ob}\mathcal{B}_i = R$ for all $i \in \mathbf{n}$. Let $f_i : \mathcal{A}_i \to \mathcal{B}_i$, $i \in \mathbf{n}$, be span morphisms such that $\operatorname{Ob}_t f_i = \operatorname{Ob}_s f_{i+1} : S \to R$ for all $1 \leqslant i < n$. Then the span morphism $f = \otimes^{i\in\mathbf{n}} f_i : \otimes_S^{i\in\mathbf{n}}\mathcal{A}_i \to \otimes_R^{i\in\mathbf{n}}\mathcal{B}_i$ with object maps $\operatorname{Ob}_s f = \operatorname{Ob}_s f_1 : S \to R$, $\operatorname{Ob}_t f = \operatorname{Ob}_t f_n : S \to R$ is defined in the obvious way.

Suppose that $\mathcal{A}_i^j$ are graded quivers, $i \in I$, $j \in \mathbf{m}$. Assume that $\operatorname{Ob} \mathcal{A}_i^j = S_i$ does not depend on j. Define $S = \prod_{i\in I} S_i$. Denote by $\otimes_{S_i}$ the tensor product in $\mathscr{Q}/S_i$. There is an isomorphism of graded quivers

$$\bar{\varkappa} : \otimes_S^{j\in\mathbf{m}} \boxtimes^{i\in I} \mathcal{A}_i^j \to \boxtimes^{i\in I} \otimes_{S_i}^{j\in\mathbf{m}} \mathcal{A}_i^j, \tag{1.1.1}$$

the identity on objects, which is a direct sum of the permutation isomorphisms

$$\sigma_{(12)} : \otimes^{j\in\mathbf{m}} \otimes^{i\in I} \mathcal{A}_i^j\bigl(X_i^{j-1}, X_i^j\bigr) \to \otimes^{i\in I} \otimes^{j\in\mathbf{m}} \mathcal{A}_i^j\bigl(X_i^{j-1}, X_i^j\bigr),$$

where $X_i^j \in S_i$, $0 \leqslant j \leqslant m$.

For a quiver $\mathcal{A}$ consider the quivers $T^n\mathcal{A} = \mathcal{A}^{\otimes n}$, $n \geqslant 0$, the tensor powers of $\mathcal{A}$ in $\mathscr{Q}/\operatorname{Ob}\mathcal{A}$, and the quivers $T\mathcal{A}$, $T^{\geqslant 1}\mathcal{A}$, $T^{\leqslant 1}\mathcal{A}$:

$$T\mathcal{A} = \bigoplus_{n=0}^{\infty} T^n\mathcal{A}, \qquad T^{\geqslant 1}\mathcal{A} = \bigoplus_{n=1}^{\infty} T^n\mathcal{A},$$
$$T^{\leqslant 1}\mathcal{A} = T^0\mathcal{A} \oplus T^1\mathcal{A} = \Bbbk\mathcal{A} \oplus \mathcal{A}.$$

Often we use right operators: the composition of two maps (or morphisms) $f : X \to Y$ and $g : Y \to Z$ is denoted by $fg = f \cdot g : X \to Z$ instead of $g \circ f$. A map is written on elements as $f : x \mapsto xf = (x)f$. However, these conventions are not used systematically, and $f(x)$ might be used instead. In all formulas with graded elements and maps we assume the Koszul rule:

$$(x \otimes y)(f \otimes g) = (-)^{yf} xf \otimes yg = (-1)^{\deg y \cdot \deg f} xf \otimes yg.$$

2. A_∞-categories and A_∞-functors

We define A_∞-categories and A_∞-functors as particular cases of differential coalgebras and their homomorphisms.

2.1. *A_∞-categories as coalgebras*

Following Keller [19] we call a graded quiver $\mathcal{A}$ a *cocomplete coalgebra* if it has a structure of a locally nilpotent coassociative coalgebra in the monoidal category $\mathscr{Q}/\operatorname{Ob}\mathcal{A}$. Thus, there is a quiver map $\bar\Delta : \mathcal{A} \to \mathcal{A} \otimes \mathcal{A}$ with $\operatorname{Ob}\bar\Delta = \mathrm{id}_{\operatorname{Ob}\mathcal{A}}$ that satisfies the usual coassociativity relation. The local nilpotency means that for each pair X, Y of objects of $\mathcal{C}$

$$\mathcal{C}(X, Y) = \bigcup_{k\geqslant 1} \operatorname{Ker}\bigl(\bar\Delta^{(k)} : \mathcal{C}(X, Y) \to \mathcal{C}^{\otimes k}(X, Y)\bigr),$$

where $\bar\Delta^{(k)}$ is the comultiplication iterated $k - 1$ times. The morphisms $f : \mathcal{A} \to \mathcal{B}$ of cocomplete coalgebras are required to preserve comultiplication:

$$\bigl(\mathcal{A} \xrightarrow{f} \mathcal{B} \xrightarrow{\bar\Delta} \mathcal{B} \otimes \mathcal{B}\bigr) = \bigl(\mathcal{A} \xrightarrow{\bar\Delta} \mathcal{A} \otimes \mathcal{A} \xrightarrow{f\otimes f} \mathcal{B} \otimes \mathcal{B}\bigr).$$

The quiver $T^{\geqslant 1}\mathcal{A}$ equipped with the cut comultiplication

$$\bar{\Delta} : T^{\geqslant 1}\mathcal{A}(X,Y) \to \bigoplus_{Z\in\operatorname{Ob}\mathcal{A}} T^{\geqslant 1}\mathcal{A}(X,Z) \bigotimes T^{\geqslant 1}\mathcal{A}(Z,Y),$$
$$h_1 \otimes h_2 \otimes \dots \otimes h_n \mapsto \sum_{k=1}^{n-1} h_1 \otimes \dots \otimes h_k \bigotimes h_{k+1} \otimes \dots \otimes h_n$$

is an example of cocomplete coalgebra.

A graded quiver $\mathcal{A}$ is an *augmented coalgebra* if it is equipped with a structure of an augmented counital coassociative coalgebra in the monoidal category $\mathscr{Q}/\operatorname{Ob}\mathcal{A}$. This means that there are quiver maps $\varepsilon : \mathcal{A} \to \Bbbk\mathcal{A}$, $\eta : \Bbbk\mathcal{A} \to \mathcal{A}$ and $\Delta_0 : \mathcal{A} \to \mathcal{A} \otimes \mathcal{A}$ with $\operatorname{Ob}\varepsilon = \operatorname{Ob}\eta = \operatorname{Ob}\Delta_0 = \mathrm{id}_{\operatorname{Ob}\mathcal{A}}$ that satisfy the usual augmented coalgebra relations. Morphisms $f : \mathcal{A} \to \mathcal{B}$ of the coalgebras considered are required to preserve comultiplication, counit and augmentation:

$$\begin{aligned}
&\bigl(\mathcal{A} \xrightarrow{f} \mathcal{B} \xrightarrow{\Delta_0} \mathcal{B} \otimes \mathcal{B}\bigr) = \bigl(\mathcal{A} \xrightarrow{\Delta_0} \mathcal{A} \otimes \mathcal{A} \xrightarrow{f\otimes f} \mathcal{B} \otimes \mathcal{B}\bigr),\\
&\bigl(\mathcal{A} \xrightarrow{f} \mathcal{B} \xrightarrow{\varepsilon} \Bbbk\mathcal{B}\bigr) = \bigl(\mathcal{A} \xrightarrow{\varepsilon} \Bbbk\mathcal{A} \xrightarrow{\Bbbk f} \Bbbk\mathcal{B}\bigr),\\
&\bigl(\Bbbk\mathcal{A} \xrightarrow{\eta} \mathcal{A} \xrightarrow{f} \mathcal{B}\bigr) = \bigl(\Bbbk\mathcal{A} \xrightarrow{\Bbbk f} \Bbbk\mathcal{B} \xrightarrow{\eta} \mathcal{B}\bigr).
\end{aligned}$$

Given a coassociative coalgebra $(\mathcal{C}, \bar{\Delta} : \mathcal{C} \to \mathcal{C} \otimes \mathcal{C})$ in the monoidal category $\mathscr{Q}/\operatorname{Ob}\mathcal{C}$ we construct an augmented counital coassociative coalgebra structure $(\Delta_0, \varepsilon, \eta)$ on the object $T^{\leqslant 1}\mathcal{C} = \Bbbk\mathcal{C} \oplus \mathcal{C}$ of the category $\mathscr{Q}/\operatorname{Ob}\mathcal{C}$ by the formulas

$$\Delta_0 = \Delta' + \mathrm{id}_{T^{\leqslant 1}\mathcal{C}} \otimes \eta + \eta \otimes \mathrm{id}_{T^{\leqslant 1}\mathcal{C}} - \varepsilon\eta \otimes \eta,$$

where

$$\Delta' = \bigl(T^{\leqslant 1}\mathcal{C} \xrightarrow{\mathrm{pr}_1} \mathcal{C} \xrightarrow{\bar{\Delta}} \mathcal{C} \otimes \mathcal{C} \xrightarrow{\mathrm{in}_1 \otimes \mathrm{in}_1} T^{\leqslant 1}\mathcal{C} \otimes T^{\leqslant 1}\mathcal{C}\bigr)$$

and

$$\eta = \mathrm{in}_0 : \Bbbk\mathcal{C} \to T^{\leqslant 1}\mathcal{C}, \qquad \varepsilon = \mathrm{pr}_0 : T^{\leqslant 1}\mathcal{C} \to \Bbbk\mathcal{C}.$$

The resulting functor $\mathcal{C} \mapsto T^{\leqslant 1}\mathcal{C}$, $f \mapsto T^{\leqslant 1} f = \Bbbk f \oplus f$, is an equivalence of the category of coassociative coalgebras $\mathrm{c}\mathscr{Q}$ and the category of augmented coalgebras $\mathrm{ac}\mathscr{Q}$. The tensor quiver $T\mathcal{A} = T^{\leqslant 1}T^{\geqslant 1}\mathcal{A}$ inherits the cut comultiplication $\Delta_0 : h_1 \otimes h_2 \otimes \dots \otimes h_n \mapsto \sum_{k=0}^{n} h_1 \otimes \dots \otimes h_k \bigotimes h_{k+1} \otimes \dots \otimes h_n$.

2.2. LEMMA. *(Cf. Keller [19].) Let $\mathcal{C}$ be a cocomplete coalgebra, and let $\mathcal{A}$ be a graded quiver. Then there are natural bijections*

$$\begin{aligned}
&\mathrm{ac}\mathscr{Q}\bigl(T^{\leqslant 1}\mathcal{C}, T\mathcal{A}\bigr) \xleftarrow{\sim} \mathrm{c}\mathscr{Q}\bigl(\mathcal{C}, T^{\geqslant 1}\mathcal{A}\bigr) \xrightarrow{\sim} \mathscr{Q}(\mathcal{C}, \mathcal{A}),\\
&\bigl(T^{\leqslant 1} f : T^{\leqslant 1}\mathcal{C} \to T\mathcal{A}\bigr) \mapsfrom \bigl(f : \mathcal{C} \to T^{\geqslant 1}\mathcal{A}\bigr) \mapsto \bigl(\mathcal{C} \xrightarrow{f} T^{\geqslant 1}\mathcal{A} \xrightarrow{\mathrm{pr}_1} \mathcal{A}\bigr).
\end{aligned}$$

When $\mathcal{C} = T^{\geqslant 1}\mathcal{B}$, we call the quiver morphisms

$$f_{nk} = \bigl(T^n\mathcal{B} \overset{\mathrm{in}_n}{\hookrightarrow} T^{\geqslant 1}\mathcal{B} \xrightarrow{f} T^{\geqslant 1}\mathcal{A} \xrightarrow{\mathrm{pr}_k} T^k\mathcal{A}\bigr)$$

the *matrix elements* of f. The matrix elements $f_{n1} : T^n\mathcal{B} \to \mathcal{A}$ are called *components* of f and are abbreviated to f_n. The coalgebra morphism f can be recovered from its components by the formula

$$f_{nk} = \sum_{i_1+\dots+i_k=n} f_{i_1} \otimes \dots \otimes f_{i_k} : T^n\mathcal{B} \to T^k\mathcal{A}.$$

Let $f, g : \mathcal{A} \to \mathcal{B}$ be augmented coalgebra morphisms. An (f, g)*-coderivation* $r : f \to g : \mathcal{A} \to \mathcal{B}$ is a span morphism $r : \mathcal{A} \to \mathcal{B}$ of some degree d such that $\operatorname{Ob}_s r = \operatorname{Ob} f$, $\operatorname{Ob}_t r = \operatorname{Ob} g$ and $r\Delta_0 = \Delta_0(f \otimes r + r \otimes g)$.

Let $\mathcal{C}$ be a cocomplete coalgebra. Let $\mathcal{A} = T^{\leqslant 1}\mathcal{C}$ and let $f, g : \mathcal{C} \to T^{\geqslant 1}\mathcal{B}$ be coalgebra morphisms. A $(T^{\leqslant 1} f, T^{\leqslant 1} g)$-coderivation $r : T^{\leqslant 1} f \to T^{\leqslant 1} g : \mathcal{A} \to T\mathcal{B}$ can be recovered from $\check{r} = (\mathcal{A} \xrightarrow{r} T\mathcal{B} \xrightarrow{\mathrm{pr}_1} \mathcal{B})$ via

$$r = \sum_{q+1+t=l} \bigl(\mathcal{A} \xrightarrow{\Delta_0^{(l)}} T^l\mathcal{A} \xrightarrow{\check{f}^{\otimes q} \otimes \check{r} \otimes \check{g}^{\otimes t}} T^l\mathcal{B} \xrightarrow{\mathrm{in}_l} T^{\geqslant 1}\mathcal{B}\bigr),$$

where

$$\check{f} = \bigl(\mathcal{A} = T^{\leqslant 1}\mathcal{C} \xrightarrow{T^{\leqslant 1} f} T\mathcal{B} \xrightarrow{\mathrm{pr}_1} \mathcal{B}\bigr)$$

and similarly for $\check{g}$. When $\mathcal{C} = T^{\geqslant 1}\mathcal{D}$, $\mathcal{A} = T\mathcal{D}$, the quiver maps $r_k = \check{r}|_{T^k\mathcal{D}} : T^k\mathcal{D} \to \mathcal{B}$, $k \geqslant 0$, of degree d are called *components* of r. The matrix elements of r are

$$r_{nm} = \sum_{\substack{i+j+k=n\\ p+1+q=m}} f_{ip} \otimes r_j \otimes g_{kq} : T^n\mathcal{A} \to T^m\mathcal{B}.$$

2.3. DEFINITION. An A_∞*-category* is a graded quiver $\mathcal{A}$ equipped with a differential $b : Ts\mathcal{A} \to Ts\mathcal{A}$ of degree 1 with $\operatorname{Ob} b = \mathrm{id}_{\operatorname{Ob}\mathcal{A}}$ such that the cut comultiplication, the counit and the augmentation are chain maps.

The condition of Δ_0 being a chain map is equivalent to b being a (1, 1)-coderivation. It can be recovered from its components as $b_{nm} = \sum_{p+k+q=n}^{p+1+q=m} 1^{\otimes p} \otimes b_k \otimes 1^{\otimes q} : T^n s\mathcal{A} \to T^m s\mathcal{A}$. Commutation of the augmentation map in_0 with the differential is equivalent to $b_0 = 0$. The equation $b^2 = 0$ is equivalent to its particular case $b^2\mathrm{pr}_1 = 0$:

$$\sum_{p+k+q=n} \bigl(1^{\otimes p} \otimes b_k \otimes 1^{\otimes q}\bigr) b_{p+1+q} = 0 : T^n s\mathcal{A} \to s\mathcal{A}, \quad n \geqslant 1.$$

The components b_n also determine operations

$$m_n = \bigl(\mathcal{A}^{\otimes n} \xrightarrow{s^{\otimes n}} (s\mathcal{A})^{\otimes n} \xrightarrow{b_n} s\mathcal{A} \xrightarrow{s^{-1}} \mathcal{A}\bigr)$$

of degree $2-n$. A differential graded category is an example of an A_∞-category for which $b_n = 0$ and $m_n = 0$ if $n > 2$.

An A_∞*-algebra* is an A_∞-category with one object. Even dealing with A_∞-algebras one encounters the (real) A_∞-category, whose objects are homomorphisms of A_∞-algebras. The latter are particular cases of A_∞-functors, as defined below.

2.4. DEFINITION. Let $\mathcal{A}$, $\mathcal{B}$ be A_∞-categories. An *A_∞-functor* $f:\mathcal{A}\to\mathcal{B}$ is an augmented coalgebra morphism $f:Ts\mathcal{A}\to Ts\mathcal{B}$ which commutes with the differential. It is called *strict* if all its components vanish except, possibly, the first, $f_1:s\mathcal{A}\to s\mathcal{B}$.

The commutation equation $fb=bf$ is equivalent to its particular case $fb\mathrm{pr}_1=bf\mathrm{pr}_1$:

$$\sum_{k>0} f_{nk}b_k=\sum_{m>0} b_{nm}f_m:T^ns\mathcal{A}\to s\mathcal{B},\qquad n\geqslant 1.$$

The identity coalgebra morphism $\mathrm{id}:Ts\mathcal{A}\to Ts\mathcal{A}$ provides an example of a strict A_∞-functor. Its first component is the identity map $\mathrm{id}:s\mathcal{A}\to s\mathcal{A}$.

2.5. *Cones*

Let $u:A\to C$, $a\mapsto au$, be a chain map. Its *cone* is defined as the complex $\mathrm{Cone}(u)=A[1]\oplus C$, $\mathrm{Cone}^k(u)=A^{k+1}\oplus C^k$, with the differential $(a,c)d=(ad^{A[1]},au+cd^C)=(-ad^A,au+cd^C)$. Cone is one of the tools of the theory of A_∞-structures.

If a chain map $u:A\to C$ is homotopically invertible, then $\mathrm{Cone}(u)$ is contractible. This can be deduced from triangulatedness of the homotopy category of complexes of $\Bbbk$-modules. For some purposes an explicit homotopy between the identity and the zero endomorphism of $\mathrm{Cone}(u)$ is useful. In order to construct it, consider a chain map $v:C\to A$ homotopically inverse to u. Thus, there are maps $h':A\to A$, $h'':C\to C$ of degree -1 such that $uv=1+h'd^A+d^Ah':A\to A$, $vu=1+h''d^C+d^Ch'':C\to C$. Define a map $h:\mathrm{Cone}(u)\to\mathrm{Cone}(u)$ of degree -1 by the formula $(a,c)h=(ah'+cv,-ch'')$. One computes easily that $hd+dh=1-f:\mathrm{Cone}(u)\to\mathrm{Cone}(u)$, where the map $f:\mathrm{Cone}(u)\to\mathrm{Cone}(u)$ is given as $(a,c)f=(0,auh''-ah'u)$. We conclude that f is a chain map homotopic to the identity map. The chain of equivalences $\mathrm{id}_{\mathrm{Cone}(u)}\sim f=\mathrm{id}_{\mathrm{Cone}(u)}f\sim f^2=0$ proves that $\mathrm{Cone}(u)$ is contractible. It gives also an explicit contracting homotopy – the map $\underline{h}=h+hf:\mathrm{Cone}(u)\to\mathrm{Cone}(u)$ of degree -1, which satisfies $\mathrm{id}_{\mathrm{Cone}(u)}=\underline{h}d+d\underline{h}$. Contractibility of $\mathrm{Cone}(u)$ is used e.g. in the proof of the following result.

2.6. PROPOSITION. *Let $(s\mathcal{C},d)$ be a differential graded quiver, let $\mathcal{B}$ be an A_∞-category, and let $f_1:(s\mathcal{C},d)\to(s\mathcal{B},b_1)$ be a chain quiver morphism such that the chain maps $f_1:s\mathcal{C}(X,Y)\to s\mathcal{B}(Xf_1,Yf_1)$ are homotopy invertible for all pairs X, Y of objects of $\mathcal{C}$. Then there is an A_∞-category structure on $\mathcal{C}$ such that $b_1^{\mathcal{C}}=d$ and an A_∞-functor $f:\mathcal{C}\to\mathcal{B}$, whose first component is the given morphism f_1.*

PROOF. Denote $\mathrm{Ob}\,f_1:\mathrm{Ob}\,\mathcal{C}\to\mathrm{Ob}\,\mathcal{B}$ simply by f. Let us construct the components of the $(1,1)$-coderivation $b:Ts\mathcal{C}\to Ts\mathcal{C}$ of degree 1 and of the augmented coalgebra homomorphism $f:Ts\mathcal{C}\to Ts\mathcal{B}$ by induction. We already know the components $b_1=d$ and f_1. Given an integer $n\geqslant 2$, assume that we have already found the components b_m, f_m of the sought for $b^{\mathcal{C}}$ and f for $m<n$, such that the equations

$$\left(b^2\right)_m=0:T^ms\mathcal{C}(X,Y)\to s\mathcal{C}(X,Y),\tag{2.6.1}$$

$$(fb)_m = (bf)_m : T^m s\mathcal{C}(X,Y) \to s\mathcal{B}(Xf,Yf) \tag{2.6.2}$$

are satisfied for all $m < n$. Introduce a $(1,1)$-coderivation $\tilde{b} : Ts\mathcal{C} \to Ts\mathcal{C}$ of degree 1 by its components $(0, b_1, \dots, b_{n-1}, 0, 0, \dots)$ and an augmented coalgebra homomorphism $\tilde{f} : Ts\mathcal{C} \to Ts\mathcal{B}$ with $\operatorname{Ob}\tilde{f} = \operatorname{Ob} f_1 = f$ by its components $(f_1, \dots, f_{n-1}, 0, 0, \dots)$. Then $\lambda \overset{\text{def}}{=} \tilde{b}^2 : Ts\mathcal{C} \to Ts\mathcal{C}$ is a $(1,1)$-coderivation of degree 2 and $\nu \overset{\text{def}}{=} -\tilde{f}b^{\mathcal{B}} + \tilde{b}\tilde{f} : Ts\mathcal{C} \to Ts\mathcal{B}$ is an $(\tilde{f},\tilde{f})$-coderivation of degree 1. Equations (2.6.1), (2.6.2) imply that $\lambda_m = 0$, $\nu_m = 0$ for $m < n$. The n-th components equal

$$\lambda_n = \sum_{a+k+c=n}^{1<k<n} \bigl(1^{\otimes a} \otimes b_k \otimes 1^{\otimes c}\bigr) b_{a+1+c},$$

$$\nu_n = -\sum_{i_1+\dots+i_k=n}^{1<k\leqslant n} (f_{i_1} \otimes \dots \otimes f_{i_k}) b_k^{\mathcal{B}} + \sum_{a+k+c=n}^{1<k<n} \bigl(1^{\otimes a} \otimes b_k \otimes 1^{\otimes c}\bigr) f_{a+1+c}.$$

Equations (2.6.1), (2.6.2) for $m = n$ take the form

$$-b_n b_1 - \sum_{a+1+c=n} \bigl(1^{\otimes a} \otimes b_1 \otimes 1^{\otimes c}\bigr) b_n = \lambda_n : T^n s\mathcal{C} \to s\mathcal{C}, \tag{2.6.3}$$

$$f_n b_1 - \sum_{a+1+c=n} \bigl(1^{\otimes a} \otimes b_1 \otimes 1^{\otimes c}\bigr) f_n - b_n f_1 = \nu_n : T^n s\mathcal{C} \to s\mathcal{B}. \tag{2.6.4}$$

Write $N = T^n s\mathcal{C}(X,Y)$ for arbitrary objects X, Y of $\mathcal{C}$ and consider the chain map

$$u = -\mathsf{C}_{\Bbbk}(N, f_1) : \mathsf{C}_{\Bbbk}\bigl(N, s\mathcal{C}(X,Y)\bigr) \to \mathsf{C}_{\Bbbk}\bigl(N, s\mathcal{B}(Xf,Yf)\bigr).$$

Since f_1 is homotopy invertible, the map u is homotopy invertible as well. Therefore, the complex $\operatorname{Cone}(u)$ is contractible, in particular, acyclic. Equations (2.6.3) and (2.6.4) have the form $-b_n d = \lambda_n$, $f_n d + b_n u = \nu_n$, that is, the element

$$(\lambda_n, \nu_n) \in \mathsf{C}^2_{\Bbbk}\bigl(N, s\mathcal{C}(X,Y)\bigr) \oplus \mathsf{C}^1_{\Bbbk}\bigl(N, s\mathcal{B}(Xf,Yf)\bigr) = \operatorname{Cone}^1(u)$$

has to be the boundary of the sought element

$$(b_n, f_n) \in \mathsf{C}^1_{\Bbbk}\bigl(N, s\mathcal{C}(X,Y)\bigr) \oplus \mathsf{C}^0_{\Bbbk}\bigl(N, s\mathcal{B}(Xf,Yf)\bigr) = \operatorname{Cone}^0(u).$$

These equations are solvable because (λ_n, ν_n) is a cycle in $\operatorname{Cone}^1(u)$. Indeed, the equations to verify $-\lambda_n d = 0$, $\nu_n d + \lambda_n u = 0$ take the form

$$-\lambda_n b_1 + \sum_{p+1+q=n} \bigl(1^{\otimes p} \otimes b_1 \otimes 1^{\otimes q}\bigr) \lambda_n = 0 : T^n s\mathcal{C} \to s\mathcal{C},$$

$$\nu_n b_1 + \sum_{p+1+q=n} \bigl(1^{\otimes p} \otimes b_1 \otimes 1^{\otimes q}\bigr) \nu_n - \lambda_n f_1 = 0 : T^n s\mathcal{C} \to s\mathcal{B}.$$

The first equation follows by composing the identity $-\lambda\tilde{b} + \tilde{b}\lambda = 0 : T^n s\mathcal{C} \to Ts\mathcal{C}$ with $\mathrm{pr}_1 : Ts\mathcal{C} \to s\mathcal{C}$. The second equation follows by composing the identity $\nu b^{\mathcal{B}} + \tilde{b}\nu - \tilde{b}^2\tilde{f} = 0 : T^n s\mathcal{C} \to Ts\mathcal{B}$ with $\mathrm{pr}_1 : Ts\mathcal{B} \to s\mathcal{B}$. Thus, the required pair (b_n, f_n) exists and we proceed by induction. □

Since one knows the contracting homotopy $\underline{h}$ of $\operatorname{Cone}(u)$ by Section 2.5, one can write down recursive formulas for b_n, f_n, cf. Gugenheim and Stasheff [13], Merkulov [28], and to express them in terms of trees, cf. Kontsevich and Soibelman [21, Section 6.4]. The above result applies, in particular, if the graded $\Bbbk$-modules $\mathcal{C}(X,Y) = H(\mathcal{B}(X,Y), m_1)$ with zero differential are homotopy isomorphic to the complexes $(\mathcal{B}(X,Y), m_1)$. In this case it is possible to transfer the A_∞-category structure from $\mathcal{B}$ to its homology $\mathcal{C} = H(\mathcal{B})$, cf. Kadeishvili [15]. Notice that $b_1^{\mathcal{C}} = 0$. A_∞-categories with this property are called *minimal*. The composition m_2 in minimal A_∞-categories is strictly associative.

In the unital case the transferred A_∞-category structure is unique up to an equivalence, see Corollary 4.8.

2.7. *A_∞-modules over A_∞-algebras*

The structure of a right or left A_∞-module over an A_∞-algebra is defined similarly to that of an A_∞-algebra itself, see Keller [17–19], Lefèvre-Hasegawa [23] and Smirnov [32]. A_∞-bimodules over an A_∞-algebra and their morphisms are defined by Tradler [35]. He also defines an ∞-inner-product on an A_∞-algebra A to be an A_∞-bimodule-map from the A_∞-bimodule A to the A_∞-bimodule A^*, and relates such inner-products with a certain graph complex.

2.8. *Unital A_∞-categories and unital A_∞-functors*

An A_∞-category $\mathcal{C}$ is called *unital* if for each object X of $\mathcal{C}$ there is a unit element, i.e., a chain map $1_X : \Bbbk \to \mathcal{C}(X,X)$ such that the chain maps $(\mathrm{id} \otimes 1_Y)m_2, (1_X \otimes \mathrm{id})m_2 : \mathcal{C}(X,Y) \to \mathcal{C}(X,Y)$ are homotopic to the identity map. In other terms, for each object X of $\mathcal{C}$ there is a *unit element* ${}_X\mathbf{i}_0^{\mathcal{C}} = 1_X s : \Bbbk \to (s\mathcal{C})^{-1}(X,X)$ – a cycle defined up to a boundary, such that the chain maps $(\mathrm{id} \otimes {}_Y\mathbf{i}_0^{\mathcal{C}})b_2, -({}_X\mathbf{i}_0^{\mathcal{C}} \otimes \mathrm{id})b_2 : s\mathcal{C}(X,Y) \to s\mathcal{C}(X,Y)$ are homotopic to the identity map.

A $\Bbbk$-linear category $H^0(\mathcal{C})$ (the homotopy category) is associated with a unital A_∞-category $\mathcal{C}$. It has the set of objects $\operatorname{Ob} H^0(\mathcal{C}) = \operatorname{Ob}\mathcal{C}$ and the $\Bbbk$-modules of morphisms $H^0(\mathcal{C})(X,Y) = H^0(\mathcal{C}(X,Y), m_1)$. The composition in $H^0(\mathcal{C})$ is induced by m_2, and the unit elements are equivalence classes of the cycles $1_X = {}_X\mathbf{i}_0^{\mathcal{C}} s^{-1}$.

An A_∞-category $\mathcal{A}$ is called *strictly unital* if for each object $X \in \operatorname{Ob}\mathcal{A}$ there is a strict unit, that is, a $\Bbbk$-linear map ${}_X\mathbf{i}_0^{\mathcal{A}} : \Bbbk \to (s\mathcal{A})^{-1}(X,X)$ such that ${}_X\mathbf{i}_0^{\mathcal{A}} b_1 = 0$ and the following conditions are satisfied: for all pairs X, Y of objects of $\mathcal{A}$ the chain maps $(\mathrm{id} \otimes {}_Y\mathbf{i}_0^{\mathcal{A}})b_2, -({}_X\mathbf{i}_0^{\mathcal{A}} \otimes \mathrm{id})b_2 : s\mathcal{A}(X,Y) \to s\mathcal{A}(X,Y)$ are equal to the identity map and $(\cdots \otimes \mathbf{i}_0^{\mathcal{A}} \otimes \cdots)b_n = 0$ if $n \geqslant 3$. For example, differential graded categories are strictly unital.

Given an A_∞-category $\mathcal{A}$, we associate a strictly unital A_∞-category $\mathcal{A}^{\mathsf{su}}$ with it. It has the same set of objects and for any pair of objects $X, Y \in \operatorname{Ob}\mathcal{A}$ the graded $\Bbbk$-module $s\mathcal{A}^{\mathsf{su}}(X,Y)$ is given by

$$s\mathcal{A}^{\mathsf{su}}(X,Y) = \begin{cases} s\mathcal{A}(X,Y), & X \neq Y, \\ s\mathcal{A}(X,X) \oplus \Bbbk\,{}_X\mathbf{i}_0^{\mathcal{A}^{\mathsf{su}}}, & X = Y, \end{cases}$$

where ${}_X\mathbf{i}_0^{\mathcal{A}^{\mathsf{su}}}$ is a new generator of degree -1. The element ${}_X\mathbf{i}_0^{\mathcal{A}^{\mathsf{su}}}$ is a strict unit by definition, and the canonical embedding $e_{\mathcal{A}} = u_{\mathsf{su}} : \mathcal{A} \hookrightarrow \mathcal{A}^{\mathsf{su}}$ is a strict A_∞-functor, that is, all compositions b_n in $\mathcal{A}$ and $\mathcal{A}^{\mathsf{su}}$ agree.

The above definition of unitality is not the only possible. We shall give two more definitions, which turn out to be equivalent to the above.

2.9. Definition *(Kontsevich).* A *unital structure* on an A_∞-category $\mathcal{A}$ is a choice of an A_∞-functor $U_{\mathsf{su}}^{\mathcal{A}} : \mathcal{A}^{\mathsf{su}} \to \mathcal{A}$ such that

$$\big(\mathcal{A} \overset{e_{\mathcal{A}}}{\hookrightarrow} \mathcal{A}^{\mathsf{su}} \xrightarrow{U_{\mathsf{su}}^{\mathcal{A}}} \mathcal{A}\big) = \mathrm{id}_{\mathcal{A}}.$$

2.10. Theorem. *(See [3].) An A_∞-category is unital if and only if it admits a unital structure.*

2.11. Definition. An A_∞-category $\mathcal{C}$ is called *homotopy unital* in the sense of Fukaya [9, Definition 5.11] if the graded $\Bbbk$-quiver

$$\mathcal{C}^+ = \mathcal{C} \oplus \Bbbk\mathcal{C} \oplus s\Bbbk\mathcal{C}$$

(with $\operatorname{Ob}\mathcal{C}^+ = \operatorname{Ob}\mathcal{C}$) has an A_∞-structure b^+ of the following kind. Denote the generators of the second and the third summands of $s\mathcal{C}^+ = s\mathcal{C} \oplus s\Bbbk\mathcal{C} \oplus s^2\Bbbk\mathcal{C}$ by ${}_X\mathbf{i}_0^{\mathcal{C}^{\mathsf{su}}} = 1s$ and $\mathbf{j}_X^{\mathcal{C}} = 1s^2$ of degree respectively -1 and -2 for $X \in \operatorname{Ob}\mathcal{C}$. The conditions on b^+ are:

(i) the elements ${}_X\mathbf{i}_0^{\mathcal{C}} \overset{\mathrm{def}}{=} {}_X\mathbf{i}_0^{\mathcal{C}^{\mathsf{su}}} - \mathbf{j}_X^{\mathcal{C}} b_1^+$ belong to $s\mathcal{C}(X, X)$ for all $X \in \operatorname{Ob}\mathcal{C}$;
(ii) the A_∞-category $\mathcal{C}^+$ is strictly unital with the strict units $\mathbf{i}_0^{\mathcal{C}^{\mathsf{su}}}$;
(iii) the embedding $\mathcal{C} \hookrightarrow \mathcal{C}^+$ is a strict A_∞-functor;
(iv) $(s\mathcal{C} \oplus s^2\Bbbk\mathcal{C})^{\otimes n} b_n^+ \subset s\mathcal{C}$ for any $n > 1$.

In particular, $\mathcal{C}^+$ contains the strictly unital envelope $\mathcal{C}^{\mathsf{su}} = \mathcal{C} \oplus \Bbbk\mathcal{C}$ of $\mathcal{C}$.

Let $\mathcal{D}$ be a strictly unital A_∞-category with strict units $\mathbf{i}_0^{\mathcal{D}}$. Then it has a canonical homotopy unital structure $(\mathcal{D}^+, b^+)$. Namely, $\mathbf{j}_X^{\mathcal{D}} b_1^+ = {}_X\mathbf{i}_0^{\mathcal{D}^{\mathsf{su}}} - {}_X\mathbf{i}_0^{\mathcal{D}}$ and b_n^+ vanishes for all $n > 1$ on all summands of $(s\mathcal{D} \oplus s^2\Bbbk\mathcal{D})^{\otimes n}$ except on $s\mathcal{D}^{\otimes n}$, where it coincides with $b_n^{\mathcal{D}}$.

2.12. Theorem. *(See [3].) An A_∞-category $\mathcal{C}$ is unital if and only if it is homotopy unital. Moreover, if $\mathcal{C}$ is unital and $\mathbf{i}_0^{\mathcal{C}}$ are its unit elements, then $\mathcal{C}$ admits a homotopy unital structure $(\mathcal{C}^+, b^+)$ with $\mathbf{j}^{\mathcal{C}} b_1^+ = \mathbf{i}_0^{\mathcal{C}^{\mathsf{su}}} - \mathbf{i}_0^{\mathcal{C}}$. Conversely, if $\mathcal{C}$ is homotopy unital, then the elements $\mathbf{i}_0^{\mathcal{C}} \overset{\mathrm{def}}{=} \mathbf{i}_0^{\mathcal{C}^{\mathsf{su}}} - \mathbf{j}^{\mathcal{C}} b_1^+$ are unit elements of $\mathcal{C}$.*

Proof of this theorem is based on the fact that for any unital A_∞-category $\mathcal{C}$ there is an A_∞-functor $\phi : \mathcal{C} \to \mathcal{D}$ to a differential graded category $\mathcal{D}$ with $\operatorname{Ob}\mathcal{D} = \operatorname{Ob}\mathcal{C}$, $\operatorname{Ob}\phi = \mathrm{id}_{\operatorname{Ob}\mathcal{C}}$ such that $\phi_1 : s\mathcal{C} \to s\mathcal{D}$ is homotopy invertible, see Section 4.8.1. Applying Proposition 2.6 to the canonically extended A_∞-category $\mathcal{D}^+$ and to a certain homotopy isomorphism $\phi_1^+ : s\mathcal{C}^+ \to s\mathcal{D}^+$ we can deduce existence of some A_∞-structure on $\mathcal{C}^+$. To satisfy all conditions of Definition 2.11 one has to refine the proof of Proposition 2.6.

2.13. Definition. Let $\mathcal{A}$, $\mathcal{B}$ be unital A_∞-categories. An A_∞-functor $f:\mathcal{A}\to\mathcal{B}$ is *unital* if ${}_X\mathbf{i}_0^{\mathcal{A}} f_1 - {}_{Xf}\,\mathbf{i}_0^{\mathcal{B}} \in \operatorname{Im} b_1$ for all objects X of $\mathcal{A}$.

To a unital A_∞-functor $f:\mathcal{A}\to\mathcal{B}$ a $\Bbbk$-linear functor $H^0(f):H^0(\mathcal{A})\to H^0(\mathcal{B})$ is assigned. Namely, $\operatorname{Ob} H^0(f)=\operatorname{Ob} f:\operatorname{Ob}\mathcal{A}\to\operatorname{Ob}\mathcal{B}$, and for each pair of objects X,Y of $\mathcal{A}$ the map $H^0(f):H^0(\mathcal{A})(X,Y)\to H^0(\mathcal{B})(Xf,Yf)$ is induced by the chain map $sf_1s^{-1}:\mathcal{A}(X,Y)\to\mathcal{B}(Xf,Yf)$.

3. Multicategories

A_∞-categories and A_∞-functors form a closed multicategory. This permits us to apply enriched multicategory methods to the study of A_∞-categories. We are going to define the relevant notions.

3.1. *Lax Monoidal categories and functors*

The term ‘Monoidal’ as opposed to ‘monoidal’ indicates that categories are equipped with n-ary tensor products, related by many associativity morphisms. The latter are invertible for Monoidal categories and not necessarily invertible for lax Monoidal categories. The same convention applies to functors.

3.2. Definition. A *lax (symmetric) Monoidal category* $(\mathcal{V},\otimes^I_{\mathcal{V}},\lambda^f_{\mathcal{V}})$ consists of

1. A category $\mathcal{V}$.
2. A functor $\otimes^I=\otimes^I_{\mathcal{V}}:\mathcal{V}^I\to\mathcal{V}$, for every set $I\in\operatorname{Ob}\mathcal{S}$, such that $\otimes^I=\mathrm{Id}_{\mathcal{V}}$ for each 1-element set I.[1] In particular, a map $\otimes^I_{\mathcal{V}}:\prod_{i\in I}\mathcal{V}(X_i,Y_i)\to\mathcal{V}(\otimes^{i\in I}X_i,\otimes^{i\in I}Y_i)$ is given.

   For a map $f:I\to J$ in $\operatorname{Mor}\mathcal{O}$ (respectively $\operatorname{Mor}\mathcal{S}$) introduce a functor $\otimes^f=\otimes^f_{\mathcal{V}}:\mathcal{V}^I\to\mathcal{V}^J$ which to a function $X:I\to\operatorname{Ob}\mathcal{V}$, $i\mapsto X_i$, assigns the function $J\to\operatorname{Ob}\mathcal{V}$, $j\mapsto\otimes^{i\in f^{-1}(j)}X_i$. The linear order on $f^{-1}(j)$ is induced by the embedding $f^{-1}(j)\hookrightarrow I$. The functor $\otimes^f_{\mathcal{V}}:\mathcal{V}^I\to\mathcal{V}^J$ acts on morphisms via the map

$$\prod_{i\in I}\mathcal{V}(X_i,Y_i) \xrightarrow{\sim} \prod_{j\in J}\prod_{i\in f^{-1}j}\mathcal{V}(X_i,Y_i) \xrightarrow{\prod_{j\in J}\otimes^{f^{-1}j}} \prod_{j\in J}\mathcal{V}\bigl(\otimes^{i\in f^{-1}j}X_i,\otimes^{i\in f^{-1}j}Y_i\bigr).$$

3. A morphism of functors

$$\lambda^f:\otimes^I\to\otimes^J\circ\otimes^f:\mathcal{V}^I\to\mathcal{V},\qquad \lambda^f:\otimes^{i\in I}X_i\to\otimes^{j\in J}\otimes^{i\in f^{-1}j}X_i,$$

   for every map $f:I\to J$ in $\operatorname{Mor}\mathcal{O}$ (respectively $\operatorname{Mor}\mathcal{S}$),

[1] See Section 1.1 above for the definition of the categories $\mathcal{S}$ and $\mathcal{O}$.

such that

(i) for all sets $I \in \operatorname{Ob}\mathcal{O}$, for all 1-element sets J

$$\lambda^{\mathrm{id}_I} = \mathrm{id}, \qquad \lambda^{I\to J} = \mathrm{id};$$

(ii) for any pair of composable maps $I \xrightarrow{f} J \xrightarrow{g} K$ from $\mathcal{O}$ (respectively from $\mathcal{S}$) the following equation holds:

$$\begin{tikzcd}
\otimes^{i\in I} X_i \arrow[rr, "\lambda^f"] \arrow[d, "\lambda^{fg}"'] & & \otimes^{j\in J} \otimes^{i\in f^{-1}j} X_i \arrow[d, "\lambda^g"] \\
\otimes^{k\in K} \otimes^{i\in f^{-1}g^{-1}k} X_i \arrow[rr, "\otimes^{k\in K}\lambda^{f|:f^{-1}g^{-1}k\to g^{-1}k}"] & = & \otimes^{k\in K} \otimes^{j\in g^{-1}k} \otimes^{i\in f^{-1}j} X_i
\end{tikzcd}$$

A *Monoidal* (respectively *symmetric Monoidal*) *category* is a lax one for which all λ^f are isomorphisms.

The basic example of a symmetric Monoidal category is the category $\mathcal{S}et$ of sets, equipped with direct product functors $(X_i)_{i\in I} \mapsto \prod_{i\in I} X_i$. Sometimes this is reflected in notation used.

3.2.1. *Monoidal categories of graded quivers* A *biunital quiver* is a quiver $\mathcal{C}$ together with a pair of morphisms $\Bbbk\mathcal{C} \xrightarrow{\eta} \mathcal{C} \xrightarrow{\varepsilon} \Bbbk\mathcal{C}$ in $\mathcal{Q}/\operatorname{Ob}\mathcal{C}$, whose composition is $\eta\varepsilon = \mathrm{id}_{\Bbbk\mathcal{C}}$.

Biunital quivers with arbitrary $\mathcal{U}$-small sets of objects form a category denoted $\mathcal{Q}_{bu}$. A *morphism* from a biunital quiver $\varepsilon : \mathcal{A} \rightleftarrows \Bbbk\mathcal{A} : \eta$ to a biunital quiver $\varepsilon : \mathcal{B} \rightleftarrows \Bbbk\mathcal{B} : \eta$ is a quiver morphism $f : \mathcal{A} \to \mathcal{B}$ such that $f\varepsilon = \varepsilon(\Bbbk f)$, $\eta f = (\Bbbk f)\eta$. The category $\mathcal{Q}_{bu}$ is equivalent to the category of quivers $\mathcal{Q}$ via the functors $\mathcal{Q}_{bu} \to \mathcal{Q}$, $(\varepsilon : \mathcal{C} \rightleftarrows \Bbbk\mathcal{C} : \eta) \mapsto \bar{\mathcal{C}} = \operatorname{Im}(1 - \varepsilon\eta) = \operatorname{Ker}\varepsilon$, and $\mathcal{Q} \to \mathcal{Q}_{bu}$, $\mathcal{A} \mapsto T^{\leqslant 1}\mathcal{A} = (\mathrm{pr}_1 : \Bbbk\mathcal{A} \oplus \mathcal{A} \rightleftarrows \Bbbk\mathcal{A} : \mathrm{in}_1)$, quasi-inverse to each other.

There is a faithful (forgetful) functor $F : \mathcal{Q}_{bu} \to \mathcal{Q}$, $(\varepsilon : \mathcal{C} \rightleftarrows \Bbbk\mathcal{C} : \eta) \mapsto \mathcal{C}$. The category $\mathcal{Q}_{bu}$ inherits a symmetric Monoidal structure from $\mathcal{Q}$ via F, namely,

$$\boxtimes^{i\in I}_{bu}\big(\mathcal{A}_i \underset{\eta}{\overset{\varepsilon}{\rightleftarrows}} \Bbbk\mathcal{A}_i\big) \overset{\text{def}}{=} \Big(\boxtimes^{i\in I}\mathcal{A}_i \underset{\boxtimes^I\eta}{\overset{\boxtimes^I\varepsilon}{\rightleftarrows}} \boxtimes^{i\in I}\Bbbk\mathcal{A}_i \underset{\sim}{\overset{\lambda^{\varnothing\to I}}{\longleftarrow}} \Bbbk\big(\boxtimes^{i\in I}\mathcal{A}_i\big)\Big).$$

The unit object is the one-object quiver

$$\varepsilon : \Bbbk = \Bbbk : \eta,$$

and the tensor product of morphisms f_i of biunital quivers is $\boxtimes^{i\in I} f_i$.

This symmetric Monoidal structure $(\mathcal{Q}_{bu}, \boxtimes^I_{bu}, \lambda^f_{bu})$ translates via the equivalence $T^{\leqslant 1} : \mathcal{Q} \to \mathcal{Q}_{bu}$ to a new symmetric Monoidal structure $\mathcal{Q}_u = (\mathcal{Q}, \boxtimes^I_u, \lambda^f_u)$ on $\mathcal{Q}$. Explicitly it is given by

$$\boxtimes^{i\in I}_u \mathcal{A}_i = \bigoplus_{j_i\in\{0,1\}, i\in I}^{\sum_i j_i > 0} \boxtimes^{i\in I} T^{j_i}\mathcal{A}_i = \bigoplus_{\varnothing\neq S\subset I} \boxtimes^{i\in I} T^{\chi(i\in S)}\mathcal{A}_i,$$

where $\chi(i \in S) = 1$ if $i \in S$, and $\chi(i \in S) = 0$ if $i \notin S$. In particular,

$$\operatorname{Ob} \boxtimes_u^{i\in I} \mathcal{A}_i = \operatorname{Ob} \boxtimes^{i\in I} \mathcal{A}_i = \prod_{i\in I} \operatorname{Ob} \mathcal{A}_i .$$

The following canonical isomorphism is implied by the additivity of $\boxtimes$,

$$\vartheta^I = \Bigl[\boxtimes^{i\in I} \bigl(T^{\leqslant 1} \mathcal{A}_i\bigr) \xrightarrow{\sim} \bigoplus_{j_i \in \{0,1\}, i\in I} \boxtimes^{i\in I} T^{j_i} \mathcal{A}_i$$

$$\xrightarrow{\sim} T^0\bigl(\boxtimes^{i\in I} \mathcal{A}_i\bigr) \oplus \boxtimes_u^{i\in I} \mathcal{A}_i = T^{\leqslant 1}\bigl(\boxtimes_u^{i\in I} \mathcal{A}_i\bigr) \Bigr],$$

and by the isomorphism $\boxtimes^{i\in I} T^0 \mathcal{A}_i \xrightarrow{\sim} T^0(\boxtimes^{i\in I} \mathcal{A}_i)$, (given by) the identifications $(\lambda^{\varnothing\to I})^{-1} : \otimes^I \Bbbk \xrightarrow{\sim} \Bbbk$. Actually, $\vartheta^I : \boxtimes_{bu}^{i\in I} T^{\leqslant 1} \mathcal{A}_i \to T^{\leqslant 1} \boxtimes_u^{i\in I} \mathcal{A}_i$ is an isomorphism of biunital quivers.

In particular, the unit object $\mathbb{1}_u = \boxtimes_u^{\varnothing}()$ of $\mathscr{Q}_u$ is the quiver with a unique object $*$ and zero module of homomorphisms, and

$$(\mathcal{A} \boxtimes_u \mathcal{B})\bigl((A, B), (A', B')\bigr)$$
$$\simeq \begin{cases} \mathcal{A}(A, A') \otimes \mathcal{B}(B, B'), \\ \quad A \neq A', \; B \neq B', \\ \mathcal{A}(A, A') \otimes \mathcal{B}(B, B') \oplus \Bbbk \otimes \mathcal{B}(B, B'), \\ \quad A = A', \; B \neq B', \\ \mathcal{A}(A, A') \otimes \mathcal{B}(B, B') \oplus \mathcal{A}(A, A') \otimes \Bbbk, \\ \quad A \neq A', \; B = B', \\ \mathcal{A}(A, A') \otimes \mathcal{B}(B, B') \oplus \mathcal{A}(A, A') \otimes \Bbbk \oplus \Bbbk \otimes \mathcal{B}(B, B'), \\ \quad A = A', B = B'. \end{cases}$$

3.3. DEFINITION. A *lax* (*symmetric*) *Monoidal functor* between lax (symmetric) Monoidal categories

$$\bigl(F, \phi^I\bigr) : \bigl(\mathcal{C}, \otimes_{\mathcal{C}}^I, \lambda_{\mathcal{C}}^f\bigr) \to \bigl(\mathcal{D}, \otimes_{\mathcal{D}}^I, \lambda_{\mathcal{D}}^f\bigr)$$

consists of

(i) a functor $F : \mathcal{C} \to \mathcal{D}$,

(ii) a functorial morphism for each set $I \in \operatorname{Ob} \mathcal{S}$

$$\phi^I : \otimes_{\mathcal{D}}^I \circ F^I \to F \circ \otimes_{\mathcal{C}}^I : \mathcal{C}^I \to \mathcal{D}, \quad \phi^I : \otimes_{\mathcal{D}}^{i\in I} F X_i \to F \otimes_{\mathcal{C}}^{i\in I} X_i,$$

such that $\phi^I = \mathrm{id}_F$ for each 1-element set I, and for every map $f : I \to J$ of $\mathcal{O}$ (respectively $\mathcal{S}$) and all families $(X_i)_{i\in I}$ of objects of $\mathcal{C}$ the following equation holds:

$$\begin{tikzcd}
\otimes_{\mathcal{D}}^{i\in I} F X_i \arrow[rr, "\phi^I"] \arrow[d, "\lambda_{\mathcal{D}}^f"'] \arrow[dr, phantom, "="] & & F \otimes_{\mathcal{C}}^{i\in I} X_i \arrow[d, "F\lambda_{\mathcal{C}}^f"] \\
\otimes_{\mathcal{D}}^{j\in J} \otimes_{\mathcal{D}}^{i\in f^{-1}j} F X_i \arrow[r, "\otimes_{\mathcal{D}}^{j\in J} \phi^{f^{-1}j}"] & \otimes_{\mathcal{D}}^{j\in J} F \otimes_{\mathcal{C}}^{i\in f^{-1}j} X_i \arrow[r, "\phi^J"] & F \otimes_{\mathcal{C}}^{j\in J} \otimes_{\mathcal{C}}^{i\in f^{-1}j} X_i
\end{tikzcd}$$

The pair

$$\bigl(T^{\leqslant 1},\vartheta^I\bigr):\mathscr{Q}_u=\bigl(\mathscr{Q},\boxtimes^I_u,\lambda^f_u\bigr)\to\bigl(\mathscr{Q},\boxtimes^I,\lambda^f\bigr)=\mathscr{Q}_p$$

is an example of a symmetric Monoidal functor.

Another example is the functor $T:\mathscr{Q}\to\mathscr{Q}$ which admits a lax symmetric Monoidal structure

$$\bigl(T,\widetilde{\tau}^I\bigr):\mathscr{Q}_u=\bigl(\mathscr{Q},\boxtimes^I_u,\lambda^f_u\bigr)\to\bigl(\mathscr{Q},\boxtimes^I,\lambda^f\bigr)=\mathscr{Q}_p.$$

Let $\mathcal{A}_i$, $i\in I$, be graded quivers. Then

$$\boxtimes^{i\in I}(T\mathcal{A}_i)=\bigoplus_{(m_i)\in\mathbb{Z}^I_{\geqslant 0}}\boxtimes^{i\in I}T^{m_i}\mathcal{A}_i$$

is a direct sum over $(m_i)\in\mathbb{Z}^I_{\geqslant 0}$. On the other hand,

$$T\bigl(\boxtimes^{i\in I}_u\mathcal{A}_i\bigr)=\bigoplus_{m=0}^{\infty}T^m\bigl(\boxtimes^{i\in I}_u\mathcal{A}_i\bigr)=\bigoplus_{m=0}^{\infty}\bigoplus_{S\subset I\times\mathbf{m}}^{\mathrm{pr}_2S=\mathbf{m}}\otimes^{p\in\mathbf{m}}\boxtimes^{i\in I}\,T^{\chi((i,p)\in S)}\mathcal{A}_i$$

decomposes into direct sum over pairs (m,S), where $m\in\mathbb{Z}_{\geqslant 0}$, and the subset $S\subset I\times\mathbf{m}$ satisfies the condition $\mathrm{pr}_2S=\mathbf{m}$. Define $\widetilde{\tau}^I:\boxtimes^{i\in I}(T\mathcal{A}_i)\to T(\boxtimes^{i\in I}_u\mathcal{A}_i)$ to be the identity map on objects $(X_i)_{i\in I}$. Define the only non-trivial matrix coefficients of $\widetilde{\tau}^I$ to be the isomorphisms

$$\begin{aligned}\widetilde{\tau}^I\ &:\boxtimes^{i\in I}T^{m_i}\mathcal{A}_i\xrightarrow{\boxtimes^{i\in I}\lambda^{S_i\hookrightarrow\mathbf{m}}}\boxtimes^{i\in I}\otimes^{p\in\mathbf{m}}\,T^{\chi((i,p)\in S)}\mathcal{A}_i\\&\xrightarrow{\bar{\varkappa}^{-1}}\otimes^{p\in\mathbf{m}}\boxtimes^{i\in I}\,T^{\chi((i,p)\in S)}\mathcal{A}_i,\end{aligned}$$

where the $S_i=\{p\in\mathbf{m}\mid(i,p)\in S\}$ satisfy the condition $|S_i|=m_i$ for all $i\in I$. Here $\bar{\varkappa}$ is given by (1.1.1). If $|S_i|\neq m_i$ for some i, the corresponding matrix coefficient of $\widetilde{\tau}^I$ vanishes.

The endofunctor $T^{\geqslant 1}:\mathscr{Q}\to\mathscr{Q}$ admits a unique lax symmetric Monoidal structure $(T^{\geqslant 1},\tau^I):\mathscr{Q}_u=(\mathscr{Q},\boxtimes^I_u,\lambda^f_u)\to(\mathscr{Q},\boxtimes^I_u,\lambda^f_u)=\mathscr{Q}_u$ such that

$$\bigl(T,\widetilde{\tau}^I\bigr)=\bigl(T^{\leqslant 1},\vartheta^I\bigr)\circ\bigl(T^{\geqslant 1},\tau^I\bigr):\bigl(\mathscr{Q},\boxtimes^I_u,\lambda^f_u\bigr)\to\bigl(\mathscr{Q},\boxtimes^I,\lambda^f\bigr).$$

The transformation τ can be computed via the following formula

$$\tau=\Biggl(\boxtimes^{i\in I}_uT^{\geqslant 1}\mathcal{A}_i=\bigoplus_{0\neq(m_i)\in\mathbb{Z}^I_{\geqslant 0}}\boxtimes^{i\in I}T^{m_i}\mathcal{A}_i\xrightarrow{\sum\widetilde{\tau}}T^{\geqslant 1}\boxtimes^{i\in I}_u\mathcal{A}_i\Biggr).$$

3.4. Definition. A *lax Monoidal transformation* (morphism of lax (symmetric) Monoidal functors)

$$t:\bigl(F,\phi^I\bigr)\to\bigl(G,\psi^I\bigr):\bigl(\mathcal{C},\otimes^I_{\mathcal{C}},\lambda^f_{\mathcal{C}}\bigr)\to\bigl(\mathcal{D},\otimes^I_{\mathcal{D}},\lambda^f_{\mathcal{D}}\bigr)$$

is a natural transformation $t : F \to G$ such that for every $I \in \operatorname{Ob}\mathcal{S}$

$$\begin{tikzcd}
\otimes^{i\in I}_{\mathcal{D}} FX_i \arrow[r, "\phi^I"] \arrow[d, "\otimes^I t"'] \arrow[dr, phantom, "="] & F\otimes^{i\in I}_{\mathcal{C}} X_i \arrow[d, "t"] \\
\otimes^{i\in I}_{\mathcal{D}} GX_i \arrow[r, "\psi^I"] & G\otimes^{i\in I}_{\mathcal{C}} X_i
\end{tikzcd}$$

3.4.1. *Lax Monoidal comonad of a tensor quiver* The functor $T^{\geqslant 1} : \mathscr{Q} \to \mathscr{Q}$ admits a comultiplication $\Delta : T^{\geqslant 1} \to T^{\geqslant 1}T^{\geqslant 1}$ and a counit $\varepsilon : T^{\geqslant 1} \to \mathrm{Id}$ which are lax Monoidal transformations. In fact, the functor $T^{\geqslant 1}$ is given by the formula

$$T^{\geqslant 1}\mathcal{C}(X,Y) = \bigoplus_{\substack{X_0,\dots,X_m\in\operatorname{Ob}\mathcal{C}\\ X_0=X,\,X_m=Y}}^{m>0} \otimes^{j\in\mathbf{m}}\mathcal{C}(X_{j-1},X_j). \tag{3.4.1}$$

Therefore, its square is

$$T^{\geqslant 1}T^{\geqslant 1}\mathcal{C}(X,Y) = \bigoplus_{\substack{X_0,\dots,X_m\in\operatorname{Ob}\mathcal{C}\\ X_0=X,\,X_m=Y}}^{\substack{n>0\\ g:\mathbf{m}\twoheadrightarrow\mathbf{n}\in\mathcal{O}}} \otimes^{p\in\mathbf{n}}\otimes^{j\in g^{-1}p}\mathcal{C}(X_{j-1},X_j), \tag{3.4.2}$$

where the summation extends over all monotonic surjections $g : \mathbf{m} \twoheadrightarrow \mathbf{n}$ with non-empty $\mathbf{n}$. The comultiplication $\Delta : T^{\geqslant 1} \to T^{\geqslant 1}T^{\geqslant 1}$ is the sum of the morphisms

$$\lambda^g : \otimes^{j\in\mathbf{m}}\mathcal{C}(X_{j-1},X_j) \to \otimes^{p\in\mathbf{n}}\otimes^{j\in g^{-1}p}\mathcal{C}(X_{j-1},X_j). \tag{3.4.3}$$

That is, for each summand of (3.4.2) labeled by a monotonic surjection $g : \mathbf{m} \twoheadrightarrow \mathbf{n}$ there exists a unique summand of (3.4.1) which is mapped to it by λ^g, namely, the summand labeled by the source $\mathbf{m}$ of g. The counit is given by the transformation $\varepsilon = \mathrm{pr}_1 : T^{\geqslant 1} \to \mathrm{Id}$, $\mathrm{pr}_1 : T^{\geqslant 1}\mathcal{A} \to \mathcal{A}$. Therefore, $((T^{\geqslant 1},\tau),\Delta,\varepsilon)$ is a lax Monoidal comonad.

Recall that a $T^{\geqslant 1}$-coalgebra is a graded quiver morphism $\delta : \mathcal{C} \to T^{\geqslant 1}\mathcal{C}$ such that

$$\big(\mathcal{C} \xrightarrow{\delta} T^{\geqslant 1}\mathcal{C} \xrightarrow{T^{\geqslant 1}\delta} T^{\geqslant 1}T^{\geqslant 1}\mathcal{C}\big) = \big(\mathcal{C} \xrightarrow{\delta} T^{\geqslant 1}\mathcal{C} \xrightarrow{\Delta} T^{\geqslant 1}T^{\geqslant 1}\mathcal{C}\big),$$
$$\big(\mathcal{C} \xrightarrow{\delta} T^{\geqslant 1}\mathcal{C} \xrightarrow{\varepsilon} \mathcal{C}\big) = \mathrm{id}.$$

In particular, $\operatorname{Ob}\delta = \mathrm{id}_{\operatorname{Ob}\mathcal{C}}$. Coassociativity of δ implies that it has the form

$$\delta = \big(1, \overline{\Delta}^{(2)}, \overline{\Delta}^{(3)}, \dots, \overline{\Delta}^{(k)}, \dots\big),$$

where $\overline{\Delta}^{(k)} : \mathcal{C} \to \mathcal{C}^{\otimes k}$ is the $(k-1)$-th iteration of some $\overline{\Delta} = \overline{\Delta}^{(2)} : \mathcal{C} \to \mathcal{C}\otimes\mathcal{C}$. One easily finds that the categories of $T^{\geqslant 1}$-coalgebras and cocomplete coalgebras are isomorphic.

The functor T has a $T^{\geqslant 1}$-comodule structure (coaction)

$$\tilde{\Delta} = T^{\leqslant 1}(\Delta) : T \to TT^{\geqslant 1} = T \circ T^{\geqslant 1}.$$

Notice that

$$TT^{\geqslant 1}\mathcal{C} = \oplus_{g:\mathbf{m}\twoheadrightarrow\mathbf{n}} \otimes^{p\in\mathbf{n}}\otimes^{j\in g^{-1}p}\mathcal{C},$$

where the summation extends over all monotonic surjections $g : \mathbf{m} \twoheadrightarrow \mathbf{n}$. The components of $\widetilde{\Delta}$ are again given by (3.4.3).

3.5. *Multicategories and multifunctors*

As defined by Lambek [22], a plain (respectively symmetric) *multicategory* C is a set $\operatorname{Ob}\mathsf{C}$ of objects, together with sets of multimorphisms $\mathsf{C}((X_i)_{i\in I}; Y)$, assigned to each map $I \sqcup \mathbf{1} \to \operatorname{Ob}\mathsf{C}$, $i \mapsto X_i$, $1 \mapsto Y$, where $I \in \operatorname{Ob}\mathcal{O}$, equipped with the units $1_X \in \mathsf{C}(X; X)$, $X \in \operatorname{Ob}\mathsf{C}$, and multiplications

$$\mu_\phi : \prod_{J\sqcup\mathbf{1}}\big[\big(\mathsf{C}\big((X_i)_{i\in\phi^{-1}(j)}; Y_j\big)\big)_{j\in J}, \mathsf{C}\big((Y_j)_{j\in J}; Z\big)\big] \to \mathsf{C}\big((X_i)_{i\in I}; Z\big),$$

given for order preserving (respectively arbitrary) map $\phi : I \to J$, together with maps $I \ni i \mapsto X_i \in \operatorname{Ob}\mathsf{C}$, $J \ni j \mapsto Y_j \in \operatorname{Ob}\mathsf{C}$, $1 \mapsto Z \in \operatorname{Ob}\mathsf{C}$. The multiplication is required to be associative. This is expressed by the equation of Fig. 1, written for each pair of composable maps $I \xrightarrow{\phi} J \xrightarrow{\psi} K$ from $\mathcal{O}$ (respectively each pair from $\mathcal{S}$), where $\phi_k = \phi|_{(\phi\psi)^{-1}(k)} : (\phi\psi)^{-1}(k) \to \psi^{-1}(k)$. Note that $\phi_k^{-1}(j) = \phi^{-1}(j)$ for any $j \in \psi^{-1}(k)$.

The units have to satisfy the axioms:

$$\big[\mathsf{C}\big((X_i)_{i\in I}; Z\big) \xrightarrow{\mathrm{id}\times 1_Z} \mathsf{C}\big((X_i)_{i\in I}; Z\big) \times \mathsf{C}(Z; Z) \xrightarrow{\mu_{I\to\mathbf{1}}} \mathsf{C}\big((X_i)_{i\in I}; Z\big)\big] = \mathrm{id},$$

$$\left[\mathsf{C}\big((X_i)_{i\in I}; Z\big) \xrightarrow{\prod_{I\sqcup\mathbf{1}}((1_{X_i})_{i\in I},\mathrm{id})} \prod_{I\sqcup\mathbf{1}}\big[\big(\mathsf{C}(X_i; X_i)\big)_{i\in I}, \mathsf{C}\big((X_i)_{i\in I}; Z\big)\big] \right.$$
$$\left. \xrightarrow{\mu_{\mathrm{id}_I}} \mathsf{C}\big((X_i)_{i\in I}; Z\big)\right] = \mathrm{id}.$$

Note that an operad is a multicategory with one object. One encounters various forms of multicategories, e.g. multilinear categories of Borcherds [4], pseudo-tensor categories of Beilinson and Drinfeld [1, Definition 1.1.1], substitudes of Day and Street [7].

Let C, D be symmetric (respectively plain) multicategories. A symmetric (respectively plain) *multifunctor* $F : \mathsf{C} \to \mathsf{D}$ is a mapping of objects $\operatorname{Ob}F : \operatorname{Ob}\mathsf{C} \to \operatorname{Ob}\mathsf{D}$, $X \mapsto FX$, together with maps

$$F_{(X_i)_{i\in I};Y} : \mathsf{C}\big((X_i)_{i\in I}; Y\big) \to \mathsf{D}\big((FX_i)_{i\in I}; FY\big),$$

given for each function $I \sqcup \mathbf{1} \to \operatorname{Ob}\mathsf{C}$, $i \mapsto X_i$, $1 \mapsto Y$, such that for each $X \in \operatorname{Ob}\mathsf{C}$

$$1_X^{\mathsf{C}} F_{X;X} = 1_{FX}^{\mathsf{D}},$$

and for each (respectively order preserving) map $\phi : I \to J$ together with a map $I \sqcup J \sqcup \mathbf{1} \to \operatorname{Ob}\mathsf{C}$, $i \mapsto X_i$, $j \mapsto Y_j$, $1 \mapsto Z$, we have

$$\begin{tikzcd}
\prod_{J\sqcup\mathbf{1}}\big[\big(\mathsf{C}\big((X_i)_{i\in\phi^{-1}j}; Y_j\big)\big)_{j\in J}, \mathsf{C}\big((Y_j)_{j\in J}; Z\big)\big] \arrow[r, "\mu_\phi^{\mathsf{C}}"] \arrow[d, "\prod_{J\sqcup\mathbf{1}}[(F_{(X_i)_{i\in\phi^{-1}j};Y_j})_{j\in J}, F_{(Y_j)_{j\in J};Z}]"'] \arrow[dr, phantom, "="] & \mathsf{C}\big((X_i)_{i\in I}; Z\big) \arrow[d, "F_{(X_i)_{i\in I};Z}"] \\
\prod_{J\sqcup\mathbf{1}}\big[\big(\mathsf{D}\big((FX_i)_{i\in\phi^{-1}j}; FY_j\big)\big)_{j\in J}, \mathsf{D}\big((FY_j)_{j\in J}; FZ\big)\big] \arrow[r, "\mu_\phi^{\mathsf{D}}"] & \mathsf{D}\big((FX_i)_{i\in I}; FZ\big)
\end{tikzcd}$$

$$\begin{tikzcd}
\prod_{J\sqcup\mathbf{1}}\Big[\big(\mathsf{C}((X_i)_{i\in\phi^{-1}j};Y_j)\big)_{j\in J},\prod_{K\sqcup\mathbf{1}}\Big(\big(\mathsf{C}((Y_j)_{j\in\psi^{-1}k};Z_k)\big)_{k\in K},\mathsf{C}((Z_k)_{k\in K};W)\Big)\Big] \arrow[r, "\prod_{J\sqcup\mathbf{1}}((1)_{j\in J},\mu_\psi)"] \arrow[d, "\wr"'] & \prod_{J\sqcup\mathbf{1}}\Big[\big(\mathsf{C}((X_i)_{i\in\phi^{-1}j};Y_j)\big)_{j\in J},\mathsf{C}((Y_j)_{j\in J};W)\Big] \arrow[d, "\mu_\phi"] \\
\prod_{K\sqcup\mathbf{1}}\Big[\Big(\prod_{\psi^{-1}k\sqcup\mathbf{1}}\big[\big(\mathsf{C}((X_i)_{i\in\phi_k^{-1}j};Y_j)\big)_{j\in\psi^{-1}k},\mathsf{C}((Y_j)_{j\in\psi^{-1}k};Z_k)\big]\Big)_{k\in K},\mathsf{C}((Z_k)_{k\in K};W)\Big] \arrow[d, "\prod_{K\sqcup\mathbf{1}}((\mu_{\phi_k})_{k\in K},1)"'] & \mathsf{C}((X_i)_{i\in I};W) \\
\prod_{K\sqcup\mathbf{1}}\big[\big(\mathsf{C}((X_i)_{i\in(\phi\psi)^{-1}(k)};Z_k)\big)_{k\in K},\mathsf{C}((Z_k)_{k\in K};W)\big] \arrow[ur, "\mu_{\phi\psi}"] &
\end{tikzcd}$$

Fig. 1. Associativity in multicategories.

A *multinatural transformation* of multifunctors $r : F \to G : \mathsf{C} \to \mathsf{D}$ is a collection of elements $r_X \in \mathsf{D}(FX; GX)$, $X \in \operatorname{Ob}\mathsf{C}$, such that

$$\begin{aligned}
&\Big[\mathsf{C}\big((X_i)_{i\in I}; Y\big) \xrightarrow{F_{(X_i)_{i\in I};Y} \times r_Y} \mathsf{D}\big((FX_i)_{i\in I}; FY\big) \times \mathsf{D}(FY; GY) \\
&\quad \xrightarrow{\mu^{\mathsf{D}}_{I\to\mathbf{1}}} \mathsf{D}\big((FX_i)_{i\in I}; GY\big)\Big] \\
&\quad = \left[\mathsf{C}\big((X_i)_{i\in I}; Y\big) \xrightarrow{\prod_{I\sqcup\mathbf{1}}[(r_{X_i})_{i\in I}, G_{(X_i)_{i\in I};Y}]} \right. \\
&\qquad \left. \prod_{I\sqcup\mathbf{1}} \Big[\big(\mathsf{D}(FX_i; GX_i)\big)_{i\in I}, \mathsf{D}\big((GX_i)_{i\in I}; GY\big)\Big] \xrightarrow{\mu^{\mathsf{D}}_{\mathrm{id}_I}} \mathsf{D}\big((FX_i)_{i\in I}; GY\big)\right]
\end{aligned}$$

for any function $I\sqcup\mathbf{1} \to \operatorname{Ob}\mathsf{C}$, $i \mapsto X_i$, $1 \mapsto Y$. A *natural transformation* of multifunctors $r : F \to G : \mathsf{C} \to \mathsf{D}$ is a natural transformation of the underlying functors. It satisfies the above equation for one-element sets I.

Plain (respectively symmetric) multicategories, multifunctors and their multinatural transformations form a 2-category $\mathcal{MC}$atm (respectively $\mathcal{SMC}$atm).

3.6. *Monoidal categories are examples of multicategories*

A plain (respectively symmetric) lax Monoidal category $\mathcal{C}$ gives rise to a plain (respectively symmetric) multicategory $\widehat{\mathcal{C}}$ with

- class of objects $\operatorname{Ob}\widehat{\mathcal{C}} = \operatorname{Ob}\mathcal{C}$,
- sets of morphisms $\widehat{\mathcal{C}}((X_i)_{i\in I}; Y) = \mathcal{C}(\otimes^{i\in I} X_i, Y)$,
- units $1^{\widehat{\mathcal{C}}}_X = 1^{\mathcal{C}}_X \in \mathcal{C}(X, X)$,
- multiplication morphisms for each map $f : I \to J$ from $\mathcal{O}$ (respectively $\mathcal{S}$)

$$\begin{aligned}
\mu_f : &\left[\prod_{j\in J} \widehat{\mathcal{C}}\big((X_i)_{i\in f^{-1}j}; Y_j\big)\right] \times \widehat{\mathcal{C}}\big((Y_j)_{j\in J}; Z\big) \\
&\xrightarrow{\otimes^J \times 1} \mathcal{C}\big(\otimes^{j\in J} \otimes^{i\in f^{-1}j} X_i, \otimes^{j\in J} Y_j\big) \times \mathcal{C}\big(\otimes^{j\in J} Y_j, Z\big) \\
&\xrightarrow{\lambda^f \cdot - \cdot -} \mathcal{C}\big(\otimes^{i\in I} X_i, Z\big) = \widehat{\mathcal{C}}\big((X_i)_{i\in I}; Z\big).
\end{aligned}$$

Any lax (symmetric) Monoidal functor $(F, \phi^I) : (\mathcal{C}, \otimes^I, \lambda^f_{\mathcal{C}}) \to (\mathcal{D}, \otimes^I, \lambda^f_{\mathcal{D}})$ between lax (symmetric) Monoidal categories gives rise to a (symmetric) multifunctor $\widehat{F} : \widehat{\mathcal{C}} \to \widehat{\mathcal{D}}$ with

- mapping of objects $\operatorname{Ob}\widehat{F} = \operatorname{Ob} F$,
- mapping of sets of morphisms

$$\begin{aligned}
\widehat{F}_{(X_i);Y} = &\left[\widehat{\mathcal{C}}\big((X_i)_{i\in I}; Y\big) = \mathcal{C}\big(\otimes^{i\in I} X_i, Y\big) \xrightarrow{F_{\otimes^{i\in I} X_i, Y}} \mathcal{D}\big(F\big(\otimes^{i\in I} X_i\big), FY\big)\right. \\
&\left.\xrightarrow{\mathcal{D}(\phi^I, FY)} \mathcal{D}\big(\otimes^{i\in I}(FX_i), FY\big) = \widehat{\mathcal{D}}\big((FX_i)_{i\in I}; FY\big)\right].
\end{aligned}$$

A lax Monoidal transformation $r:(F,\phi^I)\to(G,\psi^I):\mathcal{C}\to\mathcal{D}$ gives rise to a multi-natural transformation $\widehat{r}:\widehat{F}\to\widehat{G}:\widehat{\mathcal{C}}\to\widehat{\mathcal{D}}$, determined by the morphisms $\widehat{r}_X=r_X\in\mathcal{D}(FX,GX)$.

Thus, the 2-categories lax-Mono-cat (respectively lax-sym-Mono-cat) of lax (symmetric) Monoidal categories and the 2-categories $\mathcal{MC}$atm (respectively $\mathcal{SMC}$atm) of (symmetric) multicategories are related by the 2-functor $\mathcal{C}\mapsto\widehat{\mathcal{C}}$, $F\mapsto\widehat{F}$, $r\mapsto\widehat{r}$. Note that 2-categories and 2-functors are the same things as $\mathcal{C}$at-categories and $\mathcal{C}$at-functors. For any symmetric Monoidal category $\mathcal{V}$, in particular for $\mathcal{V}=\mathcal{C}$at, one can define lax symmetric Monoidal $\mathcal{V}$-categories.

3.7. Theorem. *(See [3].) The assignment $\mathcal{C}\mapsto\widehat{\mathcal{C}}$, $F\mapsto\widehat{F}$, $r\mapsto\widehat{r}$ is a symmetric Monoidal $\mathcal{C}$at-functor lax-Mono-cat $\to$ $\mathcal{MC}$atm, lax-sym-Mono-cat $\to$ $\mathcal{SMC}$atm. For arbitrary lax (symmetric) Monoidal categories $\mathcal{C}$, $\mathcal{D}$ the maps lax-(sym-)Mono$(\mathcal{C},\mathcal{D})\to(\mathcal{S})\mathcal{MC}atm(\widehat{\mathcal{C}},\widehat{\mathcal{D}})$, $F\mapsto\widehat{F}$, $r\mapsto\widehat{r}$ are bijective.*

3.8. *Multicategories enriched in multicategories*

Let V be a symmetric multicategory. A plain (respectively symmetric) multicategory D *enriched in* V, or a V-multicategory, is a set $\operatorname{Ob}\mathsf{D}$ of objects, equipped with objects of multimorphisms $\mathsf{D}((X_i)_{i\in I};Y)\in\operatorname{Ob}\mathsf{V}$, assigned to each map $I\sqcup\{*\}\to\operatorname{Ob}\mathsf{D}$, $i\mapsto X_i$, $*\mapsto Y$, where $I\in\operatorname{Ob}\mathcal{O}$, equipped with units $1^{\mathsf{D}}_X\in\mathsf{V}(;\mathsf{D}(X;X))$ for each $X\in\operatorname{Ob}\mathsf{D}$ and equipped with multiplications

$$\mu_\phi\in\mathsf{V}\bigl(\bigl(\mathsf{D}\bigl((X_i)_{i\in\phi^{-1}(j)};Y_j\bigr)\bigr)_{j\in J},\mathsf{D}\bigl((Y_j)_{j\in J};Z\bigr);\mathsf{D}\bigl((X_i)_{i\in I};Z\bigr)\bigr),$$

for a non-decreasing (respectively arbitrary) map $\phi:I\to J$ together with maps $I\ni i\mapsto X_i\in\operatorname{Ob}\mathsf{D}$, $J\ni j\mapsto Y_j\in\operatorname{Ob}\mathsf{D}$, $*\mapsto Z\in\operatorname{Ob}\mathsf{D}$. The multiplication is required to be associative. This is expressed by the following equation in V, written for each pair of composable maps $I\xrightarrow{\phi}J\xrightarrow{\psi}K$:

$$\begin{tikzcd}
\begin{array}{c}\bigl(\mathsf{D}\bigl((X_i)_{i\in\phi^{-1}j};Y_j\bigr)\bigr)_{j\in J},\\ \bigl(\mathsf{D}\bigl((Y_j)_{j\in\psi^{-1}k};Z_k\bigr)\bigr)_{k\in K},\\ \mathsf{D}\bigl((Z_k)_{k\in K};W\bigr)\end{array} \arrow[r, "(1)_{j\in J},\mu^{\mathsf{D}}_\psi"] \arrow[d, "\text{permutation}"'] & \begin{array}{c}\bigl(\mathsf{D}\bigl((X_i)_{i\in\phi^{-1}j};Y_j\bigr)\bigr)_{j\in J},\\ \mathsf{D}\bigl((Y_j)_{j\in J};W\bigr)\end{array} \arrow[dd, "\mu^{\mathsf{D}}_\phi"] \\
\begin{array}{c}\bigl(\bigl(\mathsf{D}\bigl((X_i)_{i\in\phi^{-1}j};Y_j\bigr)\bigr)_{j\in\psi^{-1}k},\mathsf{D}\bigl((Y_j)_{j\in\psi^{-1}k};Z_k\bigr)\bigr)_{k\in K},\\ \mathsf{D}\bigl((Z_k)_{k\in K};W\bigr)\end{array} \arrow[d, "(\mu^{\mathsf{D}}_{\phi_k})_{k\in K},1"'] & \\
\begin{array}{c}\bigl(\mathsf{D}\bigl((X_i)_{i\in(\phi\psi)^{-1}k};Z_k\bigr)\bigr)_{k\in K},\\ \mathsf{D}\bigl((Z_k)_{k\in K};W\bigr)\end{array} \arrow[r, "\mu^{\mathsf{D}}_{\phi\psi}"] & \mathsf{D}\bigl((X_i)_{i\in I};W\bigr)
\end{tikzcd}$$

where $\phi_k = \phi|_{(\phi\psi)^{-1}(k)} : (\phi\psi)^{-1}(k) \to \psi^{-1}(k)$. The above diagram means that the following equation holds:

$$\begin{aligned}&\mu^{\mathsf V}_{\alpha:\,J\sqcup K\sqcup\mathbf 1\to J\sqcup\mathbf 1}\bigl(\bigl(1_{\mathsf D((X_i)_{i\in\phi^{-1}j};Y_j)}\bigr)_{j\in J},\mu^{\mathsf D}_\psi,\mu^{\mathsf D}_\phi\bigr)\\ &\quad= \mu^{\mathsf V}_{\beta:\,J\sqcup K\sqcup\mathbf 1\to K\sqcup\mathbf 1}\bigl(\bigl(\mu^{\mathsf D}_{\phi_k}\bigr)_{k\in K},1_{\mathsf D((Z_k)_{k\in K};W)},\mu^{\mathsf D}_{\phi\psi}\bigr),\end{aligned}$$

where $\alpha = \mathrm{id}_J \sqcup \rhd : J \sqcup K \sqcup \mathbf 1 \to J \sqcup \mathbf 1$, and $\beta : J \sqcup K \sqcup \mathbf 1 \to K \sqcup \mathbf 1$ is determined by its restrictions $\beta|_{K\sqcup\mathbf 1} = \mathrm{id} : K \sqcup \mathbf 1 \to K \sqcup \mathbf 1$, $\beta|_J = (J \xrightarrow{\psi} K \hookrightarrow K \sqcup \mathbf 1)$.

The units have to satisfy the equations

$$\begin{aligned}&\mu^{\mathsf V}_{\mathbf 1\hookrightarrow\mathbf 2}\bigl(1^{\mathsf V}_{\mathsf D((X_i)_{i\in I};Z)},1^{\mathsf D}_Z,\mu^{\mathsf D}_\phi\bigr) = 1^{\mathsf V}_{\mathsf D((X_i)_{i\in I};Z)},\\ &\mu^{\mathsf V}_{\mathbf 1\hookrightarrow I\sqcup\mathbf 1}\bigl(\bigl(1^{\mathsf D}_{X_i}\bigr)_{i\in I},1^{\mathsf V}_{\mathsf D((X_i)_{i\in I};Z)},\mu^{\mathsf D}_{\mathrm{id}_I}\bigr) = 1^{\mathsf V}_{\mathsf D((X_i)_{i\in I};Z)}.\end{aligned}$$

Let $\mathsf C$, $\mathsf D$ be $\mathsf V$-multicategories. A $\mathsf V$-*multifunctor* $F : \mathsf C \to \mathsf D$ is a mapping of objects $\mathrm{Ob}\, F : \mathrm{Ob}\,\mathsf C \to \mathrm{Ob}\,\mathsf D$, $X \mapsto FX$, together with morphisms

$$F_{(X_i)_{i\in I};Y} \in \mathsf V\bigl(\mathsf C\bigl((X_i)_{i\in I};Y\bigr);\mathsf D\bigl((FX_i)_{i\in I};FY\bigr)\bigr),$$

given for each collection $(X_i)_{i\in I}$, Y of objects of $\mathsf C$, that preserve the units and multiplications:

$$\begin{tikzcd}\bigl(\mathsf C\bigl((X_i)_{i\in\phi^{-1}j};Y_j\bigr)\bigr)_{j\in J},\mathsf C((Y_j)_{j\in J};Z) \arrow[r, "\mu^{\mathsf C}_\phi"] \arrow[d, "(F_{(X_i)_{i\in\phi^{-1}j};Y_j})_{j\in J},F_{(Y_j)_{j\in J};Z}"'] \arrow[dr, phantom, "="] & \mathsf C\bigl((X_i)_{i\in I};Z\bigr) \arrow[d, "F_{(X_i)_{i\in I};Z}"] \\ \bigl(\mathsf D\bigl((FX_i)_{i\in\phi^{-1}j};FY_j\bigr)\bigr)_{j\in J},\mathsf D\bigl((FY_j)_{j\in J};FZ\bigr) \arrow[r, "\mu^{\mathsf D}_\phi"] & \mathsf D\bigl((FX_i)_{i\in I};FZ\bigr)\end{tikzcd}$$

Here compositions are taken in $\mathsf V$. A $\mathsf V$-*functor* $F : \mathsf C \to \mathsf D$ requires such data only for $I = \mathbf 1$.

A (*multi*)*natural* $\mathsf V$-*transformation* $r : F \to G : \mathsf C \to \mathsf D$ is a family of elements $r_X \in \mathsf V(;\mathsf D(FX;GX))$, $X \in \mathrm{Ob}\,\mathsf C$, that intertwines the actions of F and G on (multi)morphisms:

$$\begin{tikzcd}\mathsf C\bigl((X_i)_{i\in I};Y\bigr) \arrow[r, "F_{(X_i)_{i\in I};Y},r_Y"] \arrow[d, "(r_{X_i})_{i\in I},G_{(X_i)_{i\in I};Y}"'] \arrow[dr, phantom, "="] & \mathsf D\bigl((FX_i)_{i\in I};FY\bigr),\mathsf D(FY;GY) \arrow[d, "\mu^{\mathsf D}_{I\to\mathbf 1}"] \\ \bigl(\mathsf D(FX_i;GX_i)\bigr)_{i\in I},\mathsf D\bigl((GX_i)_{i\in I};GY\bigr) \arrow[r, "\mu^{\mathsf D}_{\mathrm{id}_I}"] & \mathsf D\bigl((FX_i)_{i\in I};GY\bigr)\end{tikzcd}$$

The compositions are taken in $\mathsf V$. In the natural $\mathsf V$-transformation case this equation is required only for $I = \mathbf 1$.

3.9. *Closed multicategories*

Closed monoidal categories are generalized in this section to the multicategory setting.

3.10. DEFINITION. A plain multicategory $\mathsf C$ is *closed* if for any collection $((X_i)_{i\in I}, Z)$, $I \in \mathrm{Ob}\,\mathcal S$, of objects of $\mathsf C$ there is an object $\underline{\mathsf C}((X_i)_{i\in I};Z)$ of $\mathsf C$ and an evaluation element

$$\mathrm{ev}^{\mathsf C}_{(X_i)_{i\in I};Z} \in \mathsf C\bigl((X_i)_{i\in I},\underline{\mathsf C}\bigl((X_i)_{i\in I};Z\bigr);Z\bigr)$$

such that the composition (with $\iota = (\mathbf{1} = \varnothing \sqcup \mathbf{1} \sqcup \varnothing \xrightarrow{\triangleleft\sqcup\mathrm{id}\sqcup\triangleleft} I \sqcup \mathbf{1} \sqcup \mathbf{1} = I \sqcup \mathbf{2})$)

$$\begin{aligned}
&\varphi_{(Y_j)_{j\in J};(X_i)_{i\in I};Z} \\
&\quad = \Big[\mathsf{C}\big((Y_j)_{j\in J}; \underline{\mathsf{C}}\big((X_i)_{i\in I}; Z\big)\big) \xrightarrow{\prod_{i\in I} 1^{\mathsf{C}}_{X_i} \times \mathrm{id} \times \mathrm{ev}^{\mathsf{C}}_{(X_i)_{i\in I};Z}} \prod_{i\in I} \mathsf{C}(X_i; X_i) \\
&\qquad \times \mathsf{C}\big((Y_j)_{j\in J}; \underline{\mathsf{C}}\big((X_i)_{i\in I}; Z\big)\big) \times \mathsf{C}\big((X_i)_{i\in I}, \underline{\mathsf{C}}\big((X_i)_{i\in I}; Z\big); Z\big) \\
&\qquad \xrightarrow{\mu^{\mathsf{C}}_{\mathrm{id}\sqcup\triangleright:\, I\sqcup J\to I\sqcup\mathbf{1}}} \mathsf{C}\big((X_i)_{i\in I}, (Y_j)_{j\in J}; Z\big)\Big]
\end{aligned}$$

is a bijection for an arbitrary sequence $(Y_j)_{j\in J}$, $J \in \operatorname{Ob}\mathcal{S}$, of objects of C.

Here $\triangleleft : \varnothing \to K$ and $\triangleright : K \to \mathbf{1}$ for an arbitrary set are of course the only maps that exist. Concatenation of sequences indexed by I and J is indexed by the disjoint union $I \sqcup J$, where $i < j$ for all $i \in I$, $j \in J$.

Note that for $I = \varnothing$ an object $\underline{\mathsf{C}}(;Z)$ and an element $\mathrm{ev}^{\mathsf{C}}_{;Z}$ with the required property always exist. Namely, we shall always take $\underline{\mathsf{C}}(;Z) = Z$ and $\mathrm{ev}^{\mathsf{C}}_{;Z} = 1_Z : Z \to Z$. With this choice $\varphi_{(Y_j)_{j\in J};Z} = \mathrm{id} : \mathsf{C}((Y_j)_{j\in J}; Z) \to \mathsf{C}((Y_j)_{j\in J}; Z)$ is the identity map.

Let C be a symmetric closed multicategory. Let $(X_i)_{i\in I}$, $(Y_j)_{j\in J}$, Z be objects of C and let $\phi : I \to J$ be an arbitrary map. Define the multiplication morphism

$$\mu^{\underline{\mathsf{C}}}_{\phi} \in \mathsf{C}\Big(\big(\underline{\mathsf{C}}\big((X_i)_{i\in\phi^{-1}j}; Y_j\big)\big)_{j\in J}, \underline{\mathsf{C}}\big((Y_j)_{j\in J}; Z\big); \underline{\mathsf{C}}\big((X_i)_{i\in I}; Z\big)\Big)$$

as a unique solution of the following equation in C:

$$\begin{tikzcd}
(X_i)_{i\in I}, \big(\underline{\mathsf{C}}\big((X_i)_{i\in\phi^{-1}j}; Y_j\big)\big)_{j\in J}, \underline{\mathsf{C}}\big((Y_j)_{j\in J}; Z\big) \arrow[r, "(\mathrm{ev}^{\mathsf{C}})_{J,1}"] \arrow[d, "(1)_I{,}\mu^{\underline{\mathsf{C}}}_{\phi}"'] \arrow[dr, phantom, "="] & (Y_j)_{j\in J}, \underline{\mathsf{C}}\big((Y_j)_{j\in J}; Z\big) \arrow[d, "\mathrm{ev}^{\mathsf{C}}"] \\
(X_i)_{i\in I}, \underline{\mathsf{C}}\big((X_i)_{i\in I}; Z\big) \arrow[r, "\mathrm{ev}^{\mathsf{C}}"] & Z
\end{tikzcd}$$

The composition in C of the left-bottom path is $\mu^{\mathsf{C}}_{\mathrm{id}\sqcup\triangleright:\, I\sqcup J\sqcup\mathbf{1}\to I\sqcup\mathbf{1}}$. The composition of the top-right path is $\mu^{\mathsf{C}}_{\beta:\, I\sqcup J\sqcup\mathbf{1}\to J\sqcup\mathbf{1}}$, where $\beta|_{J\sqcup\mathbf{1}} = \mathrm{id}$, $\beta|_I = (I \xrightarrow{\phi} J \hookrightarrow J \sqcup \mathbf{1})$.

Define for each object $X \in \operatorname{Ob}\mathsf{C}$ the unit $1^{\underline{\mathsf{C}}}_X \in \mathsf{C}(;\underline{\mathsf{C}}(X;X))$ as the unique element whose image under the bijection $\varphi_{;X;X} : \mathsf{C}(;\underline{\mathsf{C}}(X;X)) \to \mathsf{C}(X;X)$ equals 1^{C}_X.

3.11. Proposition. *(See [3].) Let* C *be a closed symmetric multicategory. The elements* $\mu^{\underline{\mathsf{C}}}_{\phi}$, $\phi : I \to J$, *and* $1^{\underline{\mathsf{C}}}_X$, $X \in \operatorname{Ob}\mathsf{C}$, *make the* C*-multiquiver* $\underline{\mathsf{C}}$, $\operatorname{Ob}\underline{\mathsf{C}} = \operatorname{Ob}\mathsf{C}$, *into a symmetric* C*-multicategory.*

A closed Monoidal category provides an example of closed multicategory. More precisely, let $\mathcal{C}$ be a symmetric closed Monoidal category. Then $\widehat{\mathcal{C}}$ is closed with inner homomorphism objects given by $\underline{\widehat{\mathcal{C}}}((X_i)_{i\in I}; Z) = \underline{\mathcal{C}}(\otimes^{i\in I} X_i, Z)$ and with evaluations represented by compositions in $\mathcal{C}$

$$\mathrm{ev}^{\widehat{\mathcal{C}}} = \big[\otimes^{I\sqcup\mathbf{1}}\big((X_i)_{i\in I}, \underline{\mathcal{C}}\big(\otimes^{i\in I} X_i, Z\big)\big)$$

$$\xrightarrow{\lambda_{\mathcal{C}}^{\triangleright\sqcup\mathrm{id}:\,I\sqcup\mathbf{1}\to\mathbf{2}}} \bigl(\otimes^{i\in I}X_i\bigr)\otimes\underline{\mathcal{C}}\bigl(\otimes^{i\in I}X_i,Z\bigr)\xrightarrow{\mathrm{ev}^{\mathcal{C}}} Z\bigr].$$

Another example is the multicategory of free coalgebras for a multicomonad. Let C be a closed multicategory, (T,Δ,ε) a multicomonad in C. This means that $T:\mathsf{C}\to\mathsf{C}$ is a multifunctor, $\Delta:T\to TT:\mathsf{C}\to\mathsf{C}$ and $\varepsilon:T\to\mathrm{Id}:\mathsf{C}\to\mathsf{C}$ are multinatural transformations, and the triple (T,Δ,ε) is a coalgebra in the strict monoidal category $\mathcal{MC}\mathrm{atm}(\mathsf{C},\mathsf{C})$. Let C_T^f denote the multicategory of free T-coalgebras. Its set of objects coincides with $\operatorname{Ob}\mathsf{C}$ and for each collection $((A_i)_{i\in I},B)$, $I\in\operatorname{Ob}\mathcal{O}$, of objects the set $\mathsf{C}_T^f((A_i)_{i\in I};B)$ consists of T-coalgebra morphisms $f:(TA_i)_{i\in I}\to TB$. The latter are morphisms $f\in\mathsf{C}((TA_i)_{i\in I};TB)$ that satisfy the following equation in C:

$$\bigl((TA_i)_{i\in I}\xrightarrow{f}TB\xrightarrow{\Delta}TTB\bigr)=\bigl((TA_i)_{i\in I}\xrightarrow{(\Delta)_{i\in I}}(TTA_i)_{i\in I}\xrightarrow{Tf}TTB\bigr).$$

Given T-coalgebras X_i, $i\in I$, and an object B of C denote by

$$\Theta_{(X_i);B}:T\underline{\mathsf{C}}\bigl((X_i)_{i\in I};B\bigr)\to\underline{\mathsf{C}}\bigl((X_i)_{i\in I};TB\bigr)$$

the unique morphism that satisfies the equation

$$\begin{tikzcd}
(X_i)_{i\in I},T\underline{\mathsf{C}}\bigl((X_i)_{i\in I};B\bigr) \arrow[r, "(\delta)_I{,}1"] \arrow[d, "(1)_I{,}\Theta_{(X_i);B}"'] \arrow[dr, phantom, "="] & (TX_i)_{i\in I},T\underline{\mathsf{C}}\bigl((X_i)_{i\in I};B\bigr) \arrow[d, "T\mathrm{ev}^{\mathsf{C}}_{(X_i);B}"] \\
(X_i)_{i\in I},\underline{\mathsf{C}}\bigl((X_i)_{i\in I};TB\bigr) \arrow[r, "\mathrm{ev}^{\mathsf{C}}_{(X_i)_{i\in I};TB}"] & TB
\end{tikzcd} \tag{3.11.1}$$

Its existence and uniqueness follows from closedness of C.

3.12. PROPOSITION. *(See* [3]*.) The multicategory C_T^f of free T-coalgebras is closed. The inner homomorphisms objects can be chosen as $\underline{\mathsf{C}_T^f}((A_i)_{i\in I};B)=\underline{\mathsf{C}}((TA_i)_{i\in I};B)$ and the evaluations as*

$$\begin{aligned}\mathrm{ev}^{\mathsf{C}_T^f}_{(A_i)_{i\in I};B}=\bigl[&(TA_i)_{i\in I},T\underline{\mathsf{C}}\bigl((TA_i)_{i\in I};B\bigr)\\ &\xrightarrow{(1)_I,\Theta_{(TA_i);B}}(TA_i)_{i\in I},\underline{\mathsf{C}}\bigl((TA_i)_{i\in I};TB\bigr)\xrightarrow{\mathrm{ev}^{\mathsf{C}}}TB\bigr].\end{aligned}$$

Closedness of the multicategory of free $T^{\geqslant1}$-coalgebras is the main ingredient in the proof of closedness of the multicategory of A_∞-categories.

3.13. *Closedness of Monoidal categories of quivers*

The symmetric Monoidal category $\mathscr{Q}_p$ is closed. In fact, define the quiver $\underline{\mathscr{Q}_p}(\mathcal{A},\mathcal{B})$ as follows. The objects of $\underline{\mathscr{Q}_p}(\mathcal{A},\mathcal{B})$ are maps $f:\operatorname{Ob}\mathcal{A}\to\operatorname{Ob}\mathcal{B}$, and the graded component $\underline{\mathscr{Q}_p}(\mathcal{A},\mathcal{B})(f,g)^d$ of the $\Bbbk$-module of morphisms consists of span morphisms $r:\mathcal{A}\to\mathcal{B}$ with $\operatorname{Ob}_s r=f$, $\operatorname{Ob}_t r=g$ of degree $\deg r=d$, that is,

$$\underline{\mathscr{Q}_p}(\mathcal{A},\mathcal{B})(f,g)=\prod_{X,Y\in\operatorname{Ob}\mathcal{A}}\underline{\mathbf{gr}}\bigl(\mathcal{A}(X,Y),\mathcal{B}(Xf,Yg)\bigr)\in\operatorname{Ob}\mathbf{gr}$$

where $\mathbf{gr}$ is the closed Monoidal category of graded $\Bbbk$-modules, and $\underline{\mathbf{gr}}$ denotes the corresponding category with inner homomorphisms objects,

$$\underline{\mathbf{gr}}(P, Q)^n = \prod_{m \in \mathbb{Z}} \operatorname{Hom}_{\Bbbk}\bigl(P^m, Q^{m+n}\bigr).$$

Evaluations are given by

$$\operatorname{ev}^{\mathscr{Q}_p} : \mathcal{A} \boxtimes \underline{\mathscr{Q}_p}(\mathcal{A}, \mathcal{B}) \to \mathcal{B}, \quad (X, f) \mapsto Xf,$$

$$\operatorname{ev}^{\mathscr{Q}_p} = \Bigl[\mathcal{A}(X, Y) \otimes \underline{\mathscr{Q}_p}(\mathcal{A}, \mathcal{B})(f, g) \xrightarrow{1 \otimes \mathrm{pr}_{(X,Y)}} \mathcal{A}(X, Y) \otimes \underline{\mathbf{gr}}\bigl(\mathcal{A}(X, Y), \mathcal{B}(Xf, Yg)\bigr) \xrightarrow{\operatorname{ev}^{\mathbf{gr}}} \mathcal{B}(Xf, Yg)\Bigr], \quad a \otimes r \mapsto a.r_{X,Y}.$$

3.14. Proposition. *(See [3].) The symmetric Monoidal category $\mathscr{Q}_u = (\mathscr{Q}, \boxtimes_u^I, \lambda_u^f)$ of quivers is closed with $\underline{\mathscr{Q}_u}(\mathcal{A}, \mathcal{B})$ being a quiver, whose objects are morphisms of quivers $\mathcal{A} \to \mathcal{B}$ and the object of morphisms between $f, g : \mathcal{A} \to \mathcal{B}$ is given by*

$$\underline{\mathscr{Q}_u}(\mathcal{A}, \mathcal{B})(f, g) = \underline{\mathscr{Q}_p}\bigl(T^{\leqslant 1}\mathcal{A}, \mathcal{B}\bigr)(\operatorname{Ob} f, \operatorname{Ob} g).$$

Denote by $\Upsilon : \underline{\mathscr{Q}_u}(\mathcal{A}, \mathcal{B}) \to \underline{\mathscr{Q}_p}(T^{\leqslant 1}\mathcal{A}, \mathcal{B})$ the quiver morphism which gives the map $f \mapsto \operatorname{Ob} f$ on objects and identity map on morphisms. The evaluation morphism $\operatorname{ev}^{\mathscr{Q}_u} : \mathcal{A} \boxtimes_u \underline{\mathscr{Q}_u}(\mathcal{A}, \mathcal{B}) \to \mathcal{B}$ acts on objects as $\operatorname{ev}^{\mathscr{Q}_u}(X, f) = Xf$, and on morphisms via

$$\operatorname{ev}' = \bigl[T^{\leqslant 1}\mathcal{A} \boxtimes \underline{\mathscr{Q}_u}(\mathcal{A}, \mathcal{B}) \xrightarrow{1 \boxtimes \Upsilon} T^{\leqslant 1}\mathcal{A} \boxtimes \underline{\mathscr{Q}_p}(T^{\leqslant 1}\mathcal{A}, \mathcal{B}) \xrightarrow{\operatorname{ev}^{\mathscr{Q}_p}} \mathcal{B}\bigr],$$

$$\operatorname{ev}'' : \mathcal{A} \boxtimes T^0\underline{\mathscr{Q}_u}(\mathcal{A}, \mathcal{B}) \to \mathcal{B},$$

$$\operatorname{ev}'' = \bigl[\mathcal{A}(X, Y) \otimes T^0\underline{\mathscr{Q}_u}(\mathcal{A}, \mathcal{B})(f, f) \xrightarrow[\sim]{(\lambda^{I\cdot})^{-1}} \mathcal{A}(X, Y) \xrightarrow{f_{X,Y}} \mathcal{B}(Xf, Yf)\bigr].$$

For an arbitrary quiver $\mathcal{C} \in \operatorname{Ob} \mathscr{Q}$ and a positive integer M introduce the morphism of quivers

$$\nu^{\leqslant M} : T\mathcal{C} \to TT^{\leqslant 1}\mathcal{C},$$

$$\nu_{kn}^{\leqslant M} : T^k\mathcal{C} \to \oplus_{S \subset \mathbf{n}} \otimes^{p \in \mathbf{n}} T^{\chi(p \in S)}\mathcal{C} = T^n T^{\leqslant 1}\mathcal{C},$$

as follows. By definition, the matrix element $\nu_{kn}^{\leqslant M}$ vanishes, if $n > k + M$. Its summand corresponding to a subset $S \subset \mathbf{n}$ vanishes unless $|S| = k$. If $n \leqslant k + M$ and $|S| = k$, the summand is defined as $\lambda^{g : \mathbf{k} \hookrightarrow \mathbf{n}} : T^k\mathcal{C} \xrightarrow{\sim} \otimes^{p \in \mathbf{n}} T^{\chi(p \in S)}\mathcal{C}$, where the image of the increasing embedding $g : \mathbf{k} \hookrightarrow \mathbf{n}$ is S. Thus, $\nu^{\leqslant M} : T \to TT^{\leqslant 1}$ is a natural transformation. We shall use it in formulas, where dependence on M is irrelevant, provided the truncation parameter M is large enough.

The quiver map

$$\begin{aligned}\theta_{(\mathcal{A}_i);\mathcal{B}} \overset{\text{def}}{=} \Theta_{(T^{\geqslant 1}s\mathcal{A}_i);s\mathcal{B}} : T^{\geqslant 1}\underline{\mathscr{Q}_u}\big(\boxtimes_u^{i\in I} T^{\geqslant 1}s\mathcal{A}_i, s\mathcal{B}\big)\\ \to \underline{\mathscr{Q}_u}\big(\boxtimes_u^{i\in I} T^{\geqslant 1}s\mathcal{A}_i, T^{\geqslant 1}s\mathcal{B}\big),\end{aligned}$$

defined by (3.11.1), assigns to an object (a morphism of quivers) $f : \boxtimes_u^{i\in I} T^{\geqslant 1}s\mathcal{A}_i \to s\mathcal{B}$, the object (a morphism of quivers)

$$\begin{aligned}\big[\boxtimes_u^{i\in I} T^{\geqslant 1}s\mathcal{A}_i \xrightarrow{\boxtimes_u^I \Delta} \boxtimes_u^{i\in I} T^{\geqslant 1}T^{\geqslant 1}s\mathcal{A}_i \xrightarrow{\tau^I} T^{\geqslant 1} \boxtimes_u^{i\in I} T^{\geqslant 1}s\mathcal{A}_i\\ \xrightarrow{T^{\geqslant 1}f} T^{\geqslant 1}s\mathcal{B}\big] = \widehat{f}.\end{aligned}$$

Starting from elements $r^t \in \underline{\mathscr{Q}_p}(\boxtimes^{i\in I} Ts\mathcal{A}_i, s\mathcal{B})(f^{t-1}, f^t)$, $1 \leqslant t \leqslant m$, which can be collected into a diagram

$$r = \big(f^0 \xrightarrow{r^1} f^1 \xrightarrow{r^2} \cdots \xrightarrow{r^{m-1}} f^{m-1} \xrightarrow{r^m} f^m\big) : \boxtimes^{i\in I} Ts\mathcal{A}_i \to s\mathcal{B},$$

we can compute the element $\widehat{r} = (r^1 \otimes \cdots \otimes r^m)\theta_{(\mathcal{A}_i);\mathcal{B}} \in \underline{\mathscr{Q}_p}(\boxtimes^{i\in I} Ts\mathcal{A}_i, T^{\geqslant 1}s\mathcal{B})$ as follows:

$$\begin{aligned}\widehat{r} = \Big[&\boxtimes^{i\in I} Ts\mathcal{A}_i \xrightarrow{\boxtimes^I \widetilde{\Delta}} \boxtimes^{i\in I} TT^{\geqslant 1}s\mathcal{A}_i \xrightarrow{\widetilde{\tau}^I} T \boxtimes_u^{i\in I} T^{\geqslant 1}s\mathcal{A}_i\\ &\xrightarrow{\nu^{\leqslant M}} TT^{\leqslant 1} \boxtimes_u^{i\in I} T^{\geqslant 1}s\mathcal{A}_i\\ &\xrightarrow[\sim]{T(\vartheta^I)^{-1}} T^n \boxtimes^{i\in I} Ts\mathcal{A}_i\\ &\xrightarrow{\sum_{k_0+\cdots+k_m+m=n} \widehat{f^0}_{k_0p_0}\otimes r^1\otimes\widehat{f^1}_{k_1p_1}\otimes\cdots\otimes r^m\otimes\widehat{f^m}_{k_mp_m}} T^{p_0+\cdots+p_m+m}s\mathcal{B}\Big],\end{aligned}$$

where $n = k_0 + \cdots + k_m + m$, $M \geqslant m$ is an arbitrary integer, and $\widehat{f^q}_{k_qp_q} : T^{k_q} \boxtimes_u^{i\in I} T^{\geqslant 1}s\mathcal{A}_i \to T^{p_q}s\mathcal{B}$ is the matrix coefficient of $\hat{f}^q$. Other presentations of this formula are

$$\begin{aligned}\widehat{r} &= \big(r^1 \otimes \cdots \otimes r^m\big)\theta_{(\mathcal{A}_i);\mathcal{B}}\\ &= \text{signed permutation and insertion of units}\\ &\quad\times \sum_{\substack{p_s\geqslant 0\\ i_t^s, j_s\in(\mathbb{Z}_{\geqslant 0})^n; i_t^s\neq 0}} \begin{aligned}&f^0_{i_1^0} \otimes \cdots \otimes f^0_{i_{p_0}^0} \otimes r^1_{j_1} \otimes f^1_{i_1^1} \otimes \cdots \otimes f^1_{i_{p_1}^1} \otimes \cdots\\ &\otimes r^m_{j_m} \otimes f^m_{i_1^m} \otimes \cdots \otimes f^m_{i_{p_m}^m},\end{aligned}\end{aligned}$$

$$\begin{aligned}\widehat{r} = \big(&\boxtimes^{i\in I} Ts\mathcal{A}_i \xrightarrow{\Delta_0^{(2m+1)}} (\boxtimes^{i\in I} Ts\mathcal{A}_i)^{\otimes(2m+1)}\\ &\xrightarrow{f^0\otimes\check{r}^1\otimes f^1\otimes\cdots\otimes\check{r}^m\otimes f^m} Ts\mathcal{B} \otimes s\mathcal{B} \otimes Ts\mathcal{B} \otimes \cdots \otimes s\mathcal{B} \otimes Ts\mathcal{B} \xrightarrow{\mu} Ts\mathcal{B}\big).\end{aligned}$$

Here μ is the multiplication in the tensor algebra quiver $Ts\mathcal{B}$. If $I = \mathbf{1}$, then the matrix elements of $\widehat{r} = (r^1 \otimes \cdots \otimes r^m)\theta_{\mathcal{A};\mathcal{B}}$ are

$$\begin{aligned}\widehat{r}_{kl} = \sum_{\substack{p_0+j_1+p_1+\cdots+j_n+p_n=k\\ m_0+m_1+\cdots+m_n=l}} f^0_{p_0m_0} \otimes r^1_{j_1} \otimes f^1_{p_1m_1} \otimes \cdots \otimes r^n_{j_n} \otimes f^n_{p_nm_n} :\\ T^ks\mathcal{A} \to T^ls\mathcal{B}.\end{aligned}$$

3.15. *Closing transformations of multifunctors*

Let C, D be closed symmetric multicategories. Let $F : \mathsf{C} \to \mathsf{D}$ be a (symmetric) multifunctor. Define a morphism in D

$$\underline{F}_{(X_i);Z} : F\underline{\mathsf{C}}\big((X_i)_{i\in I}; Z\big) \to \underline{\mathsf{D}}\big((FX_i)_{i\in I}; FZ\big) \tag{3.15.1}$$

via the commutative diagram

$$\begin{tikzcd}
(FX_i)_{i\in I}, F\underline{\mathsf{C}}\big((X_i)_{i\in I}; Y\big) \arrow[r, "(1)_I{,}\underline{F}_{(X_i);Y}"] \arrow[rd, "F\mathrm{ev}_{(X_i);Y}"'] & (FX_i)_{i\in I}, \underline{\mathsf{D}}\big((FX_i)_{i\in I}; FY\big) \arrow[d, "\mathrm{ev}_{(FX_i);FY}"] \\
& FY
\end{tikzcd}$$

The natural transformation (3.15.1) is called *closing transformation* of the multifunctor F.

4. Closed multicategories of A_∞-categories

Here we construct an A_∞-analogue of the category of functors between ordinary categories. The closedness of multicategories of A_∞-categories and of unital A_∞-categories permits it to find in them general features of closed multicategories.

4.1. *The A_∞-category of A_∞-functors*

The notions of A_∞-functors and A_∞-transformations can be generalized to the case of several arguments.

4.2. DEFINITION. Let $\mathcal{A}_i, \mathcal{B}$ be A_∞-categories, $i \in \mathbf{n}$. An *A_∞-functor* (with n arguments) $f : (\mathcal{A}_i)_{i\in\mathbf{n}} \to \mathcal{B}$ is an augmented coalgebra morphism $f : \boxtimes^{i\in\mathbf{n}} Ts\mathcal{A}_i \to Ts\mathcal{B}$ commuting with the differential.

The commutation condition is equivalent to its composition with pr_1:

$$\begin{aligned}
&\big(\boxtimes^{i\in\mathbf{n}} Ts\mathcal{A}_i \xrightarrow{f} Ts\mathcal{B} \xrightarrow{\check{b}} s\mathcal{B}\big) \\
&\quad = \big(\boxtimes^{i\in\mathbf{n}} Ts\mathcal{A}_i \xrightarrow{\sum_{i=1}^n 1^{\boxtimes(i-1)}\boxtimes b\boxtimes 1^{\boxtimes(n-i)}} \boxtimes^{i\in\mathbf{n}} Ts\mathcal{A}_i \xrightarrow{\check{f}} s\mathcal{B}\big),
\end{aligned}$$

where

$$\check{f} = \big(\boxtimes^{i\in\mathbf{n}} Ts\mathcal{A}_i \xrightarrow{f} Ts\mathcal{B} \xrightarrow{\mathrm{pr}_1} s\mathcal{B}\big). \tag{4.2.1}$$

Composition of A_∞-functors as coalgebra morphisms gives an A_∞-functor as well. Therefore, A_∞-categories as objects and A_∞-functors as morphisms form a multicategory, denoted A_∞.

4.3. DEFINITION. Let $f, g : (\mathcal{A}_i)_{i\in I} \to \mathcal{B}$ be A_∞-functors. An *A_∞-transformation* $r : f \to g : (\mathcal{A}_i)_{i\in I} \to \mathcal{B}$ is an (f, g)-coderivation $r : f \to g : \boxtimes^{i\in I} Ts\mathcal{A}_i \to Ts\mathcal{B}$.

4.4. *Closedness of the multicategory* A_∞

The multicategory A_∞ turns out to be closed. The inner homomorphism object

$$\underline{\mathsf{A}_\infty}\big((\mathcal{A}_i)_{i\in\mathbf{n}};\mathcal{B}\big)$$

is described as follows. Its objects are A_∞-functors $f:(\mathcal{A}_i)_{i\in\mathbf{n}}\to\mathcal{B}$. The elements of the graded component $\underline{\mathsf{A}_\infty}((\mathcal{A}_i)_{i\in\mathbf{n}};\mathcal{B})(f,g)^{d+1}$ are (f,g)-coderivations $r:\boxtimes^{i\in\mathbf{n}}Ts\mathcal{A}_i\to Ts\mathcal{B}$ of degree d. We have an isomorphism

$$s\underline{\mathsf{A}_\infty}\big((\mathcal{A}_i)_{i\in\mathbf{n}};\mathcal{B}\big)(f,g)\xrightarrow{\sim}\underline{\mathscr{Q}_p}\big(\boxtimes^{i\in\mathbf{n}}Ts\mathcal{A}_i,s\mathcal{B}\big)(\check{f},\check{g}),$$

$r\mapsto\check{r}=r\cdot\mathrm{pr}_1$ where $\check{f}$, $\check{g}$ are given by (4.2.1). The inverse isomorphism expresses a coderivation via its components.

The components of the differential $B:Ts\underline{\mathsf{A}_\infty}((\mathcal{A}_i)_{i\in\mathbf{n}};\mathcal{B})\to Ts\underline{\mathsf{A}_\infty}((\mathcal{A}_i)_{i\in\mathbf{n}};\mathcal{B})$ are defined as follows. For $m>1$, B_m takes an element

$$r^1\otimes\dots\otimes r^m\in T^m\underline{\mathscr{Q}_p}\big(\boxtimes^{i\in\mathbf{n}}Ts\mathcal{A}_i;s\mathcal{B}\big)$$

to the composition in $\underline{\mathscr{Q}_p}$

$$\big(r^1\otimes\dots\otimes r^m\big)B_m=\big[\boxtimes^{i\in\mathbf{n}}Ts\mathcal{A}_i\xrightarrow{(r^1\otimes\dots\otimes r^m)\theta}T^{\geqslant1}s\mathcal{B}\xrightarrow{\check{b}}s\mathcal{B}\big].$$

The component B_1 takes an element r to the difference of compositions in $\underline{\mathscr{Q}_p}$

$$\begin{aligned}(r)B_1&=\big[\boxtimes^{i\in\mathbf{n}}Ts\mathcal{A}_i\xrightarrow{r}Ts\mathcal{B}\xrightarrow{\check{b}}s\mathcal{B}\big]\\&\quad-(-)^r\big[\boxtimes^{i\in\mathbf{n}}Ts\mathcal{A}_i\xrightarrow{\sum_{j=1}^n1^{\boxtimes(j-1)}\boxtimes b\boxtimes1^{\boxtimes(n-j)}}\boxtimes^{i\in\mathbf{n}}Ts\mathcal{A}_i\xrightarrow{\check{r}}s\mathcal{B}\big].\end{aligned}$$

For $n=1$ we have

$$\begin{aligned}&(r)B_1=[r,b]^\vee=r\check{b}-(-)^rb\check{r},\\&\big[(r^1\otimes\dots\otimes r^m)B_m\big]_k=\sum_l\big(r^1\otimes\dots\otimes r^m\big)\theta_{kl}b_l,\quad\text{for }m>1.\end{aligned}$$

The evaluation A_∞-functor is given by

$$\begin{aligned}&\mathrm{ev}^{\mathsf{A}_\infty}_{(\mathcal{A}_i);\mathcal{B}}:\big(\boxtimes^{i\in I}Ts\mathcal{A}_i\big)\boxtimes Ts\underline{\mathsf{A}_\infty}\big((\mathcal{A}_i)_{i\in I};\mathcal{B}\big)\to Ts\mathcal{B},\\&a\boxtimes\big(r^1\otimes\dots\otimes r^m\big)\\&\qquad\mapsto a\big[\big(\check{r}^1\otimes\dots\otimes\check{r}^m\big)\theta\big]=a\Delta_0^{(2m+1)}\big(f^0\otimes\check{r}^1\otimes f^1\otimes\dots\otimes\check{r}^m\otimes f^m\big)\mu,\end{aligned}$$

where a is a morphism of $\boxtimes^{i\in I}Ts\mathcal{A}_i$ and $r^1\otimes\dots\otimes r^m$ is a morphism of

$$T^ms\underline{\mathsf{A}_\infty}\big((\mathcal{A}_i)_{i\in I};\mathcal{B}\big).$$

When $n=1$, we denote $\underline{\mathsf{A}_\infty}(\mathcal{A};\mathcal{B})$ also by $A_\infty(\mathcal{A},\mathcal{B})$. Being a closed multicategory, A_∞ has a uniquely determined multiplication in $\underline{\mathsf{A}_\infty}$:

$$
\begin{aligned}
M : \big(\boxtimes^{i\in\mathbf{n}} Ts\underline{\mathsf{A}_\infty}\big((\mathcal{A}_i^j)_{j\in\mathbf{m}_i};\mathcal{B}_i\big)\big) \boxtimes Ts\underline{\mathsf{A}_\infty}\big((\mathcal{B}_i)_{i\in\mathbf{n}};\mathcal{C}\big)& \\
\to Ts\underline{\mathsf{A}_\infty}\big(\mathcal{A}_1^1,\dots,\mathcal{A}_1^{m_1},\dots,\mathcal{A}_n^1,\dots,\mathcal{A}_n^{m_n};\mathcal{C}\big)&,
\end{aligned}
$$

determined uniquely by its components

$$
\begin{aligned}
M_{k_1\dots k_n l} : \big(\boxtimes^{i\in\mathbf{n}} T^{k_i} s\underline{\mathsf{A}_\infty}\big((\mathcal{A}_i^j)_{j\in\mathbf{m}_i};\mathcal{B}_i\big)\big) \boxtimes T^l s\underline{\mathsf{A}_\infty}\big((\mathcal{B}_i)_{i\in\mathbf{n}};\mathcal{C}\big)& \\
\to s\underline{\mathsf{A}_\infty}\big(((\mathcal{A}_i^j)_{j\in\mathbf{m}_i})_{i\in\mathbf{n}};\mathcal{C}\big)&.
\end{aligned}
$$

It turns out that $M_{k_1\dots k_n l} = 0$ for $l > 1$.

If $n = 1$, we find the components of M as follows:

$$
\begin{aligned}
\big[(r^1\otimes\dots\otimes r^k \boxtimes g)M_{k0}\big]_m &= \sum_p (\check{r}^1\otimes\dots\otimes\check{r}^k)\theta_{mp} g_p, \\
\big[(r^1\otimes\dots\otimes r^k \boxtimes t)M_{k1}\big]_m &= \sum_p (\check{r}^1\otimes\dots\otimes\check{r}^k)\theta_{mp} t_p.
\end{aligned}
$$

Let $f : (\mathcal{A}_i)_{i\in I} \to \mathcal{B}$ be an A_∞-functor and let j be an element of I. Choose a family of objects $(X_i)_{i\neq j} \in \prod_{i\neq j}\operatorname{Ob}\mathcal{A}_i$. We view them as A_∞-functors $X_i : () \to \mathcal{A}_i$. Define an A_∞-functor $f|_j^{(X_i)_{i\neq j}} : \mathcal{A}_j \to \mathcal{B}$, the restriction of f to the argument j, as the image of $(\mathrm{id}_{\mathcal{A}_j}, (X_i)_{i\neq j}, f)$ under the composition map

$$
\mu^{\mathsf{A}_\infty}_{\{j\}\hookrightarrow I} : \mathsf{A}_\infty(\mathcal{A}_j;\mathcal{A}_j) \times \prod_{i\neq j}\mathsf{A}_\infty(;\mathcal{A}_i) \times \mathsf{A}_\infty\big((\mathcal{A}_i)_{i\in I};\mathcal{B}\big) \to \mathsf{A}_\infty(\mathcal{A}_j;\mathcal{B}).
$$

The A_∞-functor $f|_j^{(X_i)_{i\neq j}}$ takes an object $X_j \in \operatorname{Ob}\mathcal{A}_j$ to the object $((X_i)_{i\in I})f \in \operatorname{Ob}\mathcal{B}$. The k-th component is

$$
\begin{aligned}
&\big(f|_j^{(X_i)_{i\neq j}}\big)_k : T^k s\mathcal{A}_j(X_j, Y_j) \\
&\quad \xrightarrow[\sim]{\lambda^{\{j\}\hookrightarrow I}} \otimes^{i\in I}\big[\big(T^0 s\mathcal{A}_i(X_i,X_i)\big)_{i<j}, T^k s\mathcal{A}_j(X_j,Y_j), \big(T^0 s\mathcal{A}_i(X_i,X_i)\big)_{i>j}\big] \\
&\quad \xrightarrow{f_{ke_j}} s\mathcal{B}\big((X_1,\dots,X_n)f, (X_1,\dots,X_{j-1},Y_j,X_{j+1},\dots,X_n)f\big),
\end{aligned}
$$

where $ke_j = (0,\dots,0,k,0,\dots,0) \in \mathbb{Z}^n$ has k on j-th place. Similarly one can define restriction of f to a subset of arguments $J \subset I$.

4.5. Definition. Let $\mathcal{A}_i, \mathcal{B}$ be unital A_∞-categories, $i \in I$. An A_∞-functor

$$
f : (\mathcal{A}_i)_{i\in I} \to \mathcal{B}
$$

is called *unital* if the A_∞-functors $f|_j^{(X_i)_{i\neq j}} : \mathcal{A}_j \to \mathcal{B}$ are unital for all $j \in I$.

The subsets $\mathsf{A}^{\mathsf{u}}_\infty((\mathcal{A}_i)_{i\in I};\mathcal{B}) \subset \mathsf{A}_\infty((\mathcal{A}_i)_{i\in I};\mathcal{B})$ form a submulticategory $\mathsf{A}^{\mathsf{u}}_\infty \subset \mathsf{A}_\infty$ with unital A_∞-categories as objects, and unital A_∞-functors as multimorphisms.

4.6. Proposition. *(See* [3].*) The multicategory $\mathsf{A}^{\mathsf{u}}_\infty$ is closed. If $(\mathcal{A}_i)_{i\in I}$, $\mathcal{B}$ are unital A_∞-categories, then $\underline{\mathsf{A}^{\mathsf{u}}_\infty}((\mathcal{A}_i)_{i\in I};\mathcal{B}) \subset \underline{\mathsf{A}_\infty}((\mathcal{A}_i)_{i\in I};\mathcal{B})$ is the full A_∞-subcategory,* *whose objects are unital A_∞-functors $f : (\mathcal{A}_i)_{i \in I} \to \mathcal{B}$. The evaluation A_∞-functor* $\mathrm{ev}^{\mathsf{A}^{\mathsf{u}}_\infty} :$ $(\mathcal{A}_i)_{i\in I}, \underline{\mathsf{A}^{\mathsf{u}}_\infty}((\mathcal{A}_i)_{i\in I}; \mathcal{B}) \to \mathcal{B}$ *is the restriction of* $\mathrm{ev}^{\mathsf{A}_\infty} : (\mathcal{A}_i)_{i\in I}, \underline{\mathsf{A}_\infty}((\mathcal{A}_i)_{i\in I}; \mathcal{B}) \to \mathcal{B}$.

When $n = 1$, we denote $\underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{A}; \mathcal{B})$ also by $A^u_\infty(\mathcal{A}, \mathcal{B})$.

Let $\mathcal{A}$ be a unital A_∞-category. Since $\underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{A}; \mathcal{A})$ is unital, there is a unit element $\mathbf{i}^{\mathcal{A}} \in \underline{\mathsf{A}_\infty}(\mathcal{A}; \mathcal{A})(\mathrm{id}_{\mathcal{A}}, \mathrm{id}_{\mathcal{A}})$ called also a unit transformation $\mathbf{i}^{\mathcal{A}} : \mathrm{id} \to \mathrm{id} : \mathcal{A} \to \mathcal{A}$. For any unital A_∞-functor $f : (\mathcal{A}_i)_{i\in I} \to \mathcal{B}$ the unit element in $\underline{\mathsf{A}^{\mathsf{u}}_\infty}((\mathcal{A}_i)_{i\in I}; \mathcal{B})(f, f)$ can be chosen as the transformation $f\mathbf{i}^{\mathcal{B}} : f \to f : (\mathcal{A}_i)_{i\in I} \to \mathcal{B}$ [24, Section 7].

4.7. *2-category of A_∞-categories*

Unital A_∞-categories form a 2-category $\overline{A^u_\infty}$, whose 2-morphism sets are $\Bbbk$-modules, and whose vertical and horizontal compositions of 2-morphisms are (poly)linear. Indeed, the objects of $\overline{A^u_\infty}$ are unital A_∞-categories, 1-morphisms are unital A_∞-functors, 2-morphisms rs^{-1} are equivalence classes of natural A_∞-transformations r. An A_∞-transformation $r : f \to g : \mathcal{A} \to \mathcal{B}$, $r \in A_\infty(\mathcal{A}, \mathcal{B})(f, g)[1]$ is *natural*, if $\deg r = -1$ and $rB_1 = 0$. Two natural A_∞-transformations $p, r : f \to g : \mathcal{A} \to \mathcal{B}$ are *equivalent* ($p \equiv r$) if they are homologous, that is, differ by a B_1-boundary. If the A_∞-transformations $f \xrightarrow{r} g \xrightarrow{p} h : \mathcal{A} \to \mathcal{B}$ are natural, the vertical composition of the 2-morphisms rs^{-1} and ps^{-1} is represented by $(r \otimes p)B_2 s^{-1}$. For any unital A_∞-functor $f : \mathcal{A} \to \mathcal{B}$ the natural A_∞-transformations $f\mathbf{i}^{\mathcal{B}} \equiv \mathbf{i}^{\mathcal{A}} f : f \to f : \mathcal{A} \to \mathcal{B}$ represent the identity 2-morphism 1_f. Equivalences in the 2-category $\overline{A^u_\infty}$ are called A_∞-equivalences.

4.8. COROLLARY *(to [24, Theorem 8.8]). Let $\mathcal{C}$ be an A_∞-category and let $\mathcal{A}$ be a unital A_∞-category. Let $f : \mathcal{C} \to \mathcal{A}$ be an A_∞-functor such that for all objects X, Y of $\mathcal{C}$ the chain map $f_1 : (s\mathcal{C}(X, Y), b_1) \to (s\mathcal{A}(X\phi, Y\phi), b_1)$ is homotopy invertible, and* $\mathrm{Ob}\, f : \mathrm{Ob}\, \mathcal{C} \to \mathrm{Ob}\, \mathcal{A}$ *is surjective. Then $\mathcal{C}$ is unital and f is an A_∞-equivalence.*

The theorem mentioned is proved similarly to Proposition 2.6. Combining this proposition with the above corollary we see that if in the assumptions of Proposition 2.6 $\mathcal{B}$ is unital, then the obtained A_∞-category $\mathcal{C}$ is unital as well and it is A_∞-equivalent to the full A_∞-subcategory $\mathcal{A}$ of $\mathcal{B}$ with $\mathrm{Ob}\, \mathcal{A} = \mathrm{Im}(\mathrm{Ob}\, f)$.

A natural A_∞-transformation $r : f \to g : \mathcal{A} \to \mathcal{B}$ of unital A_∞-functors $f, g : \mathcal{A} \to \mathcal{B}$ determines a natural transformation $H^0(r) : H^0(f) \to H^0(g) : H^0(\mathcal{A}) \to H^0(\mathcal{B})$ given by the family of elements $[{}_X r_0 s^{-1}] \in H^0(\mathcal{B})(Xf, Xg)$, $X \in \mathrm{Ob}\, \mathcal{A}$. Homologous natural A_∞-transformations determine the same natural transformation. Together with the correspondences $\mathcal{C} \mapsto H^0(\mathcal{C})$, $f \mapsto H^0(f)$ this defines a strict 2-functor H^0 from the 2-category $\overline{A^u_\infty}$ to the 2-category of $\Bbbk$-linear categories, $\Bbbk$-linear functors, and natural transformations.

4.8.1. *Yoneda lemma for unital A_∞-categories*

An A_∞-analogue of the Yoneda lemma for strictly unital A_∞-categories was proved by Fukaya [9, Theorem 9.1]. The general case of unital A_∞-categories was considered in [25, Appendix A]. Let $\mathcal{A}$ be an A_∞-category. The differential graded category of complexes of $\Bbbk$-modules is denoted $\mathsf{C}_\Bbbk$. For any object X of $\mathcal{A}$ we define an A_∞-functor $h^X : \mathcal{A} \to \mathsf{C}_\Bbbk$ as follows. It maps an object Z to the complex $h^X Z = (s\mathcal{A}(X,Z), b_1)$. The A_∞-functor h^X is completely specified by its components h^X_k for $k \geqslant 1$:

$$\begin{aligned} h^X_k = \bigl[& s\mathcal{A}(Z_0,Z_1) \otimes \dots \otimes s\mathcal{A}(Z_{k-1},Z_k) \\ & \xrightarrow{\text{coev}} \mathsf{C}_\Bbbk\bigl(h^X Z_0, h^X Z_0 \otimes h^{Z_0} Z_1 \otimes \dots \otimes h^{Z_{k-1}} Z_k\bigr) \\ & \xrightarrow{\mathsf{C}_\Bbbk(1,b_{k+1})} \mathsf{C}_\Bbbk\bigl(s\mathcal{A}(X,Z_0), s\mathcal{A}(X,Z_k)\bigr) \\ & \xrightarrow{\sigma} s\mathsf{C}_\Bbbk\bigl(s\mathcal{A}(X,Z_0), s\mathcal{A}(X,Z_k)\bigr)\bigr], \end{aligned}$$

where for a graded $\Bbbk$-module P we define the map $\sigma : P \to P[1]$ by $p \mapsto (-)^{\deg p} ps$. The coevaluation map coev is a consequence of the closed monoidal structure of $\mathsf{C}_\Bbbk$.

The *opposite quiver* $\mathcal{A}^{\text{op}}$ is defined as the quiver with the same class of objects $\operatorname{Ob}\mathcal{A}^{\text{op}} = \operatorname{Ob}\mathcal{A}$, and with graded $\Bbbk$-modules of morphisms $\mathcal{A}^{\text{op}}(X,Y) = \mathcal{A}(Y,X)$. Let $\gamma : Ts\mathcal{A}^{\text{op}} \to Ts\mathcal{A}$ denote the following coalgebra anti-isomorphism:

$$\begin{aligned} \gamma = (-1)^k \omega^0_c : & s\mathcal{A}^{\text{op}}(X_0,X_1) \otimes \dots \otimes s\mathcal{A}^{\text{op}}(X_{k-1},X_k) \\ & \to s\mathcal{A}(X_k,X_{k-1}) \otimes \dots \otimes s\mathcal{A}(X_1,X_0), \end{aligned}$$

where the permutation

$$\omega^0 = \begin{pmatrix} 1 & 2 & \dots & k-1 & k \\ k & k-1 & \dots & 2 & 1 \end{pmatrix}$$

is the longest element of $\mathfrak{S}_k$, and ω^0_c is the corresponding signed permutation, and where the action of ω^0 on tensor products is the standard one. The *opposite A_∞-category* $\mathcal{A}^{\text{op}}$ to an A_∞-category $\mathcal{A}$ is the opposite quiver, equipped with the differential $b^{\text{op}} = \gamma b \gamma : Ts\mathcal{A}^{\text{op}} \to Ts\mathcal{A}^{\text{op}}$. The components of b^{op} are computed as follows:

$$\begin{aligned} b^{\text{op}}_k = (-)^{k+1}\bigl[& s\mathcal{A}^{\text{op}}(X_0,X_1) \otimes \dots \otimes s\mathcal{A}^{\text{op}}(X_{k-1},X_k) \\ & \xrightarrow{\omega^0_c} s\mathcal{A}(X_k,X_{k-1}) \otimes \dots \otimes s\mathcal{A}(X_1,X_0) \\ & \xrightarrow{b_k} s\mathcal{A}(X_k,X_0) = s\mathcal{A}^{\text{op}}(X_0,X_k)\bigr]. \end{aligned}$$

For an arbitrary A_∞-category $\mathcal{A}$ there is the Yoneda A_∞-functor $Y : \mathcal{A}^{\text{op}} \to \underline{\mathsf{A}_\infty}(\mathcal{A};\mathsf{C}_\Bbbk)$, $X \mapsto h^X$, cf. Fukaya [9]. If $\mathcal{A}$ is unital, then Y is an A_∞-equivalence of $\mathcal{A}^{\text{op}}$ with the full subcategory of $\underline{\mathsf{A}_\infty}(\mathcal{A};\mathsf{C}_\Bbbk)$ whose objects are A_∞-functors $h^X : \mathcal{A} \to \mathsf{C}_\Bbbk$ [25, Theorem A.11]. Since $\underline{\mathsf{A}_\infty}(\mathcal{A};\mathsf{C}_\Bbbk)$ is a differential graded category, it follows that each $\mathscr{U}$-small unital A_∞-category $\mathcal{A}$ is A_∞-equivalent to a $\mathscr{U}$-small differential graded category.

4.9. $\mathsf{A}^{\mathsf{u}}_\infty$-*functors and* $\mathsf{A}^{\mathsf{u}}_\infty$-*transformations*

According to the general picture of closed multicategories, the multicategory $\underline{\mathsf{A}^{\mathsf{u}}_\infty}$ is enriched in the multicategory $\mathsf{A}^{\mathsf{u}}_\infty$. This means that there exists a unital A_∞-functor

$$\mu^{\underline{\mathsf{A}^{\mathsf{u}}_\infty}}_{\mathbf{1}\to\mathbf{1}} : \underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{A};\mathcal{B}), \underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{B};\mathcal{C}) \to \underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{A};\mathcal{C}),$$

the composition, or equivalently, an augmented differential coalgebra morphism

$$M : Ts\underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{A};\mathcal{B}) \boxtimes Ts\underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{B};\mathcal{C}) \to Ts\underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{A};\mathcal{C}).$$

This morphism satisfies the associativity equation.

An $\mathsf{A}^{\mathsf{u}}_\infty$*-functor* $F : \underline{\mathsf{A}^{\mathsf{u}}_\infty} \to \underline{\mathsf{A}^{\mathsf{u}}_\infty}$ consists of the map $\operatorname{Ob} F : \operatorname{Ob} \mathsf{A}^{\mathsf{u}}_\infty \to \operatorname{Ob} \mathsf{A}^{\mathsf{u}}_\infty$ and of unital A_∞-functors $F : \underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{A};\mathcal{B}) \to \underline{\mathsf{A}^{\mathsf{u}}_\infty}(F\mathcal{A};F\mathcal{B})$. These A_∞-functors have to map the unit $\mathrm{id}_\mathcal{A}$ to the unit $\mathrm{id}_{F\mathcal{A}}$ and to satisfy the equation

$$\begin{tikzcd}
\underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{A};\mathcal{B}), \underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{B};\mathcal{C}) \arrow[r, "F,F"] \arrow[d, "\mu^{\mathsf{A}^{\mathsf{u}}_\infty}_{\mathbf{1}\to\mathbf{1}}"'] \arrow[dr, phantom, "="] & \underline{\mathsf{A}^{\mathsf{u}}_\infty}(F\mathcal{A};F\mathcal{B}), \underline{\mathsf{A}^{\mathsf{u}}_\infty}(F\mathcal{B};F\mathcal{C}) \arrow[d, "\mu^{\mathsf{A}^{\mathsf{u}}_\infty}_{\mathbf{1}\to\mathbf{1}}"'] \\
\underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{A};\mathcal{C}) \arrow[r, "F"] & \underline{\mathsf{A}^{\mathsf{u}}_\infty}(F\mathcal{A};F\mathcal{C})
\end{tikzcd}$$

An $\mathsf{A}^{\mathsf{u}}_\infty$*-transformation* $\lambda : G \to F : \underline{\mathsf{A}^{\mathsf{u}}_\infty} \to \underline{\mathsf{A}^{\mathsf{u}}_\infty}$ is a collection of unital A_∞-functors $\lambda_\mathcal{A} : G\mathcal{A} \to F\mathcal{A}$, $\mathcal{A} \in \operatorname{Ob} \mathsf{A}^{\mathsf{u}}_\infty$, such that the following equation holds:

$$\begin{tikzcd}
\underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{C};\mathcal{D}) \arrow[r, "G"] \arrow[d, "F"'] \arrow[dr, phantom, "="] & \underline{\mathsf{A}^{\mathsf{u}}_\infty}(G\mathcal{C};G\mathcal{D}) \arrow[d, "(1\boxtimes\lambda_\mathcal{D})M = \underline{\mathsf{A}^{\mathsf{u}}_\infty}(1;\lambda_\mathcal{D})"] \\
\underline{\mathsf{A}^{\mathsf{u}}_\infty}(F\mathcal{C};F\mathcal{D}) \arrow[r, "(\lambda_\mathcal{C}\boxtimes 1)M", "\overset{\|}{\underline{\mathsf{A}^{\mathsf{u}}_\infty}(\lambda_\mathcal{C};1)}"'] & \underline{\mathsf{A}^{\mathsf{u}}_\infty}(G\mathcal{C};F\mathcal{D})
\end{tikzcd}$$

In [24,26,25] the $\mathsf{A}^{\mathsf{u}}_\infty$-category $\underline{\mathsf{A}^{\mathsf{u}}_\infty}$, $\mathsf{A}^{\mathsf{u}}_\infty$-functors and $\mathsf{A}^{\mathsf{u}}_\infty$-transformations were called respectively A^u_∞, A^u_∞-2-functors and A^u_∞-2-transformations. $\mathsf{A}^{\mathsf{u}}_\infty$-functors $\underline{\mathsf{A}^{\mathsf{u}}_\infty} \to \underline{\mathsf{A}^{\mathsf{u}}_\infty}$ and their $\mathsf{A}^{\mathsf{u}}_\infty$-transformations form a strict monoidal category A^u_∞-2. Algebras in this monoidal category are called A^u_∞*-2-monads*. Analogous notions without the unitality requirement are called A_∞-2-functors, A_∞-2-transformations and A_∞-2-monads.

An example of an $\mathsf{A}^{\mathsf{u}}_\infty$-functor is given by the A_∞-category of A_∞-functors. Let $\mathcal{D}$ be a unital A_∞-category. Then the correspondence $\mathcal{A} \mapsto \underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{D};\mathcal{A})$ extends to an $\mathsf{A}^{\mathsf{u}}_\infty$-functor $\underline{\mathsf{A}^{\mathsf{u}}_\infty} \to \underline{\mathsf{A}^{\mathsf{u}}_\infty}$.

For an arbitrary A_∞-category $\mathcal{C}$ there is an $\mathsf{A}^{\mathsf{u}}_\infty$-functor $\underline{\mathsf{A}^{\mathsf{u}}_\infty} \to \underline{\mathsf{A}^{\mathsf{u}}_\infty}$, $\mathcal{A} \mapsto \underline{\mathsf{A}_\infty}(\mathcal{C};\mathcal{A})$. By a theorem from [3] it is representable by the pair $(\mathcal{C}^{\mathsf{su}}, u_{\mathsf{su}} : \mathcal{C} \hookrightarrow \mathcal{C}^{\mathsf{su}})$, that is, the $\mathsf{A}^{\mathsf{u}}_\infty$-transformation $\underline{\mathsf{A}_\infty}(u_{\mathsf{su}};1) : \underline{\mathsf{A}^{\mathsf{u}}_\infty}(\mathcal{C}^{\mathsf{su}};\mathcal{A}) \to \underline{\mathsf{A}_\infty}(\mathcal{C};\mathcal{A})$ consists of A_∞-equivalences. Morally this means that the strictly unital envelope of an A_∞-category is simultaneously its unital envelope in the weak sense.

A pair $(\mathcal{D}, e \in \operatorname{Ob} F\mathcal{D})$ representing a representable $\mathsf{A}^{\mathsf{u}}_\infty$-functor $F : \underline{\mathsf{A}^{\mathsf{u}}_\infty} \to \underline{\mathsf{A}^{\mathsf{u}}_\infty}$ is unique up to an equivalence. Indeed, the same pair $(\mathcal{D}, e)$ represents the strict 2-functor $\underline{A^u_\infty} \to \mathcal{C}\mathrm{at}$, $\mathcal{A} \mapsto H^0(F\mathcal{A})$. Such a representing pair is unique up to an equivalence by the 2-category analogue of the Yoneda lemma [25, Appendix C.17].

5. Quotients of A_∞-categories

Two different constructions of quotients of A_∞-categories, one via relatively free A_∞-categories, and another via subquivers of the tensor quiver, turn out to be equivalent.

5.1. *Quotients via representability of an $\mathsf{A}^{\mathrm{u}}_\infty$-functor*

An A_∞-functor $g:\mathcal{B}\to\mathcal{A}$ from a unital A_∞-category $\mathcal{B}$ is *contractible* if for all objects X, Y of $\mathcal{B}$ the chain map $g_1:s\mathcal{B}(X,Y)\to s\mathcal{A}(Xg,Yg)$ is null-homotopic. If $g:\mathcal{B}\to\mathcal{A}$ is a unital A_∞-functor, then it is contractible if and only if for any $X\in\operatorname{Ob}\mathcal{B}$ and any $V\in\operatorname{Ob}\mathcal{A}$ the complexes $s\mathcal{A}(Xg,V)$ and $s\mathcal{A}(V,Xg)$ are contractible, see [27, Section 6].

Let $\mathcal{B}$ be a full A_∞-subcategory of a unital A_∞-category $\mathcal{C}$. Let $\mathcal{A}$ be an arbitrary unital A_∞-category. Denote by $\underline{\mathsf{A}^{\mathrm{u}}_\infty}(\mathcal{C};\mathcal{A})_{\mathrm{mod}\mathcal{B}}$ the full A_∞-subcategory of $\underline{\mathsf{A}^{\mathrm{u}}_\infty}(\mathcal{C};\mathcal{A})$, whose objects are unital A_∞-functors $\mathcal{C}\to\mathcal{A}$, whose restriction to $\mathcal{B}$ is contractible. It is allowed to consider A_∞-categories with the empty set of objects. The correspondence $\mathcal{A}\mapsto\underline{\mathsf{A}^{\mathrm{u}}_\infty}(\mathcal{C};\mathcal{A})_{\mathrm{mod}\mathcal{B}}$ is an $\mathsf{A}^{\mathrm{u}}_\infty$-functor $\underline{\mathsf{A}^{\mathrm{u}}_\infty}\to\underline{\mathsf{A}^{\mathrm{u}}_\infty}$. It turns out to be representable, that is, there is an $\mathsf{A}^{\mathrm{u}}_\infty$-transformation $\underline{\mathsf{A}^{\mathrm{u}}_\infty}(\mathcal{D};\mathcal{A})\to\underline{\mathsf{A}^{\mathrm{u}}_\infty}(\mathcal{C};\mathcal{A})_{\mathrm{mod}\,\mathcal{B}}$ consisting of A_∞-equivalences.

5.2. THEOREM. *(See* [25, *Theorem* 1.3]*.) Under the assumptions above there exists a unital A_∞-category $\mathcal{D}$ and a unital A_∞-functor $e:\mathcal{C}\to\mathcal{D}$ such that*

(1) *the composition $\mathcal{B}\hookrightarrow\mathcal{C}\xrightarrow{e}\mathcal{D}$ is contractible*;

(2) *the strict A_∞-functor given by composition with e*

$$(e\boxtimes 1)M:\underline{\mathsf{A}^{\mathrm{u}}_\infty}(\mathcal{D};\mathcal{A})\to\underline{\mathsf{A}^{\mathrm{u}}_\infty}(\mathcal{C};\mathcal{A})_{\mathrm{mod}\,\mathcal{B}},\qquad f\mapsto ef,$$

is an A_∞-equivalence for an arbitrary unital A_∞-category $\mathcal{A}$.

The pair $(\mathcal{D},e:\mathcal{C}\hookrightarrow\mathcal{D})$ representing the $\mathsf{A}^{\mathrm{u}}_\infty$-functor $\mathcal{A}\mapsto\underline{\mathsf{A}^{\mathrm{u}}_\infty}(\mathcal{C};\mathcal{A})_{\mathrm{mod}\,\mathcal{B}}$ is unique up to an equivalence due to the general result concerning uniqueness of a representing object, see Section 4.9. It is called the quotient of $\mathcal{C}$ by $\mathcal{B}$ and is denoted $\mathsf{q}(\mathcal{C}|\mathcal{B})$.

The proof of Theorem 5.2 reduces to the case when $\mathcal{C}$ is strictly unital. Under these assumptions the representing A_∞-category $\mathcal{D}=\mathsf{q}(\mathcal{C}|\mathcal{B})$ is constructed explicitly as an A_∞-category, freely generated over $\mathcal{C}$ by the application of contracting homotopies H to morphisms, whose source or target is in $\mathcal{B}$. The universality of $\mathcal{D}$ is based on the fact that it is *relatively free* over $\mathcal{C}$, that is, it admits a filtration

$$\mathcal{C}=\mathcal{D}_0\subset\mathcal{Q}_1\subset\mathcal{D}_1\subset\mathcal{Q}_2\subset\mathcal{D}_2\subset\mathcal{Q}_3\subset\cdots\subset\operatorname{colim}_j\mathcal{D}_j=\mathcal{D}$$

by A_∞-subcategories $\mathcal{D}_j$ and differential graded subquivers $\mathcal{Q}_j$, such that the graded subquiver $\mathcal{D}_j\subset\mathcal{Q}_{j+1}$ has a direct complement $\mathcal{N}_{j+1}$ (a graded subquiver of $\mathcal{Q}_{j+1}$), and such that $\mathcal{D}_{j+1}$ is generated by $\mathcal{N}_{j+1}$ over $\mathcal{D}_j$. This filtration permits us to write down a sequence of restriction A_∞-functors

$$\begin{aligned}\underline{\mathsf{A}^{\mathrm{u}}_\infty}(\mathcal{C};\mathcal{A})_{\mathrm{mod}\,\mathcal{B}}\leftarrow A^{\psi u}_{\infty1}(\mathcal{D}_0,\mathcal{Q}_1;\mathcal{A})&\leftarrow A^{\psi u}_{\infty}(\mathcal{D}_1,\mathcal{A})\\ \leftarrow A^{\psi u}_{\infty1}(\mathcal{D}_1,\mathcal{Q}_2;\mathcal{A})&\leftarrow A^{\psi u}_{\infty}(\mathcal{D}_2,\mathcal{A})\\ \leftarrow A^{\psi u}_{\infty1}(\mathcal{D}_2,\mathcal{Q}_3;\mathcal{A})&\leftarrow\cdots\end{aligned}\tag{5.2.1}$$

and to prove that each of these A_∞-functors is an equivalence, surjective on objects. The A_∞-category $A^{\psi u}_{\infty 1}(\mathcal{D}_j, \mathcal{Q}_{j+1}; \mathcal{A})$ is defined via the pull-back square

$$\begin{tikzcd}
A^{\psi u}_{\infty 1}(\mathcal{D}_j, \mathcal{Q}_{j+1}; \mathcal{A}) \arrow[r] \arrow[d] \arrow[dr, phantom, "\lrcorner", very near start] & \underline{\mathcal{Q}_p}(\mathcal{Q}_{j+1}, \mathcal{A}) \arrow[d] \\
A^{\psi u}_{\infty}(\mathcal{D}_j, \mathcal{A}) \arrow[r] & \underline{\mathcal{Q}_p}(\mathcal{D}_j, \mathcal{A})
\end{tikzcd}$$

where the quivers in the right column get an A_∞-structure from $\mathcal{A}$. The A_∞-categories $\mathcal{D}_j$ are not unital, but only pseudounital – there are distinguished cycles $\mathbf{i}_0^{\mathcal{C}} \in (s\mathcal{D}_j)^{-1}$, which are not unit elements of $\mathcal{D}_j$ if $j > 0$. The index ψu in $A^{\psi u}_\infty$ indicates that we consider pseudounital A_∞-functors – a generalization of unital ones. Their first components preserve the distinguished cycles up to a boundary. The A_∞-equivalence $\underline{\mathsf{A}^{\mathrm{u}}_\infty}(\mathcal{D}; \mathcal{A}) = A^{\psi u}_\infty(\mathcal{D}, \mathcal{A}) \to \underline{\mathsf{A}^{\mathrm{u}}_\infty}(\mathcal{C}; \mathcal{A})_{\mathrm{mod}\ \mathcal{B}}$ is the limit case of (5.2.1).

5.3. *Quotients via the bar resolution*

There is a construction from [27] which gives a unital A_∞-category $\mathsf{D}(\mathcal{C}|\mathcal{B})$ that is also a kind of quotient. It is a generalization of Drinfeld's quotient construction for differential graded categories [8, Section 3]. In order to introduce $\mathsf{D}(\mathcal{C}|\mathcal{B})$ we endow $s^{-1}T^{\geqslant 1}s\mathcal{C}$ with a structure of A_∞-category, given by $\underline{b} : T(T^{\geqslant 1}s\mathcal{C}) \to T(T^{\geqslant 1}s\mathcal{C})$ with the components $\underline{b}_0 = 0$, $\underline{b}_1 = b : T^{\geqslant 1}s\mathcal{C} \to T^{\geqslant 1}s\mathcal{C}$, $\underline{b}_k = 0$ for $k > 1$.

There is an augmented coalgebra automorphism $\boldsymbol{\mu} : TT^{\geqslant 1}s\mathcal{C} \to TT^{\geqslant 1}s\mathcal{C}$, $\mathrm{Ob}\,\boldsymbol{\mu} = \mathrm{id}_{\mathrm{Ob}\,\mathcal{C}}$, specified by its components $\boldsymbol{\mu}_k = \mu^{\mathbf{k}} : T^k T^{\geqslant 1}s\mathcal{C} \to T^{\geqslant 1}s\mathcal{C}$, $k \geqslant 1$, where $\mu : T^{\geqslant 1}s\mathcal{C} \otimes T^{\geqslant 1}s\mathcal{C} \to T^{\geqslant 1}s\mathcal{C}$ is the multiplication in the tensor algebra, $\mu^{\mathbf{1}} = 1 : T^{\geqslant 1}s\mathcal{C} \to T^{\geqslant 1}s\mathcal{C}$, $\mu^{\mathbf{2}} = \mu$, $\mu^{\mathbf{3}} = (\mu \otimes 1)\mu : (T^{\geqslant 1}s\mathcal{C})^{\otimes 3} \to T^{\geqslant 1}s\mathcal{C}$ and so on. Its inverse is the cocategory automorphism $\boldsymbol{\mu}^{-1} = \boldsymbol{\mu}^{-} : TT^{\geqslant 1}s\mathcal{C} \to TT^{\geqslant 1}s\mathcal{C}$ with the components $\boldsymbol{\mu}^{-}_k = (-)^{k-1}\mu^{\mathbf{k}} : T^k T^{\geqslant 1}s\mathcal{C} \to T^{\geqslant 1}s\mathcal{C}$.

The conjugate codifferential $\bar{b} = \boldsymbol{\mu}\underline{b}\boldsymbol{\mu}^{-1} : T(T^{\geqslant 1}s\mathcal{C}) \to T(T^{\geqslant 1}s\mathcal{C})$ has the following components: $\bar{b}_0 = 0$, $\bar{b}_1 = b$ and for $n \geqslant 2$

$$\bar{b}_n = \mu^{\mathbf{n}} \sum_{m;q<k;t<l} 1^{\otimes q} \otimes b_m \otimes 1^{\otimes t} : T^k s\mathcal{C} \otimes (T^{\geqslant 1}s\mathcal{C})^{\otimes n-2} \otimes T^l s\mathcal{C} \to T^{\geqslant 1}s\mathcal{C},$$

$$\begin{aligned}\bar{b}_n = \mu^{\mathbf{n}}b &- \big(1 \otimes \mu^{\mathbf{n-1}}b\big)\mu - \big(\mu^{\mathbf{n-1}}b \otimes 1\big)\mu \\ &+ \big(1 \otimes \mu^{\mathbf{n-2}}b \otimes 1\big)\mu^{\mathbf{3}} : \big(T^{\geqslant 1}s\mathcal{C}\big)^{\otimes n} \to T^{\geqslant 1}s\mathcal{C},\end{aligned}$$

for all $n \geqslant 0$. Thus, we have an A_∞-category $(s^{-1}T^{\geqslant 1}s\mathcal{C}, \bar{b})$.

As a graded $\Bbbk$-quiver $\mathcal{E} = \mathsf{D}(\mathcal{C}|\mathcal{B})$ has the set of objects $\mathrm{Ob}\,\mathcal{E} = \mathrm{Ob}\,\mathcal{C}$, the morphisms for $X, Y \in \mathrm{Ob}\,\mathcal{C}$ are

$$\begin{aligned}s\mathcal{E}(X, Y) = \oplus_{C_1,\dots,C_{n-1}\in\mathcal{B}} s\mathcal{C}(X, C_1) &\otimes s\mathcal{C}(C_1, C_2) \otimes \cdots \\ &\otimes s\mathcal{C}(C_{n-2}, C_{n-1}) \otimes s\mathcal{C}(C_{n-1}, Y),\end{aligned}$$

where the summation runs over all sequences of objects $(C_1, \dots, C_{n-1})$ of $\mathcal{B}$. To the empty sequence ($n = 1$) corresponds the summand $s\mathcal{C}(X, Y)$. Thus, $s\mathcal{E}$ is a direct summand of $T^{\geqslant 1}s\mathcal{C}$. The operations $\bar{b}_n$ restrict to maps $s\mathcal{E}^{\otimes n} \to s\mathcal{E}$ via the natural embedding $s\mathcal{E} \subset T^{\geqslant 1}s\mathcal{C}$ [27, Proposition 2.2]. Hence, $\bar{b}$ turns $\mathcal{E}$ into an A_∞-category.

5.4. THEOREM. *(See [25].) Let $\mathcal{B}$ be a full A_∞-subcategory of a unital A_∞-category $\mathcal{C}$. Then the unital A_∞-categories $\mathsf{q}(\mathcal{C}|\mathcal{B})$ and $\mathsf{D}(\mathcal{C}|\mathcal{B})$ are A_∞-equivalent.*

6. The action of differential graded categories on A_∞-categories

Some objects of Monoidal categories have an algebra structure and may act on other objects. Similarly some objects of symmetric Monoidal $\mathcal{C}$at-categories have a structure imitating the lax Monoidal category structure described in Definition 3.2. They may act on other objects likewise a monoidal category acts on another category. For instance, the category of symmetric multicategories $\mathcal{SMC}$atm is a symmetric Monoidal $\mathcal{C}$at-category; the symmetric multicategory $\widehat{\mathbf{dg}\text{-}\mathcal{C}\mathrm{at}}$ possesses an analogue of a symmetric Monoidal category structure, coming from the tensor product $\boxtimes$ in $\mathbf{dg}$-$\mathcal{C}$at; and the multicategory $\widehat{\mathbf{dg}\text{-}\mathcal{C}\mathrm{at}}$ acts on the multicategory $\mathsf{A}^{\mathsf{u}}_\infty$. The action is a generalization of the tensor product of differential graded categories. In particular, there is a multifunctor

$$\boxdot : \mathsf{A}^{\mathsf{u}}_\infty \boxtimes \widehat{\mathbf{dg}\text{-}\mathcal{C}\mathrm{at}} \to \mathsf{A}^{\mathsf{u}}_\infty. \tag{6.0.1}$$

To a unital A_∞-category $\mathcal{A}$ and a differential graded category $\mathcal{C}$ this multifunctor assigns the quiver $\mathcal{A} \boxdot \mathcal{C} \in \operatorname{Ob} \mathscr{Q}$ such that $s(\mathcal{A} \boxdot \mathcal{C}) = s\mathcal{A} \boxtimes \mathcal{C}$. It is equipped with the differential $b^{\mathcal{A}\boxdot\mathcal{C}} : \mathrm{id} \to \mathrm{id} : \mathcal{A} \boxdot \mathcal{C} \to \mathcal{A} \boxdot \mathcal{C}$, specified by its components:

$$b_1^{\mathcal{A}\boxdot\mathcal{C}} = b_1 \boxtimes 1 - 1 \boxtimes d : s\mathcal{A} \boxtimes \mathcal{C} \to s\mathcal{A} \boxtimes \mathcal{C},$$

$$b_n^{\mathcal{A}\boxdot\mathcal{C}} = \bigl[T^n(s\mathcal{A} \boxtimes \mathcal{C}) \xrightarrow{\varkappa} T^n s\mathcal{A} \boxtimes T^n\mathcal{C} \xrightarrow{b_n^{\mathcal{A}} \boxtimes \mu_{\mathcal{C}}^{\mathbf{n}}} s\mathcal{A} \boxtimes \mathcal{C}\bigr], \quad n > 1.$$

Here $\mu_{\mathcal{C}}^{\mathbf{n}}$ is iterated composition in the category $\mathcal{C}$. The differential $b^{\mathcal{A}\boxdot\mathcal{C}}$ turns the quiver $\mathcal{A} \boxdot \mathcal{C}$ into a unital A_∞-category. A unit element for the object (X, U) of $\mathcal{A} \boxdot \mathcal{C}$ can be chosen as ${}_X\mathbf{i}_0^{\mathcal{A}} \otimes 1_U \in s\mathcal{A}(X, X) \otimes \mathcal{C}(U, U)$. To a unital A_∞-functor $f : \boxtimes^{i\in I} Ts\mathcal{A}_i \to Ts\mathcal{B}$ and a differential graded functor $g : \boxtimes^{i\in I}\mathcal{C}_i \to \mathcal{D}$ the action $\boxdot$ assigns the unital A_∞-functor $f \boxdot g : \boxtimes^{i\in I} Ts(\mathcal{A}_i \boxdot \mathcal{C}_i) \to Ts(\mathcal{B} \boxdot \mathcal{D})$ such that

$$\begin{aligned}(f \boxdot g)^\vee = \bigl[&\boxtimes^{i\in I} T(s\mathcal{A}_i \boxtimes \mathcal{C}_i) \xrightarrow{\boxtimes^I \varkappa} \boxtimes^{i\in I}(Ts\mathcal{A}_i \boxtimes T\mathcal{C}_i)\\ &\xrightarrow{\sigma_{(12)}} \bigl(\boxtimes^{i\in I} Ts\mathcal{A}_i\bigr) \boxtimes \bigl(\boxtimes^{i\in I} T\mathcal{C}_i\bigr)\\ &\xrightarrow{\check{f} \boxtimes (\boxtimes^I \mu)} s\mathcal{B} \boxtimes \bigl(\boxtimes^{i\in I}\mathcal{C}_i\bigr) \xrightarrow{1\boxtimes g} s\mathcal{B} \boxtimes \mathcal{D}\bigr].\end{aligned}$$

The action obeys certain associativity constrains similar to the isomorphisms λ^f in a Monoidal category.

7. Pretriangulated A_∞-categories

We introduce the monad of pretriangulated A_∞-categories as a composition of two monads: the A^u_∞-2-monad of shifts and the Maurer–Cartan A^u_∞-2-monad. The first monad adds to an object of an A_∞-category its formal shifts. The second adds iterated cones, solutions of the Maurer–Cartan equation. The commutation morphism between the two monads turns their composition into an A^u_∞-2-monad as well.

7.1. *The multifunctor of shifts*

The category of differential graded categories **dg**-$\mathcal{C}$at equipped with the tensor product $\boxtimes$ is a symmetric monoidal category. The differential graded quiver $\mathcal{Z}$ with $\operatorname{Ob}\mathcal{Z} = \mathbb{Z}$, $\mathcal{Z}(m,n) = \Bbbk[n-m]$ and zero differential has an obvious structure of a $\Bbbk$-linear differential graded category. We equip the object $\mathcal{Z}$ of $\mathscr{D} = (\mathbf{dg}\text{-}\mathcal{C}\mathrm{at}, \boxtimes)$ with an algebra structure, given by multiplication, a differential graded functor

$$\begin{aligned}
&\otimes_{\mathcal{Z}} : \mathcal{Z} \boxtimes \mathcal{Z} \to \mathcal{Z}, \quad m \times n \mapsto m+n,\\
&\otimes_{\mathcal{Z}} = (-1)^{k(m-l)} : (\mathcal{Z} \boxtimes \mathcal{Z})(n \times m, k \times l) = \mathcal{Z}(n,k) \otimes \mathcal{Z}(m,l)\\
&\qquad\qquad\qquad \to \mathcal{Z}(n+m, k+l).
\end{aligned}$$

Clearly, the algebra $(\mathcal{Z}, \otimes_{\mathcal{Z}})$ is unital with unit $\eta_{\mathcal{Z}} : \mathbb{1}_p \to \mathcal{Z}$, $* \mapsto 0$, $\mathrm{id} : \mathbb{1}_p(*,*) = \Bbbk \to \mathcal{Z}(0,0)$. We could say that $\mathcal{Z}$ is a strict monoidal differential graded category.

The algebra $\mathcal{Z}$ together with the action (6.0.1) provide the multifunctor of shifts $-^{[]} : \mathsf{A}^{\mathsf{u}}_\infty \to \mathsf{A}^{\mathsf{u}}_\infty$. It takes a unital A_∞-category $\mathcal{A}$ to the unital A_∞-category $\mathcal{A}^{[]} = \mathcal{A} \boxdot \mathcal{Z}$. Objects of $\mathcal{A} \boxdot \mathcal{Z}$ are pairs $(X,n) = X[n]$, $X \in \operatorname{Ob}\mathcal{A}$, $n \in \mathbb{Z}$. The $\Bbbk$-modules of morphisms are $(\mathcal{A} \boxdot \mathcal{Z})(X[n], Y[m])[1] = \mathcal{A}(X,Y)[1] \otimes \mathcal{Z}(n,m)$. The multifunctor $-^{[]}$ operates on morphisms via the map

$$\begin{aligned}
-^{[]} = \bigl[&\mathsf{A}^{\mathsf{u}}_\infty\bigl((\mathcal{A}_i)_{i\in I}; \mathcal{B}\bigr) \simeq \mathsf{A}^{\mathsf{u}}_\infty\bigl((\mathcal{A}_i)_{i\in I}; \mathcal{B}\bigr) \times \{1\}\\
&\xrightarrow{1\times\widehat{\mathcal{Z}}} \mathsf{A}^{\mathsf{u}}_\infty\bigl((\mathcal{A}_i)_{i\in I}; \mathcal{B}\bigr) \times \widehat{\mathscr{D}}\bigl((\mathcal{Z})_{i\in I}; \mathcal{Z}\bigr)\\
&\xrightarrow{\boxdot} \mathsf{A}^{\mathsf{u}}_\infty\bigl((\mathcal{A}_i \boxdot \mathcal{Z})_{i\in I}; \mathcal{B} \boxdot \mathcal{Z}\bigr)\bigr], \quad f \mapsto f \boxdot \bigl(\otimes^I_{\mathcal{Z}}\bigr),
\end{aligned}$$

where $\widehat{\mathcal{Z}}(1) = \otimes^I_{\mathcal{Z}} \in \mathscr{D}(\boxtimes^I \mathcal{Z}, \mathcal{Z}) = \widehat{\mathscr{D}}((\mathcal{Z})_{i\in I}; \mathcal{Z})$.

The algebra morphism $\eta_{\mathcal{Z}} : \mathbb{1}_p \to \mathcal{Z}$ gives a multinatural transformation of multifunctors $u_{[]} : \mathrm{Id}_{\mathsf{A}^{\mathsf{u}}_\infty} \to -^{[]}$. Since the algebra $\mathcal{Z}$ is not commutative, the multiplication $\otimes_{\mathcal{Z}} : \mathcal{Z} \boxtimes \mathcal{Z} \to \mathcal{Z}$ gives a natural transformation $m_{[]} : -^{[][]} \to -^{[]}$ which is not multinatural. The triple $(-^{[]}, m_{[]}, u_{[]})$ is a monad in $\mathsf{A}^{\mathsf{u}}_\infty$ and an $\mathsf{A}^{\mathsf{u}}_\infty$-monad in $\underline{\mathsf{A}^{\mathsf{u}}_\infty}$, called the monad of shifts.

7.2. Definition. We say that a unital A_∞-category $\mathcal{C}$ is *closed under shifts* if every object $X[n]$ of $\mathcal{C}^{[]}$ is isomorphic in $H^0(\mathcal{C}^{[]})$ to some object $Y[0]$ for $Y \in \operatorname{Ob}\mathcal{C}$.

A unital A_∞-category $\mathcal{C}$ is closed under shifts if and only if the A_∞-functor $u_{[]} : \mathcal{C} \to \mathcal{C}^{[]}$ is an equivalence. If $\mathcal{C}$ is a unital A_∞-category, then $\mathcal{C}^{[]}$ is closed under shifts. Furthermore, the A_∞-functors $u_{[]}, u_{[]}^{[]} : \mathcal{C}^{[]} \to \mathcal{C}^{[]\,[]}$ and $m_{[]} : \mathcal{C}^{[]\,[]} \to \mathcal{C}^{[]}$ are equivalences, quasi-inverse to each other. Moreover, for an arbitrary A_∞-category $\mathcal{C}$ closed under shifts there exists an A_∞-equivalence $U_{[]} = U_{[]}^{\mathcal{C}} : \mathcal{C}^{[]} \to \mathcal{C}$ such that $u_{[]} \cdot U_{[]} = \mathrm{id}_{\mathcal{C}}$. In particular, $U_{[]}$ is quasi-inverse to $u_{[]}$. For an arbitrary unital A_∞-category $\mathcal{A}$, the strict A_∞-functor $A_\infty^u(u_{[]}, \mathcal{C}) = (u_{[]} \boxtimes 1)M : A_\infty^u(\mathcal{A}^{[]}, \mathcal{C}) \to A_\infty^u(\mathcal{A}, \mathcal{C})$ is an A_∞-equivalence which admits a one-sided inverse

$$F_{[]} = \big[A_\infty^u(\mathcal{A}, \mathcal{C}) \xrightarrow{-^{[]}} A_\infty^u(\mathcal{A}^{[]}, \mathcal{C}^{[]}) \xrightarrow{A_\infty^u(\mathcal{A}^{[]}, U_{[]})} A_\infty^u(\mathcal{A}^{[]}, \mathcal{C})\big]$$

(quasi-inverse to $A_\infty^u(u_{[]}, \mathcal{C})$), namely, $F_{[]} \cdot A_\infty^u(u_{[]}, \mathcal{C}) = \mathrm{id}_{A_\infty^u(\mathcal{A}, \mathcal{C})}$.

Adding shifts commutes with taking quotients. More precisely, let $\mathcal{C}$ be an A_∞-category, $\mathcal{B} \subset \mathcal{C}$ a full subcategory. Then the A_∞-categories $\mathsf{q}(\mathcal{C}^{[]}|\mathcal{B}^{[]})$ and $\mathsf{q}(\mathcal{C}|\mathcal{B})^{[]}$ are A_∞-equivalent. Furthermore, if $\mathcal{B}, \mathcal{C}$ are closed under shifts, then so is $\mathsf{q}(\mathcal{C}|\mathcal{B})$.

Let $\mathcal{C}$ be an A_∞-category closed under shifts. Then for an arbitrary A_∞-category $\mathcal{A}$ the A_∞-category $A_\infty(\mathcal{A}, \mathcal{C})$ is closed under shifts. If $\mathcal{A}$ is unital, then the A_∞-category $A_\infty^u(\mathcal{A}, \mathcal{C})$ is closed under shifts as well.

7.3. *The Maurer–Cartan monad*

Let I be the set $\mathbf{m}$ for some $m \in \mathbb{Z}_{\geqslant 0}$. We turn I into an A_∞-category with the set of objects $\mathbf{m}$. The only non-zero graded modules of morphisms are $I(i, i+1) = \Bbbk[-1]$ concentrated in degree 1. Thus, the compositions b_n in this A_∞-category vanish for all $n \geqslant 1$.

Let $\mathcal{C}$ be an A_∞-category. The A_∞-category $\mathcal{C}^{\mathsf{mc}}$ of bounded complexes in $\mathcal{C}$ is defined as follows. Objects of $\mathcal{C}^{\mathsf{mc}}$ are A_∞-functors $X : I \to \mathcal{C}$, $\mathrm{Ob}\, X : i \mapsto X_i$, where I is some set $\mathbf{m}$ for $m \in \mathbb{Z}_{\geqslant 0}$. If $I = \varnothing$, the A_∞-functor $\varnothing \to \mathcal{C}$ gives the zero object of $\mathcal{C}^{\mathsf{mc}}$. Besides X_i the A_∞-functor X is determined by $\Bbbk$-linear maps of degree 0 for all $i < j$, $i, j \in I$,

$$X_{ij} : \Bbbk = sI(i, i+1) \otimes \cdots \otimes sI(j-1, j) \to s\mathcal{C}(X_i, X_j),$$

that is, by elements $x_{ij} \in s\mathcal{C}(X_i, X_j)$ of degree 0. They give an A_∞-functor if and only if the Maurer–Cartan equation holds for all $i < j$, $i, j \in I$:

$$\sum_{i<k_1<\cdots<k_{m-1}<j}^{m>0} (X_{ik_1} \otimes X_{k_1k_2} \otimes \cdots \otimes X_{k_{m-1}j}) b_m^{\mathcal{C}} = 0 :$$

$$\Bbbk = \bigotimes_{k=i}^{j-1} sI(k, k+1) \to s\mathcal{C}(X_i, X_j).$$

The graded $\Bbbk$-module of morphisms between objects $X : I \to \mathcal{C}$ and $Y : J \to \mathcal{C}$ of $\mathcal{C}^{\mathsf{mc}}$ is defined as

$$s\mathcal{C}^{\mathsf{mc}}(X, Y) = \prod_{i \in I, j \in J} s\mathcal{C}(X_i, Y_j).$$

An element r of $s\mathcal{C}^{\mathsf{mc}}(X, Y)$ is viewed as an $I \times J$-matrix $r = (r_{ij})^{i\in I}_{j\in J}$. The differential b^{mc} in $\mathcal{C}^{\mathsf{mc}}$ is given by its components:

$$\begin{aligned} &b^{\mathsf{mc}}_n : s\mathcal{C}^{\mathsf{mc}}\bigl(X^0, X^1\bigr) \otimes \cdots \otimes s\mathcal{C}^{\mathsf{mc}}\bigl(X^{n-1}, X^n\bigr) \to s\mathcal{C}^{\mathsf{mc}}\bigl(X^0, X^n\bigr),\\ &\bigl(r^1 \otimes \cdots \otimes r^n\bigr)b^{\mathsf{mc}}_n\\ &\quad = \sum_{t_0,\dots,t_n\geqslant 0} \bigl[(X^0)^{\otimes t_0} \otimes r^1 \otimes (X^1)^{\otimes t_1} \otimes \cdots \otimes r^n \otimes (X^n)^{\otimes t_n}\bigr] b^{\mathcal{C}}_{t_0+\cdots+t_n+n}, \end{aligned}$$

where we adopt the following matrix notations: the tensor product $A \otimes B$ of two matrices $A = (A_{kl})^{k\in K}_{l\in L}$, $B = (B_{lm})^{l\in L}_{m\in M}$ means $(A \otimes B)_{km} = \sum_{l\in L} A_{kl} \otimes B_{lm}$.

Let $f : \mathcal{A} \to \mathcal{B}$ be an A_∞-functor. It gives rise to an A_∞-functor $f^{\mathsf{mc}} : \mathcal{A}^{\mathsf{mc}} \to \mathcal{B}^{\mathsf{mc}}$. An object $X : I \to \mathcal{A}$ of $\mathcal{A}^{\mathsf{mc}}$ is mapped by f^{mc} to the object $Xf : I \to \mathcal{B}$ of $\mathcal{B}^{\mathsf{mc}}$. It is represented by the matrix $Xf^{\mathsf{mc}} = \sum_{n\geqslant 1}(X^{\otimes n})f_n = \sum_{1\leqslant n<|I|}(X^{\otimes n})f_n$. The components of f^{mc} are

$$\begin{aligned} &f^{\mathsf{mc}}_n : s\mathcal{A}^{\mathsf{mc}}\bigl(X^0, X^1\bigr) \otimes \cdots \otimes s\mathcal{A}^{\mathsf{mc}}\bigl(X^{n-1}, X^n\bigr) \to s\mathcal{B}^{\mathsf{mc}}\bigl(X^0 f^{\mathsf{mc}}, X^n f^{\mathsf{mc}}\bigr),\\ &\bigl(r^1 \otimes \cdots \otimes r^n\bigr)f^{\mathsf{mc}}_n\\ &\quad = \sum_{t_0,\dots,t_n\geqslant 0} \bigl[\bigl(X^0\bigr)^{\otimes t_0} \otimes r^1 \otimes \bigl(X^1\bigr)^{\otimes t_1} \otimes \cdots \otimes r^n \otimes \bigl(X^n\bigr)^{\otimes t_n}\bigr] f_{t_0+\cdots+t_n+n}. \end{aligned} \tag{7.3.1}$$

Let $p : f \to g : \mathcal{A} \to \mathcal{B}$ be an A_∞-transformation. It gives rise to an A_∞-transformation $p^{\mathsf{mc}} : f^{\mathsf{mc}} \to g^{\mathsf{mc}} : \mathcal{A}^{\mathsf{mc}} \to \mathcal{B}^{\mathsf{mc}}$ specified by its components:

$$\begin{aligned} &r^{\mathsf{mc}}_n : s\mathcal{A}^{\mathsf{mc}}\bigl(X^0, X^1\bigr) \otimes \cdots \otimes s\mathcal{B}^{\mathsf{mc}}\bigl(X^{n-1}, X^n\bigr) \to s\mathcal{B}^{\mathsf{mc}}\bigl(X^0 f^{\mathsf{mc}}, X^n g^{\mathsf{mc}}\bigr),\\ &\bigl(r^1 \otimes \cdots \otimes r^n\bigr)p^{\mathsf{mc}}_n\\ &\quad = \sum_{t_0,\dots,t_n\geqslant 0} \bigl[\bigl(X^0\bigr)^{\otimes t_0} \otimes r^1 \otimes \bigl(X^1\bigr)^{\otimes t_1} \otimes \cdots \otimes r^n \otimes \bigl(X^n\bigr)^{\otimes t_n}\bigr] p_{t_0+\cdots+t_n+n}. \end{aligned} \tag{7.3.2}$$

The correspondences $f \mapsto f^{\mathsf{mc}}$, $p \mapsto p^{\mathsf{mc}}$ given by (7.3.1) and (7.3.2) define a strict A_∞-functor $-^{\mathsf{mc}} : A_\infty(\mathcal{A}, \mathcal{B}) \to A_\infty(\mathcal{A}^{\mathsf{mc}}, \mathcal{B}^{\mathsf{mc}})$. If $\mathcal{C}$ is a unital A_∞-category, then $\mathcal{C}^{\mathsf{mc}}$ is unital with the unit transformation $\mathbf{i}^{(\mathcal{C}^{\mathsf{mc}})} = (\mathbf{i}^{\mathcal{C}})^{\mathsf{mc}}$. For a unital A_∞-functor $f : \mathcal{A} \to \mathcal{B}$ the A_∞-functor $f^{\mathsf{mc}} : \mathcal{A}^{\mathsf{mc}} \to \mathcal{B}^{\mathsf{mc}}$ is unital as well. If $\mathcal{A}$, $\mathcal{B}$ are unital A_∞-categories, then the A_∞-functor $-^{\mathsf{mc}} : A^u_\infty(\mathcal{A}, \mathcal{B}) \to A^u_\infty(\mathcal{A}^{\mathsf{mc}}, \mathcal{B}^{\mathsf{mc}})$ is unital. The Maurer–Cartan construction gives an example of an A_∞-2-functor $-^{\mathsf{mc}} : A_\infty \to A_\infty$ which restricts to an A^u_∞-2-functor $-^{\mathsf{mc}} : A^u_\infty \to A^u_\infty$.

Denote by $u_{\mathsf{mc}} = u^{\mathcal{A}}_{\mathsf{mc}} : \mathcal{A} \to \mathcal{A}^{\mathsf{mc}}$ the strict A_∞-functor $X \mapsto (\mathbf{1} \ni 1 \mapsto X, 0)$, with the first component given by $(u_{\mathsf{mc}})_1 = \mathrm{id} : s\mathcal{A}(X, Y) \to s\mathcal{A}^{\mathsf{mc}}((1 \mapsto X, 0), (1 \mapsto Y, 0)) = s\mathcal{A}(X, Y)$. If $\mathcal{A}$ is unital, then the A_∞-functor $u^{\mathcal{A}}_{\mathsf{mc}}$ is unital. The collection of $u^{\mathcal{A}}_{\mathsf{mc}}$ defines a strict A_∞-2-transformation $u_{\mathsf{mc}} : \mathrm{Id} \to (-)^{\mathsf{mc}}$.

Let $\mathcal{A}$ be an A_∞-category. We want to define a strict A_∞-functor

$$m^{\mathcal{A}}_{\mathsf{mc}} = \mathrm{Tot}_{\mathcal{A}} : \bigl(\mathcal{A}^{\mathsf{mc}}\bigr)^{\mathsf{mc}} \to \mathcal{A}^{\mathsf{mc}}.$$

Objects of $(\mathcal{A}^{\mathsf{mc}})^{\mathsf{mc}}$ are A_∞-functors

$$X : I \to \mathcal{A}^{\mathsf{mc}}, \quad i \mapsto \bigl(X^i : J^i \to \mathcal{A}\bigr),$$

$$x^{ii'} \in \big[s\mathcal{A}^{\mathsf{mc}}\big(X^i, X^{i'}\big)\big]^0 \quad \text{for } i, i' \in I,\ i < i'.$$

For each $i \in I$ the A_∞-functor $X^i : J^i \to \mathcal{A}$ consists of $\operatorname{Ob} X^i : J^i \ni j \mapsto X^i_j \in \operatorname{Ob}\mathcal{A}$ and of $x^{ii} \overset{\text{def}}{=} x^i = (x^i_{jj'})_{j,j' \in J^i}$. Let K denote the partition $\bigsqcup_{i\in I} J^i = J^0 \sqcup \dots \sqcup J^n$. The A_∞-functors X with fixed sets I, $(J^i)_{i\in I}$ are in bijection with A_∞-functors $\tilde{X} : K \to \mathcal{A}$. The map $\operatorname{Ob}\operatorname{Tot}_{\mathcal{A}} : \operatorname{Ob}(\mathcal{A}^{\mathsf{mc}})^{\mathsf{mc}} \to \operatorname{Ob}\mathcal{A}^{\mathsf{mc}}$, $X \mapsto \tilde{X}$ is given by the assignment $J^i \ni j \mapsto X^i_j$, $\tilde{x}^{ii'}_{jj'} = x^{ii'}_{jj'}$, where $j \in J^i$, $j' \in J^{i'}$, $i \leqslant i'$.

Given two objects of $(\mathcal{A}^{\mathsf{mc}})^{\mathsf{mc}}$, $X : I \to \mathcal{A}^{\mathsf{mc}}$, $i \mapsto (X^i : J^i \to \mathcal{A}, j \mapsto X^i_j)$, and $Y : L \to \mathcal{A}^{\mathsf{mc}}$, $l \mapsto (Y^l : M^l \to \mathcal{A}, m \mapsto Y^l_m)$, we describe the $\Bbbk$-module of morphisms between them:

$$\begin{aligned}\big(\mathcal{A}^{\mathsf{mc}}\big)^{\mathsf{mc}}(X, Y) &= \prod_{i\in I, l\in L} \mathcal{A}^{\mathsf{mc}}\big(X^i, Y^l\big) = \prod_{i\in I, l\in L} \prod_{j\in J^i, m\in M^l} \mathcal{A}\big(X^i_j, Y^l_m\big) \\ &\simeq \prod_{k\in K, n\in N} \mathcal{A}\big(\tilde{X}^k, \tilde{Y}^n\big) = \mathcal{A}^{\mathsf{mc}}(\tilde{X}, \tilde{Y}),\end{aligned}$$

where $K = \bigsqcup_{i\in I} J^i$, $N = \bigsqcup_{l\in L} M^l$ and $\tilde{X} = X\operatorname{Tot}$, $\tilde{Y} = Y\operatorname{Tot}$.

We make $\operatorname{Tot}_{\mathcal{A}}$ into a strict A_∞-functor setting $(\operatorname{Tot}_{\mathcal{A}})_1$ to be the above isomorphism and $(\operatorname{Tot}_{\mathcal{A}})_k = 0$, $k > 1$. If $\mathcal{A}$ is unital, then the A_∞-functor $\operatorname{Tot}_{\mathcal{A}} : \mathcal{A}^{\mathsf{mcmc}} \to \mathcal{A}^{\mathsf{mc}}$ is unital. When $\mathcal{A}$ runs over all A_∞-categories, the collection of functors $\operatorname{Tot}_{\mathcal{A}}$ determines an A_∞-2-transformation $m_{\mathsf{mc}} = \operatorname{Tot} : ((-)^{\mathsf{mc}})^{\mathsf{mc}} \to (-)^{\mathsf{mc}}$. The triple $(-^{\mathsf{mc}}, \operatorname{Tot}, u_{\mathsf{mc}})$ is an A_∞-2-monad, which restricts to an A^u_∞-2-monad.

7.4. Definition. We say that a unital A_∞-category $\mathcal{C}$ is mc-*closed* if every object X of $\mathcal{C}^{\mathsf{mc}}$ is isomorphic in $H^0(\mathcal{C}^{\mathsf{mc}})$ to Yu_{mc} for some object $Y \in \operatorname{Ob}\mathcal{C}$.

7.5. Proposition. *(See [3].) Let $\mathcal{C}$ be a unital A_∞-category. Then the following conditions are equivalent*:

(i) *$\mathcal{C}$ contains a contractible object, and each object $\big(W : \mathbf{2} \to \mathcal{C}, \left(\begin{smallmatrix} 0 & f \\ 0 & 0 \end{smallmatrix}\right)\big)$ of $\mathcal{C}^{\mathsf{mc}}$ is isomorphic in $H^0(\mathcal{C}^{\mathsf{mc}})$ to Cu_{mc} for some object $C \in \operatorname{Ob}\mathcal{C}$*;

(ii) *$\mathcal{C}$ is mc-closed*;

(iii) *the A_∞-functor $u_{\mathsf{mc}} : \mathcal{C} \to \mathcal{C}^{\mathsf{mc}}$ is an equivalence.*

Equivalence of (i) and (ii) can be put in words as follows: any solution to the Maurer–Cartan equation is an iterated cone.

If $\mathcal{C}$ is a unital A_∞-category, then $\mathcal{C}^{\mathsf{mc}}$ is mc-closed. Furthermore, the A_∞-functors $u_{\mathsf{mc}}, u^{\mathsf{mc}}_{\mathsf{mc}} : \mathcal{C}^{\mathsf{mc}} \to \mathcal{C}^{\mathsf{mc}\,\mathsf{mc}}$ and $m_{\mathsf{mc}} : \mathcal{C}^{\mathsf{mc}\,\mathsf{mc}} \to \mathcal{C}^{\mathsf{mc}}$ are equivalences, quasi-inverse to each other. Moreover, for an arbitrary mc-closed A_∞-category $\mathcal{C}$ there exists an A_∞-equivalence $U_{\mathsf{mc}} = U^{\mathcal{C}}_{\mathsf{mc}} : \mathcal{C}^{\mathsf{mc}} \to \mathcal{C}$ such that $u_{\mathsf{mc}} \cdot U_{\mathsf{mc}} = \mathrm{id}_{\mathcal{C}}$. In particular, U_{mc} is quasi-inverse to u_{mc}. For an arbitrary unital A_∞-category $\mathcal{A}$, the strict A_∞-functor $A^u_\infty(u_{\mathsf{mc}}, \mathcal{C}) = (u_{\mathsf{mc}} \boxtimes 1)M : A^u_\infty(\mathcal{A}^{\mathsf{mc}}, \mathcal{C}) \to A^u_\infty(\mathcal{A}, \mathcal{C})$ is an A_∞-equivalence which admits a one-sided inverse

$$F_{\mathsf{mc}} = \big[A^u_\infty(\mathcal{A}, \mathcal{C}) \xrightarrow{-^{\mathsf{mc}}} A^u_\infty\big(\mathcal{A}^{\mathsf{mc}}, \mathcal{C}^{\mathsf{mc}}\big) \xrightarrow{A^u_\infty(\mathcal{A}^{\mathsf{mc}}, U_{\mathsf{mc}})} A^u_\infty(\mathcal{A}^{\mathsf{mc}}, \mathcal{C})\big]$$

(quasi-inverse to $A^u_\infty(u_{\mathsf{mc}}, \mathcal{C})$), namely, $F_{\mathsf{mc}} \cdot A^u_\infty(u_{\mathsf{mc}}, \mathcal{C}) = \mathrm{id}_{A^u_\infty(\mathcal{A},\mathcal{C})}$.

Adding iterated cones commutes with taking quotients. This means that if $\mathcal{C}$ is an A_∞-category and $\mathcal{B} \subset \mathcal{C}$ is a full subcategory, then the A_∞-categories $\mathsf{q}(\mathcal{C}^{\mathsf{mc}}|\mathcal{B}^{\mathsf{mc}})$ and $\mathsf{q}(\mathcal{C}|\mathcal{B})^{\mathsf{mc}}$ are A_∞-equivalent. Furthermore, if $\mathcal{B}, \mathcal{C}$ are mc-closed, then so is $\mathsf{q}(\mathcal{C}|\mathcal{B})$.

Let $\mathcal{C}$ be an mc-closed A_∞-category. Then for an arbitrary A_∞-category $\mathcal{A}$ the $\mathcal{A}_\infty$-category $A_\infty(\mathcal{A}, \mathcal{C})$ is mc-closed. If $\mathcal{A}$ is unital, then the $\mathcal{A}$-category $A^u_\infty(\mathcal{A}, \mathcal{C})$ is mc-closed as well.

7.6. *The monad of pretriangulated A_∞-categories*

A commutation morphism $\mathfrak{c} = \mathfrak{c}_{\mathcal{C}} : \mathcal{C}^{\mathsf{mc}[]} \to \mathcal{C}^{[]\mathsf{mc}}$ between the A_∞-2-monads mc and $[]$ is defined as follows. The strict A_∞-functor $\mathfrak{c}_{\mathcal{C}}$ takes an object $X[n]$ of $\mathcal{C}^{\mathsf{mc}[]}$, where $X : I \to \mathcal{C}$ is an A_∞-functor, to the object $X' = X[n]\mathfrak{c} : I \to \mathcal{C}^{[]}$, $i \mapsto X_i[n]$, $x'_{ij} = x_{ij} \in s\mathcal{C}^{[]}(X_i[n], X_j[n]) = s\mathcal{C}(X_i, X_j)$. The first component of $\mathfrak{c}$ is

$$\begin{aligned}
\mathfrak{c}_1 \ : s\mathcal{C}^{\mathsf{mc}[]}\bigl(X[n], Y[m]\bigr) &= s\mathcal{C}^{\mathsf{mc}}(X, Y)[m-n] \\
&= \Bigl\{ \prod_{i\in I, j\in J} s\mathcal{C}(X_i, Y_j) \Bigr\} [m-n] \\
&= \prod_{i\in I, j\in J} \bigl\{ s\mathcal{C}(X_i, Y_j)[m-n] \bigr\} = \prod_{i\in I, j\in J} s\mathcal{C}^{[]}\bigl(X_i[n], Y_j[m]\bigr) \\
&= s\mathcal{C}^{[]\mathsf{mc}}\bigl(X[n]\mathfrak{c}, Y[m]\mathfrak{c}\bigr).
\end{aligned}$$

The A_∞-functor $\mathfrak{c}_{\mathcal{C}} : \mathcal{C}^{\mathsf{mc}[]} \to \mathcal{C}^{[]\mathsf{mc}}$ is unital if $\mathcal{C}$ is. When $\mathcal{C}$ runs over all A_∞-categories, the collection of A_∞-functors $\mathfrak{c}_{\mathcal{C}} : \mathcal{C}^{\mathsf{mc}[]} \to \mathcal{C}^{[]\mathsf{mc}}$ determines an A_∞-2-transformation.

The A_∞-2-functor $\mathsf{tr} : A_\infty \to A_\infty$ (respectively A^u_∞-2-functor $\mathsf{tr} : A^u_\infty \to A^u_\infty$), $\mathcal{C} \mapsto \mathcal{C}^{\mathsf{tr}} = \mathcal{C}^{[]\mathsf{mc}}$, equipped with the unit

$$u_{\mathsf{tr}} = \bigl(\mathrm{Id} \xrightarrow{u_{[]}} \_^{[]} \xrightarrow{u_{\mathsf{mc}}} \_^{[]\mathsf{mc}}\bigr)$$

and with the multiplication

$$m_{\mathsf{tr}} = \bigl(\_^{[]\mathsf{mc}[]\mathsf{mc}} \xrightarrow{\mathfrak{c}^{\mathsf{mc}}} \_^{[][]\mathsf{mcmc}} \xrightarrow{m_{\mathsf{mc}}} \_^{[][]\mathsf{mc}} \xrightarrow{m^{\mathsf{mc}}_{[]}} \_^{[]\mathsf{mc}}\bigr)$$

is an A_∞-2-monad (respectively A^u_∞-2-monad).

Let us describe the strict A_∞-functor $m_{\mathsf{tr}} : \mathcal{C}^{\mathsf{tr}\,\mathsf{tr}} \to \mathcal{C}^{\mathsf{tr}}$ explicitly. First we consider its action on objects. An object X of $\mathcal{C}^{\mathsf{tr}\,\mathsf{tr}} = \mathcal{C}^{[]\mathsf{mc}[]\mathsf{mc}}$ is specified by the following data:

- finite linearly ordered sets $I \in \operatorname{Ob}\mathcal{O}$, $J^i \operatorname{Ob}\mathcal{O}$ for every $i \in I$;
- objects X^i_j of $\mathcal{C}$ for all $i \in I$, $j \in J^i$;
- integers m^i, n^i_j for all $i \in I$, $j \in J^i$;
- and the following matrices x, x^{ii}, $i \in I$ of morphisms:

$$\begin{aligned}
&\bigl(X : I \ni i \mapsto \bigl(X^i : J^i \ni j \mapsto X^i_j[n^i_j], x^{ii} = \bigl(x^{ii}_{jj'}\bigr)_{j,j'\in J^i}\bigr)[m^i], \\
&\quad x = \bigl(x^{ii'}\bigr)_{i,i'\in I}\bigr), \\
&\quad x^{ii}_{jj'} \in \mathcal{C}^{[]}\bigl(X^i_j[n^i_j], X^i_{j'}[n^i_{j'}]\bigr)[1]^0 = \mathcal{C}\bigl(X^i_j, X^i_{j'}\bigr)\bigl[n^i_{j'} - n^i_j + 1\bigr]^0, \\
&\quad x^{ii'} \in \mathcal{C}^{[]\mathsf{mc}[]}\bigl(X^i[m^i], X^{i'}[m^{i'}]\bigr)[1]^0 = \mathcal{C}\bigl(X^i, X^{i'}\bigr)\bigl[m^{i'} - m^i + 1\bigr]^0,
\end{aligned}$$

where X^i and X specify A_∞-functors. This means that $j \geqslant j'$ implies $x^{ii}_{jj'} = 0$, $i \geqslant i'$ implies $x^{ii'} = 0$, the Maurer–Cartan equations $\sum_{t \geqslant 0} x^{\otimes t} b_t^{[]\mathsf{mc}[]} = 0$ and $\sum_{t \geqslant 0} (x^{ii})^{\otimes t} b_t^{[]} = 0$ hold for all $i \in I$.

The elements $x^{ii'}$ are identified with matrices $(x^{ii'}_{jj'})^{j' \in J^{i'}}_{j \in J^i}$, where

$$\begin{aligned} x^{ii'}_{jj'} &\in \mathcal{C}^{[]}\big(X^i_j[n^i_j], X^{i'}_{j'}[n^{i'}_{j'}]\big)\big[m^{i'} - m^i + 1\big]^0 \\ &= \mathcal{C}\big(X^i_j, X^{i'}_{j'}\big)\big[n^{i'}_{j'} + m^{i'} - n^i_j - m^i + 1\big]^0. \end{aligned}$$

Write $IJ = \bigsqcup_{i \in I} J^i = \{(i, j) \in I \times \bigcup_{i \in I} J^i \mid j \in J^i\}$, and consider the $IJ \times IJ$-matrix $\tilde{x} = x + \operatorname{diag}(x^{ii})$, so that

$$\begin{aligned} \tilde{x}^{ii'}_{jj'} = x^{ii'}_{jj'} &\in \mathcal{C}^{[][]}\big(X^i_j[n^i_j][m^i], X^{i'}_{j'}[n^{i'}_{j'}][m^{i'}]\big)[1]^0 \\ &= \mathcal{C}\big(X^i_j, X^{i'}_{j'}\big)\big[m^{i'} + n^{i'}_{j'} - m^i - n^i_j + 1\big]^0 \end{aligned}$$

if $i < i'$, or if $i = i'$ and $j < j'$, and otherwise $\tilde{x}^{ii'}_{jj'} = 0$. The object X is taken by m_{tr} to the objects

$$\begin{gathered} \big(Xm_{\mathsf{tr}} : IJ \ni (i, j) \mapsto X^i_j\big[n^i_j + m^i\big],\ \bar{x} = \big(\bar{x}^{ii'}_{jj'}\big)_{(i,j),(i',j') \in IJ}\big), \\ \bar{x}^{ii'}_{jj'} = (-1)^{n^{i'}_{j'}(m^i - m^{i'})} x^{ii'}_{jj'} \in \mathcal{C}^{[]}\big(X^i_j\big[n^i_j + m^i\big], X^{i'}_{j'}\big[n^{i'}_{j'} + m^{i'}\big]\big)[1]^0. \end{gathered}$$

Now we consider the first component of the strict A_∞-functor m_{tr}. Let Y be another object of $\mathcal{C}^{\mathsf{tr}\,\mathsf{tr}}$,

$$\begin{aligned} &\big(Y : K \ni k \mapsto \big(Y^k : L^k \ni l \mapsto Y^k_l\big[p^k_l\big], \\ &\quad y^{kk} = \big(y^{kk}_{ll'}\big)_{l,l' \in L^k}\big)\big[q^k\big], y = \big(y^{kk'}\big)_{k,k' \in K}\big). \end{aligned}$$

A morphism $f \in \mathcal{C}^{\mathsf{tr}\,\mathsf{tr}}(X, Y)[1]$ identifies with the matrix $(f^{ik}_{jl})^{i \in I, k \in K}_{j \in J^i, l \in L^k}$

$$f^{ik}_{jl} \in \mathcal{C}\big(X^i_j, Y^k_l\big)\big[p^k_l + q^k - n^i_j - m^i + 1\big] \simeq \mathcal{C}^{\mathsf{tr}}\big(X^i, Y^k\big)\big[q^k - m^i + 1\big].$$

Write $KL = \bigsqcup_{k \in K} L^k$. The morphism f is mapped by $(m_{\mathsf{tr}})_1$ to the morphism $f(m_{\mathsf{tr}})_1 \in \mathcal{C}^{\mathsf{tr}}(Xm_{\mathsf{tr}}, Ym_{\mathsf{tr}})[1]$, determined by the matrix

$$\begin{aligned} &f(m_{\mathsf{tr}})_1 = \big((-1)^{p^k_l(m^i - q^k)} f^{ik}_{jl}\big)_{(i,j) \in IJ, (k,l) \in KL}, \\ &\quad f^{ik}_{jl} \in \mathcal{C}^{[]}\big(X^i_j\big[n^i_j + m^i\big], Y^k_l\big[p^k_l + q^k\big]\big)[1]. \end{aligned}$$

7.7. Definition. We say that a unital A_∞-category $\mathcal{C}$ is *pretriangulated* if every object X of $\mathcal{C}^{\mathsf{tr}}$ is isomorphic in $H^0(\mathcal{C}^{\mathsf{tr}})$ to Yu_{tr} for some object Y of $\mathcal{C}$.

Another (but similar) notion of a triangulated category due to Kontsevich and Soibelman can be found in [33].

7.8. Proposition. *(See* [3].*) Let $\mathcal{C}$ be a unital A_∞-category. Then the following conditions are equivalent*:

(i) *$\mathcal{C}$ is pretriangulated*;

(ii) *the A_∞-functor $u_{\mathsf{tr}} : \mathcal{C} \to \mathcal{C}^{\mathsf{tr}}$ is an equivalence*;

(iii) *$\mathcal{C}$ is closed under shifts and* mc*-closed*;

(iv) *the A_∞-functors $u_{[]} : \mathcal{C} \to \mathcal{C}^{[]}$ and $u_{\mathsf{mc}} : \mathcal{C} \to \mathcal{C}^{\mathsf{mc}}$ are equivalences.*

Let $\mathcal{C}$ be a unital A_∞-category. Then the A_∞-category $\mathcal{C}^{\mathsf{tr}}$ is pretriangulated, closed under shifts and mc-closed. The A_∞-functors $u_{\mathsf{tr}}, u_{\mathsf{tr}}^{\mathsf{tr}} : \mathcal{C}^{\mathsf{tr}} \to \mathcal{C}^{\mathsf{tr}\,\mathsf{tr}}$ and $m_{\mathsf{tr}} : \mathcal{C}^{\mathsf{tr}\,\mathsf{tr}} \to \mathcal{C}^{\mathsf{tr}}$ are equivalences, quasi-inverse to each other. Moreover, for an arbitrary pretriangulated A_∞-category $\mathcal{C}$ there exists an A_∞-equivalence $U_{\mathsf{tr}} = U_{\mathsf{tr}}^{\mathcal{C}} = (\mathcal{C}^{\mathsf{tr}} \xrightarrow{U_{[]}^{\mathsf{mc}}} \mathcal{C}^{\mathsf{mc}} \xrightarrow{U_{\mathsf{mc}}} \mathcal{C})$ such that $u_{\mathsf{tr}} \cdot U_{\mathsf{tr}} = \mathrm{id}_{\mathcal{C}}$. In particular, U_{tr} is quasi-inverse to u_{tr}. For an arbitrary A_∞-category $\mathcal{A}$ the strict A_∞-functor $A_\infty^u(u_{\mathsf{tr}}, \mathcal{C}) = (u_{\mathsf{tr}} \boxtimes 1)M : A_\infty^u(\mathcal{A}^{\mathsf{tr}}, \mathcal{C}) \to A_\infty^u(\mathcal{A}, \mathcal{C})$ is an A_∞-equivalence which admits a one-sided inverse

$$F_{\mathsf{tr}} \overset{\text{def}}{=} \bigl[A_\infty^u(\mathcal{A}, \mathcal{C}) \xrightarrow{-^{\mathsf{tr}}} A_\infty^u(\mathcal{A}^{\mathsf{tr}}, \mathcal{C}^{\mathsf{tr}}) \xrightarrow{A_\infty^u(\mathcal{A}^{\mathsf{tr}}, U_{\mathsf{tr}})} A_\infty^u(\mathcal{A}^{\mathsf{tr}}, \mathcal{C})\bigr] = F_{[]} \cdot F_{\mathsf{mc}};$$

indeed, $F_{\mathsf{tr}} \cdot A_\infty^u(u_{\mathsf{tr}}, \mathcal{C}) = \mathrm{id}_{A_\infty^u(\mathcal{A},\mathcal{C})}$.

Let $\mathcal{A}_i$, $\mathcal{B}$ be unital A_∞-categories, $i \in I$. Suppose that $\mathcal{B}$ is pretriangulated. Then the A_∞-category $\underline{\mathsf{A}_\infty^{\mathsf{u}}}((\mathcal{A}_i)_{i \in I}; \mathcal{B})$ is pretriangulated.

Let $\mathsf{A}_\infty^{\mathsf{tr}}$ denote the full submulticategory of $\mathsf{A}_\infty^{\mathsf{u}}$, whose objects are pretriangulated A_∞-categories. The multicategory $\mathsf{A}_\infty^{\mathsf{tr}}$ is closed with the inner object of morphism $\underline{\mathsf{A}_\infty^{\mathsf{tr}}}((\mathcal{A}_i)_{i \in I}; \mathcal{B}) = \underline{\mathsf{A}_\infty^{\mathsf{u}}}((\mathcal{A}_i)_{i \in I}; \mathcal{B})$ and the evaluation

$$\mathrm{ev}^{\mathsf{A}_\infty^{\mathsf{tr}}} = \mathrm{ev}^{\mathsf{A}_\infty^{\mathsf{u}}} : (\mathcal{A}_i)_{i \in I}, \underline{\mathsf{A}_\infty^{\mathsf{tr}}}\bigl((\mathcal{A}_i)_{i \in I}; \mathcal{B}\bigr) \to \mathcal{B}.$$

Since the A_∞-2-functors $-^{[]}$ and $-^{\mathsf{mc}}$ commute with taking quotients, so does their composition, the A_∞-2-functor $-^{\mathsf{tr}}$. Furthermore, the quotient of a pretriangulated A_∞-category over a full pretriangulated A_∞-subcategory is pretriangulated as well.

7.9. *Triangulated categories*

We define triangulated categories with weak versions of suspension and desuspension functors. These are presented in the 2-category setting.

Let $\mathfrak{C}$ be a 2-category, $\mathcal{C}$, $\mathcal{D}$ objects of $\mathfrak{C}$. An *adjunction* from $\mathcal{C}$ to $\mathcal{D}$ is a quadruple $(F, U, \eta, \varepsilon)$, where $F : \mathcal{C} \to \mathcal{D}$ and $U : \mathcal{D} \to \mathcal{C}$ are 1-morphisms, $\eta : \mathrm{Id}_{\mathcal{C}} \to FU$ and $\varepsilon : UF \to \mathrm{Id}_{\mathcal{D}}$ are 2-morphisms such that the following equations hold:

$$\bigl(F \xrightarrow{\eta F} FUF \xrightarrow{F\varepsilon} F\bigr) = \mathrm{id}_F,$$
$$\bigl(U \xrightarrow{U\eta} UFU \xrightarrow{\varepsilon U} U\bigr) = \mathrm{id}_U.$$

Let $\mathcal{C}$ be an object of $\mathfrak{C}$. The category $\mathfrak{C}(\mathcal{C}, \mathcal{C}) = \mathrm{End}(\mathcal{C})$ is a strict monoidal category with the tensor product given by composition of 1-morphisms. The set of integers $\mathbb{Z}$ can be viewed as a discrete category. It is a strict monoidal category with the tensor product given by addition.

7.10. Definition. A *translation structure* on $\mathcal{C}$ is a monoidal functor $(\Sigma,\varsigma):\mathbb{Z}\to\operatorname{End}(\mathcal{C})$. More specifically, a translation structure on $\mathcal{C}$ consists of the following data: for every $n\in\mathbb{Z}$ a 1-morphism $\Sigma^n=\Sigma(n):\mathcal{C}\to\mathcal{C}$; for each pair $m,n\in\mathbb{Z}$ a 2-isomorphism $\varsigma_{m,n}:\Sigma^m\Sigma^n\xrightarrow{\sim}\Sigma^{m+n}$; a 2-isomorphism $\varsigma_0:\mathrm{Id}_{\mathcal{C}}\xrightarrow{\sim}\Sigma^0$. These data must satisfy the following coherence relations:

(i) cocycle condition: for $k,m,n\in\mathbb{Z}$

$$\begin{aligned}&\bigl(\Sigma^k\Sigma^m\Sigma^n\xrightarrow{\Sigma^k\varsigma_{m,n}}\Sigma^k\Sigma^{m+n}\xrightarrow{\varsigma_{k,m+n}}\Sigma^{k+m+n}\bigr)\\ &\quad=\bigl(\Sigma^k\Sigma^m\Sigma^n\xrightarrow{\varsigma_{k,m}\Sigma^n}\Sigma^{k+m}\Sigma^n\xrightarrow{\varsigma_{k+m,n}}\Sigma^{k+m+n}\bigr);\end{aligned}$$

(ii) for every $n\in\mathbb{Z}$:

$$\begin{aligned}&\bigl(\Sigma^n\xrightarrow{\varsigma_0\Sigma^n}\Sigma^0\Sigma^n\xrightarrow{\varsigma_{0,n}}\Sigma^n\bigr)=\mathrm{id}_{\Sigma^n},\\ &\bigl(\Sigma^n\xrightarrow{\Sigma^n\varsigma_0}\Sigma^n\Sigma^0\xrightarrow{\varsigma_{n,0}}\Sigma^n\bigr)=\mathrm{id}_{\Sigma^n}.\end{aligned}$$

Let us define a new 2-category $\operatorname{Trans}\mathfrak{C}$. An object of this category is an object of $\mathfrak{C}$ with a translation structure on it. A 1-morphism from $\mathcal{C}$ to $\mathcal{D}$ consists of the following data: a 1-morphism $F:\mathcal{C}\to\mathcal{D}$; for every $n\in\mathbb{Z}$ a 2-isomorphism $\phi_n:\Sigma^nF\xrightarrow{\sim}F\Sigma^n$ such that the following diagram commutes for each pair $m,n\in\mathbb{Z}$:

$$\begin{tikzcd}
\Sigma^m\Sigma^nF \arrow[r, "\Sigma^m\phi_n"] \arrow[d, "\varsigma_{m,n}F"'] & \Sigma^mF\Sigma^n \arrow[r, "\phi_m\Sigma^n"] & F\Sigma^m\Sigma^n \arrow[d, "F\varsigma_{m,n}"] \\
\Sigma^{m+n}F \arrow[rr, "\phi_{m+n}"] & & F\Sigma^{m+n}
\end{tikzcd}$$

and the following equation holds:

$$\bigl(F\xrightarrow{\varsigma_0F}\Sigma^0F\xrightarrow{\phi_0}F\Sigma^0\bigr)=\bigl(F\xrightarrow{F\varsigma_0}F\Sigma^0\bigr).$$

A 2-morphism $\nu:(F,\phi_n)\to(G,\psi_n):\mathcal{C}\to\mathcal{D}$ is a 2-morphism $\nu:F\to G$ such that the following diagram commutes for every $n\in\mathbb{Z}$:

$$\begin{tikzcd}
\Sigma^nF \arrow[r, "\Sigma^n\nu"] \arrow[d, "\phi_n"'] & \Sigma^nG \arrow[d, "\psi_n"] \\
F\Sigma^n \arrow[r, "\nu\Sigma^n"] & G\Sigma^n
\end{tikzcd}$$

Let $\mathcal{C}$, $\mathcal{D}$ be objects of $\mathfrak{C}$, (F,U,η,ε) an adjunction from $\mathcal{C}$ to $\mathcal{D}$. The correspondence $T\mapsto FTU$ extends to a lax monoidal functor $\Gamma:\operatorname{End}(\mathcal{D})\to\operatorname{End}(\mathcal{C})$. In particular, if (F,U,η,ε) is an adjunction-equivalence from $\mathcal{C}$ to $\mathcal{D}$ and $\mathcal{D}$ is equipped with a translation structure, then the composite functor $\Sigma_{\mathcal{C}}=(\mathbb{Z}\xrightarrow{\Sigma_{\mathcal{D}}}\operatorname{End}(\mathcal{D})\xrightarrow{\Gamma}\operatorname{End}(\mathcal{C}))$ is monoidal and defines a translation structure on $\mathcal{C}$. A given equivalence $F:\mathcal{C}\to\mathcal{D}$ can be completed to an adjunction in a non-unique way, but $\Sigma_{\mathcal{C}}$ depends functorially on the choice of adjunction data.

Let $A_\infty^{[]\text{-closed}}$ denote the full 2-subcategory of the 2-category $\overline{A_\infty^u}$, whose objects are unital A_∞-categories closed under shifts.

7.11. Proposition. *(See [3].) The embedding* $A_\infty^{[]\text{-closed}} \hookrightarrow \overline{A_\infty^u}$ *lifts naturally to a* 2*-functor* $\tilde{-} : A_\infty^{[]\text{-closed}} \to \operatorname{Trans} \overline{A_\infty^u}$ *that is the identity on* 2*-morphisms.*

As a corollary, we obtain a composite 2-functor

$$\tilde{H}^0 = \bigl(A_\infty^{[]\text{-closed}} \xrightarrow{\tilde{-}} \operatorname{Trans} \overline{A_\infty^u} \xrightarrow{\operatorname{Trans} H^0} \operatorname{Trans} \Bbbk\text{-}\mathcal{C}\mathrm{at}\bigr).$$

Given an A_∞-category $\mathcal{B}$ closed under shifts, we denote by $\tilde{H}^0(\mathcal{B})$ the corresponding $\Bbbk$-linear category $H^0(\mathcal{B})$ equipped with the natural translation structure. Similarly, an A_∞-functor $f : \mathcal{B} \to \mathcal{C}$ between such A_∞-categories induces a functor $\tilde{H}^0(f) = (H^0(f), \phi_n) : \tilde{H}^0(\mathcal{B}) \to \tilde{H}^0(\mathcal{C})$ between categories with translation structure.

Let from now on $\mathfrak{C}$ be the 2-category of additive $\Bbbk$-linear categories, $\Bbbk$-linear functors, and natural transformations. Let $\mathcal{C}$ be an additive $\Bbbk$-linear category with a translation structure $(\Sigma^n, \varsigma_{m,n}, \varsigma_0)$ on it. We put $\Sigma = \Sigma^1$ for brevity.

7.12. Definition. A *triangle* (X, Y, Z, a, b, c) in $\mathcal{C}$ is the sequence of maps

$$X \xrightarrow{a} Y \xrightarrow{b} Z \xrightarrow{c} X\Sigma.$$

A *morphism of triangles*

$$(f, g, h) : (X, Y, Z, a, b, c) \to (X', Y', Z', a', b', c')$$

is a commutative diagram

$$\begin{tikzcd}
X \arrow[r, "a"] \arrow[d, "f"'] & Y \arrow[r, "b"] \arrow[d, "g"'] & Z \arrow[r, "c"] \arrow[d, "h"'] & X\Sigma \arrow[d, "f\Sigma"] \\
X' \arrow[r, "a'"] & Y' \arrow[r, "b'"] & Z' \arrow[r, "c'"] & X'\Sigma
\end{tikzcd}$$

A *system of triangles* in $\mathcal{C}$ is a collection of triangles, called *distinguished triangles*, which satisfies the axiom

(TR0) Each triangle isomorphic to a distinguished triangle is distinguished. Each morphism $u : X \to Y$ is contained in a distinguished triangle $(X, Y, *, u, *, *)$.

Let $\mathcal{C}, \mathcal{D}$ be categories equipped with a system of triangles. A functor $(F, \phi_n) : \mathcal{C} \to \mathcal{D}$ between categories with translation structure is called *triangulated* if for every distinguished triangle

$$X \xrightarrow{u} Y \xrightarrow{v} Z \xrightarrow{w} X\Sigma$$

in $\mathcal{C}$, the triangle

$$XF \xrightarrow{uF} YF \xrightarrow{vF} ZF \xrightarrow{(wF)\phi_1} XF\Sigma$$

is distinguished in $\mathcal{D}$.

Note that a functor $(F, \phi_n) : \mathcal{C} \to \mathcal{D}$ between categories with translation structure maps isomorphic triangles to isomorphic ones.

Let $\mathcal{B}$ be a pretriangulated A_∞-category. Let us construct a system of triangles in $\tilde{H}^0(\mathcal{B})$. Let $\mathcal{A}_2$ be the A_∞-category with two objects $\underline{1}$ and $\underline{2}$ and with $\mathcal{A}_2(\underline{1}, \underline{2}) = \Bbbk\{e_{12}\} \simeq \Bbbk$,

$\deg e_{12} = 0$. The other three $\Bbbk$-modules $\mathcal{A}_2(*, *)$ vanish. All operations $b_n^{\mathcal{A}_2}$, $n \geqslant 1$ vanish, since $T^{\geqslant 1}s\mathcal{A}_2 = s\mathcal{A}_2$ is concentrated in degree -1. Denote by $\mathcal{D}_2$ the differential graded category $\mathcal{A}_2^{\mathsf{su}}$ (with zero differential). Consider the object $\operatorname{Cone}(e_{12}) = \bigl(1 \mapsto \underline{1}[1], 2 \mapsto \underline{2}[0], \bigl(\begin{smallmatrix} 0 & e_{12} \\ 0 & 0 \end{smallmatrix}\bigr)\bigr)$ of $\mathcal{D}_2^{\mathsf{tr}}$. Denote by i_2 and j_1 the cycles $i_2 = (0, 1) \in \mathcal{D}_2^{\mathsf{tr}}(\underline{2}, \operatorname{Cone}(e_{12}))$ and $j_X = \bigl(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\bigr) \in \mathcal{D}_2^{\mathsf{tr}}(\operatorname{Cone}(e_{12}), \underline{1}[1])$, respectively. Take the standard distinguished triangle in $H^0(\mathcal{D}_2^{\mathsf{tr}})$:

$$\triangle = \bigl(\underline{1} \xrightarrow{e_{12}} \underline{2} \xrightarrow{i_2} \operatorname{Cone}(e_{12}) \xrightarrow{j_1} \underline{1}[1]\bigr).$$

An arbitrary A_∞-functor $F \in \mathsf{A}_\infty^{\mathsf{u}}(\mathcal{D}_2^{\mathsf{tr}}; \mathcal{B})$ gives rise to a morphism of categories with translation structure $\tilde{H}^0(F) = (H^0(F), \phi_n) : \tilde{H}^0(\mathcal{D}_2^{\mathsf{tr}}) \to \tilde{H}^0(\mathcal{B})$, and thus determines a triangle

$$\triangle\tilde{H}^0(F) = \bigl(\underline{1}F \xrightarrow{e_{12}.H^0(F)} \underline{2}F \xrightarrow{i_2.H^0(F)} \operatorname{Cone}(e_{12})F \xrightarrow{j_1.H^0(F)\cdot\phi_1} \underline{1}F\Sigma\bigr).$$

Isomorphic A_∞-functors determine isomorphic triangles.

7.13. Definition. The triangles $\triangle\tilde{H}^0(F)$ where $F : \mathcal{D}_2^{\mathsf{tr}} \to \mathcal{B}$ runs over all unital A_∞-functors are called *standard distinguished triangles* in $\tilde{H}^0(\mathcal{B})$. *Distinguished triangles* in $\tilde{H}^0(\mathcal{B})$ are defined as those isomorphic to a standard distinguished triangle. This is the system of triangles associated with $\mathcal{B}$.

Any unital A_∞-functor $G : \mathcal{B} \to \mathcal{C}$ between pretriangulated A_∞-categories induces a triangulated functor $\tilde{H}^0(G) : H^0(\mathcal{B}) \to H^0(\mathcal{C})$ in homology.

In the setting of translation structures the definition of a triangulated category takes a form very similar to classical one, given by Verdier [36].

7.14. Definition. A *triangulated category* is an additive $\Bbbk$-linear category $\mathcal{C}$ with a translation structure on it and a system of triangles which satisfies the following axioms.

(TR1) The triangle $(A, A, 0, \mathrm{id}_A, 0, 0)$ is distinguished.

(TR2) (X, Y, Z, u, v, w) is distinguished if and only if $(Y, Z, X\Sigma, v, w, -u\Sigma)$ is distinguished.

(TR3) Given a diagram as follows,

$$\begin{tikzcd} X \arrow[r, "u"] \arrow[d, "f"'] & Y \arrow[r, "v"] \arrow[d, "g"'] & Z \arrow[r, "w"] & X\Sigma \arrow[d, "f\Sigma"] \\ X' \arrow[r, "u'"] & Y' \arrow[r, "v'"] & Z' \arrow[r, "w'"] & X'\Sigma \end{tikzcd}$$

whose rows are distinguished triangles such that the left square is commutative, there is a map $h : Z \to Z'$ making the entire diagram commutative.

(TR4) If we have three distinguished triangles $(X, Y, Z', u, i, *)$, $(Y, Z, X', v, *, j)$, and $(X, Z, Y', w, *, *)$, with $w = uv$, then there are morphisms $f : Z' \to Y'$, $g : Y' \to X'$ such that:

(a) (id_X, v, f) is a morphism of triangles;

(b) (u, id_Z, g) is a morphism of triangles;

(c) $(Z', Y', X', f, g, j(i\Sigma))$ is a distinguished triangle.

Every triangulated category is equivalent to a triangulated category whose translation structure is generated by an automorphism of this category, cf. [37].

7.15. THEOREM. *(See [3].) Let $\mathcal{C}$ be a pretriangulated A_∞-category. Then its homotopy category $\tilde{H}^0(\mathcal{C})$ is triangulated.*

The question about pretriangulated A_∞-categories is reduced to a question about a few small differential graded pretriangulated categories as in Soibelman's approach [33]. After that it remains to prove that some cycles are boundaries.

References

[1] A.A. Beilinson, V.G. Drinfeld, Chiral Algebras, Amer. Math. Soc. Colloq. Publ., vol. 51, Amer. Math. Soc., Providence, RI, 2004, http://www.math.uchicago.edu/~arinkin/.

[2] A.I. Bondal, M.M. Kapranov, Enhanced triangulated categories, Mat. Sb. 181 (5) (1990) 669–683; English transl. in: Math. USSR Sb. 70 (1) (1991) 93–107.

[3] Y. Bespalov, V.V. Lyubashenko, O. Manzyuk, Pretriangulated A_∞-categories, book in progress, 2007.

[4] R.E. Borcherds, Vertex algebras, in: M. Kashiwara, A. Matsuo, K. Saito, I. Satake (Eds.), Topological Field Theory, Primitive Forms and Related Topics, 1996, in: Progr. Math., vol. 160, Birkhäuser Boston, Boston, MA, 1998, pp. 35–77, http://arXiv.org/abs/q-alg/9706008, 1996.

[5] P. Bressler, Y.S. Soibelman, Homological mirror symmetry, deformation quantization and noncommutative geometry, J. Math. Phys. 45 (10) (2004) 3972–3982.

[6] K. Costello, Topological conformal field theories and Calabi–Yau categories, Adv. Math. 210 (1) (2007) 165–214, http://arXiv.org/abs/math.QA/0412149.

[7] B.J. Day, R.H. Street, Abstract substitution in enriched categories, J. Pure Appl. Algebra 179 (1–2) (2003) 49–63.

[8] V.G. Drinfeld, DG quotients of DG categories, J. Algebra 272 (2) (2004) 643–691, http://arXiv.org/abs/math.KT/0210114.

[9] K. Fukaya, Floer homology and mirror symmetry. II, in: Minimal Surfaces, Geometric Analysis and Symplectic Geometry, Baltimore, MD, 1999, in: Adv. Stud. Pure Math., vol. 34, Math. Soc. Japan, Tokyo, 2002, pp. 31–127.

[10] K. Fukaya, Deformation theory, homological algebra and mirror symmetry, in: Geometry and Physics of Branes, Como, 2001, in: Ser. High Energy Phys. Cosmol. Gravit., IOP, Bristol, 2003, pp. 121–209.

[11] E. Getzler, J.D.S. Jones, A_∞-algebras and the cyclic bar complex, Illinois J. Math. 34 (2) (1990) 256–283.

[12] A. Grothendieck, J.-L. Verdier, Préfaisceaux, in: SGA 4: Théorie des Topos et Cohomologie Étale des Schémas, Tome 1. Théorie des Topos, in: Lecture Notes in Math., vol. 269, Springer-Verlag, Heidelberg, 1972–1973, pp. 1–184.

[13] V.K.A.M. Gugenheim, J.D. Stasheff, On perturbations and A_∞-structures, Bull. Soc. Math. Belg. Sér. A 38 (1986) 237–246.

[14] T.V. Kadeishvili, On the theory of homology of fiber spaces, Uspekhi Mat. Nauk 35 (3) (1980) 183–188; translated in: Russian Math. Surveys 35 (3) (1980) 231–238.

[15] T.V. Kadeishvili, The algebraic structure in the homology of an $A(\infty)$-algebra, Soobshch. Akad. Nauk Gruzin. SSR 108 (2) (1982) 249–252 (in Russian).

[16] H. Kajiura, Noncommutative homotopy algebras associated with open strings, Rev. Math. Phys. 19 (1) (2007) 1–99, http://arXiv.org/abs/math.QA/0306332.

[17] B. Keller, Introduction to A-infinity algebras and modules, Homology Homotopy Appl. 3 (1) (2001) 1–35, http://arXiv.org/abs/math.RA/9910179, http://www.math.rutgers.edu/hha/.

[18] B. Keller, Addendum to: "Introduction to A-infinity algebras and modules", Homology Homotopy Appl. 3 (1) (2001) 1–35, MR1854636 (2004a:18008a); Homology Homotopy Appl. 4 (1) (2002) 25–28 (electronic).

[19] B. Keller, A-infinity algebras, modules and functor categories, in: Trends in Representation Theory of Algebras and Related Topics, in: Contemp. Math., vol. 406, Amer. Math. Soc., Providence, RI, 2006, pp. 67–93, http://arXiv.org/abs/math.RT/0510508.

[20] M. Kontsevich, Homological algebra of mirror symmetry, in: Proc. Internat. Cong. Math., Zürich, Switzerland, 1994, vol. 1, Birkhäuser, Basel, 1995, pp. 120–139, http://arXiv.org/abs/math.AG/9411018.

[21] M. Kontsevich, Y.S. Soibelman, Homological mirror symmetry and torus fibrations, in: Symplectic Geometry and Mirror Symmetry, Seoul, 2000, World Sci. Publ., River Edge, NJ, 2001, pp. 203–263, http://arXiv.org/abs/math.SG/0011041.

[22] J. Lambek, Deductive systems and categories II: Standard constructions and closed categories, in: P.J. Hilton (Ed.), Category Theory, Homology Theory and Their Applications, in: Lecture Notes in Math., vol. 86, Springer-Verlag, 1969, pp. 76–122.

[23] K. Lefèvre-Hasegawa, Sur les A_∞-catégories, Ph.D. thesis, Université Paris 7, U.F.R. de Mathématiques, http://arXiv.org/abs/math.CT/0310337, 2003.

[24] V.V. Lyubashenko, Category of A_∞-categories, Homology, Homotopy Appl. 5 (1) (2003) 1–48, http://arXiv.org/abs/math.CT/0210047, http://www.rmi.acnet.ge/hha/.

[25] V.V. Lyubashenko, O. Manzyuk, Quotients of unital A_∞-categories, Max-Planck-Institut fur Mathematik, preprint, MPI 04-19, http://arXiv.org/abs/math.CT/0306018, 2004.

[26] V.V. Lyubashenko, O. Manzyuk, Free A_∞-categories, Theory Appl. Categ. 16 (9) (2006) 174–205, http://arXiv.org/abs/math.CT/0312339.

[27] V.V. Lyubashenko, S.A. Ovsienko, A construction of quotient A_∞-categories, Homology, Homotopy Appl. 8 (2) (2006) 157–203, http://arXiv.org/abs/math.CT/0211037.

[28] S.A. Merkulov, Strong homotopy algebras of a Kähler manifold, Int. Math. Res. Not. 3 (1999) 153–164, http://arXiv.org/abs/math.AG/9809172.

[29] A. Polishchuk, A_∞-structures, Brill–Noether loci and the Fourier–Mukai transform, Compos. Math. 140 (2) (2004) 459–481, http://arXiv.org/abs/math.AG/0204092.

[30] A. Polishchuk, A_∞-structures on an elliptic curve, Comm. Math. Phys. 247 (3) (2004) 527–551, http://arXiv.org/abs/math.AG/0001048.

[31] V.A. Smirnov, Homology of fiber spaces, in: International Topology Conference, Moscow State Univ., Moscow, 1979, Uspekhi Mat. Nauk 35 (3) (1980) 227–230, translated in: Russian Math. Surveys 35 (3) (1980) 294–298.

[32] V.A. Smirnov, Simplicial and Operad Methods in Algebraic Topology, Transl. Math. Monogr., vol. 198, Amer. Math. Soc., Providence, RI, 2001; translated from the Russian manuscript by G.L. Rybnikov.

[33] Y.S. Soibelman, Mirror symmetry and noncommutative geometry of A_∞-categories, J. Math. Phys. 45 (10) (2004) 3742–3757.

[34] J.D. Stasheff, Homotopy associativity of H-spaces I and II, Trans. Amer. Math. Soc. 108 (1963) 275–292, 293–312.

[35] T. Tradler, Infinity-inner-products on A-infinity-algebras, http://arXiv.org/abs/math.AT/0108027, 2001.

[36] J.-L. Verdier, Catégories dérivées, in: Cohomologie Étale, in: Lecture Notes in Math., vol. 569, Springer-Verlag, Berlin, 1977, pp. 262–311.

[37] J.-L. Verdier, Des catégories dérivées des catégories abéliennes, Astérisque 239 (1996), xii+253 pp., (1997), with a preface by L. Illusie, edited and with a note by G. Maltsiniotis.

0-Cohomology of Semigroups

B. Novikov

Department of Mathematics, University of Kharkov, Ukraine
E-mail: *boris.v.novikov@univer.kharkov.ua*

Contents

HANDBOOK OF ALGEBRA, VOL. 5
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(07)05004-8

This page intentionally left blank

In 1974 my teacher L.M. Gluskin suggested to me to study projective representations of semigroups. As one could expect, cohomology appeared in this problem, however it "slightly" differed from Eilenberg–MacLane cohomology. I have named it "0-cohomology", have studied its properties insofar as this was necessary for the initial problem, and thought that I would probably never meet this notion again.

However, in the last 30 years I returned to 0-cohomology again and again since I met problems, in which it appeared.

This chapter is a survey of 0-cohomology. The main attention is devoted to applications. The necessary definitions and results from homological algebra and theory of semigroups can be found in [5], [7], and [19].

I have to note that quite a lot of various cohomology theories have been adapted to solution of specific problems in semigroups [6,12,18,23–25,42,44,46]; see also my review [40]. So the invention of one more cohomology for this purpose is not any novelty. However, 0-cohomology seems attractive (at least to me!) because it links semigroups with several different branches of the algebra.

1. Eilenberg–MacLane cohomology

The definition of semigroup cohomology does not differ from group cohomology [5]: for a semigroup S and a (left) S-module A the *n-dimensional cohomology group* is the group $H^n(S, A) = \mathrm{Ext}^n_{\mathbb{Z}S}(\mathbb{Z}, A)$ where $\mathbb{Z}$ is considered as a trivial $\mathbb{Z}S$-module. We will call this cohomology *Eilenberg–MacLane cohomology* or briefly *EM-cohomology*.

In what follows another well-known definition will be used. $C^n(S, A)$ denotes the group of all n-place mappings $f : \underbrace{S \times \cdots \times S}_{n \text{ times}} \to A$ (the group of n-dimensional cochains); a coboundary operator $\partial^n : C^n(S, A) \to C^{n+1}(S, A)$ is defined as follows:

$$\begin{aligned}\partial^n f(x_1, \ldots, x_{n+1}) = {} & x_1 f(x_2, \ldots, x_{n+1}) \\ & + \sum_{i=1}^{n} (-1)^i f(x_1, \ldots, x_i x_{i+1}, \ldots, x_{n+1}) \\ & + (-1)^{n+1} f(x_1, \ldots, x_n).\end{aligned} \tag{1}$$

Then $\partial^n \partial^{n-1} = 0$, i.e.

$$\begin{aligned}\mathrm{Im}\,\partial^{n-1} &= B^n(S, A) \quad \text{(the group of } n\text{-dimensional coboundaries)} \\ &\subseteq \mathrm{Ker}\,\partial^n = Z^n(S, A) \quad \text{(the group of } n\text{-dimensional cocycles)}\end{aligned}$$

and the cohomology groups are defined as $H^n(S, A) = Z^n(S, A)/B^n(S, A)$.

However, for the cohomology of semigroups one does not manage to obtain results which are comparable to theory of cohomology of groups. In this section we give several results about the cohomology of semigroups, illustrating its specifics.

Here is a typical example: since a projective module over a semigroup is not necessarily projective over a subsemigroup of the semigroup, the lemma of Shapiro [4], which expresses the cohomology of a subgroup through the cohomology of the containing group, does not hold for semigroups. So a result, obtained in [5], is of interest:

THEOREM 1. *Let T be a subsemigroup of a group G. If G is a group of fractions for T (i.e. each element from G can be written in the form $x^{-1}y$ for some $x, y \in T$), then the homomorphisms $i^n : H^n(G, A) \to H^n(T, A)$, induced by the embedding $i : T \to G$, are bijective for any G-module A.*

Let I be an ideal of a semigroup S. What can one say about homomorphisms $\varepsilon^n : H^n(S, A) \to H^n(I, A)$, induced by the embedding $\varepsilon : I \to S$? It is easy to show that ε^0 is always an isomorphism and that ε^1 is a monomorphism. Using a technique of adjoint functors, Adams and Rieffel [1] proved

THEOREM 2. *Let I be a left ideal of a semigroup S, having an identity e. Then for any S-module A and for any $n \geqslant 0$*

$$H^n(S, A) \cong H^n(I, A) \cong H^n(I, eA).$$

In particular, if S contains zero then $H^n(S, A) = 0$ for $n > 0$.

In [1], using Theorem 2, a sufficient condition was obtained for an associative algebra over $\mathbb{R}$ be a semigroup algebra.

The connection between $H^n(S, A)$ and $H^n(I, A)$ becomes closer when we take for I the so-called Sushkevich kernel (the least two-sided ideal). This situation was considered in detail by W. Nico [27]. I will not formulate this result, but only note that it implies a description of the cohomology of completely simple semigroups:

THEOREM 3. *Let S be a completely simple semigroup, G its basic group, e the identity of G, A an S-module. Then $H^n(S, A) \cong H^n(G, eA)$ for any $n \geqslant 3$.*

A number of papers is devoted to the study of the cohomological dimension of semigroups. *Cohomological dimension* (cd S) of a semigroup S is defined as the least integer n such that $H^{n+1}(S, A) = 0$ for any S-module A.

It is well known [5] that cohomological dimension of both a free group and of a free semigroup (monoid) does not exceed 1. In the case of groups the converse statement is also true (this is the famous Stallings–Swan theorem [4]). For semigroups the situation is entirely different: joining an extra zero to any semigroup makes all its cohomology groups trivial, except the 0-dimensional one. So it is natural to confine ourselves to the class of semigroups with cancellation. Mitchell [26] has shown that the free product of a free group and a free monoid (which he called a partially free monoid) has cohomological dimension 1. In the same paper he has formulated a suggestion: if S is a cancellative monoid with $\operatorname{cd} S = 1$ then S is partially free.

In [32] (and later in [34] more generally) a counter-example to the Mitchell suggestion was constructed and a "weakened Mitchell conjecture" was proposed: if S is a cancellative monoid with $\operatorname{cd} S = 1$ then S can be embedded into a free group. This suggestion was proved in [39]. Probably, it is difficult to get more exact information: it is shown in [35] that a semigroup anti-isomorphic to the counter-example from [32] (and therefore also embeddable into a free group) has cohomological dimension 2. A good answer is known only in the commutative case: the cohomological dimensions of all subsemigroups of $\mathbb{Z}$ are equal to 1 [36].

2. Properties of 0-cohomology

Before constructing 0-cohomology we need to define a suitable Abelian category. In what follows S is a semigroup with zero.

DEFINITION 1. A 0-module over S is an Abelian (additive) group A equipped with a multiplication $(S \setminus 0) \times A \to A$ satisfying the following conditions for all $s, t \in S \setminus 0$, $a, b \in A$:

$$s(a+b) = sa + sb,$$
$$st \neq 0 \quad \Rightarrow \quad s(ta) = (st)a.$$

A morphism of 0-modules is a homomorphism of Abelian groups $\varphi : A \to B$ such that $\varphi(sa) = s\varphi(a)$ for $s \in S \setminus 0,\ a \in A$.

We will denote the category of 0-modules thus defined by $\mathrm{Mod}_0\, S$. It is easy to see that for the semigroup $T^0 = T \cup 0$ with an extra zero the category $\mathrm{Mod}_0\, T^0$ is isomorphic to the category $\mathrm{Mod}\, T$ of the usual modules over T. It turns out that quite generally $\mathrm{Mod}_0\, S$ is also isomorphic to a category of (standard) modules over some semigroup.

Denote by $\overline{S}$ the set of all finite sequences $(x_1, \ldots, x_m)$ such that $x_i \in S \setminus 0$ $(1 \leqslant i \leqslant m)$ and $x_i x_{i+1} = 0$ $(1 \leqslant i < m)$; thus, all one-element sequences, except (0), are contained in $\overline{S}$. Define a binary relation ρ on $\overline{S}$ via $(x_1, \ldots, x_m)\rho(y_1, \ldots, y_n)$ if and only if one of the following conditions is fulfilled:

(1) $m = n$ and there exists i $(1 \leqslant i \leqslant m-1)$ that $x_i = y_i u$, $y_{i+1} = u x_{i+1}$ for some $u \in S$, and $x_j = y_j$ for $j \neq i$, $j \neq i+1$;

(2) $m = n+1$ and there exists an i $(2 \leqslant i \leqslant m-1)$ such that $x_i = uv$, $y_{i-1} = x_{i-1}u$, $y_i = v x_{i+1}$ for some $u, v \in S$, $x_j = y_j$ for $1 \leqslant j \leqslant i-2$, and $x_j = y_{j-1}$ for $i+2 \leqslant j \leqslant m$.

Let $\bar{\rho}$ be the least equivalence containing ρ, $\tilde{S}$ the quotient set $\overline{S}/\bar{\rho}$. The image of $(x_1, \ldots, x_m) \in \overline{S}$ in $\tilde{S}$ will be denoted by $[x_1, \ldots, x_m]$.

Define a multiplication on $\tilde{S}$:

$$[x_1, \ldots, x_m][y_1, \ldots, y_n] = \begin{cases} [x_1, \ldots, x_m y_1, \ldots, y_n], & \text{if } x_m y_1 \neq 0, \\ [x_1, \ldots, x_m, y_1, \ldots, y_n], & \text{if } x_m y_1 = 0. \end{cases}$$

Then $\tilde{S}$ becomes a semigroup, which is called the *gown* of S.

Each 0-module over S can be transformed into a (usual) module over $\tilde{S}$ by

$$[x_1, \ldots, x_m]a = x_1\big(\ldots(x_m a)\ldots\big)$$

for $x_1, \ldots, x_m \in S \setminus 0$, $a \in A$. Hence we obtain

PROPOSITION 1. $\mathrm{Mod}_0\, S \cong \mathrm{Mod}\, \tilde{S}$.

COROLLARY 1. *The category* $\mathrm{Mod}_0\, S$ *is Abelian.*

Here are some simplest properties of the gown of a semigroup [33]:

(i) If $S = T^0$ is a semigroup with an extra zero then $\tilde{S} \cong S \setminus 0 = T$.

(ii) It follows from the definition of the relation ρ that the map $S \setminus 0 \to \tilde{S}$, $x \to [x]$, is bijective.

(iii) The subset $J = \{[x_1, \dots, x_m] \in \tilde{S}\colon m > 1\}$ is an ideal in $\tilde{S}$ and $\tilde{S}/J \cong S$.

It is easy to find the gown if the semigroup S is given by defining relations. We will write $S = \langle a_1, \dots, a_m \mid P_i = Q_i, 1 \leqslant i \leqslant n\rangle$ if S is generated by elements $a_1, \dots, a_m$ and is defined by equalities $P_1 = Q_1, \dots, P_n = Q_n$. If the value of a word P_i (or the same, of Q_i) in a semigroup S is 0 then the equality $P_i = Q_i$ is called a *zero* relation.

PROPOSITION 2. *Let $S = \langle a_1, \dots, a_m \mid P_i = Q_i, 1 \leqslant i \leqslant n\rangle$ be a semigroup with a zero, in which none of the generating elements is* 0. *If one deletes all zero defining relations then the obtained semigroup will be isomorphic to the gown $\tilde{S}$.*

EXAMPLE 1. If S is a semigroup with zero multiplication ($S^2 = 0$) then $\tilde{S}$ is a free semigroup.

Let now A be a 0-module over S.

DEFINITION 2. A partial n-place mapping from S into A, defined on all n-tuples $(x_1, \dots, x_n)$ such that $x_1 \cdots x_n \neq 0$, is called a n-dimensional cochain. The group of n-dimensional cochains is denoted by $C_0^n(S, A)$. A coboundary operator $\partial^n : C_0^n(S, A) \to C_0^{n+1}(S, A)$ is given as above, by the formula (1). Then $\partial^n \partial^{n-1} = 0$ and the 0-cohomology groups are defined as $H_0^n(S, A) = Z_0^n(S, A)/B_0^n(S, A)$, where $Z_0^n(S, A) = \operatorname{Ker} \partial^n$ is the group of n-dimensional 0-cocycles, $B_0^n(S, A) = \operatorname{Im} \partial^{n-1}$ is the group of n-dimensional 0-coboundaries.

EXAMPLE 2. Let $S = T^0 = T \cup 0$ be a semigroup with an extra zero. Then $\tilde{S} \cong T$ and one can easily check that $H_0^n(S, A) \cong H^n(T, A)$[1] while $H^n(S, A) = 0$.

This example shows that 0-cohomology is a generalization of EM-cohomology. In view of Proposition 1, in general case it is naturally to compare the groups $H_0^n(S, A)$ and $H^n(\tilde{S}, A)$. We describe this comparison in more detail.

Since the sequence of functors $\{H_0^n(S, \_)\}_{n \geqslant 0}$ is connected in the terminology of [19] and $\{H^n(\tilde{S}, \_)\}_{n \geqslant 0}$ is a sequence of derived functors in the isomorphic categories $\mathrm{Mod}_0\, S$ and $\mathrm{Mod}\, \tilde{S}$ respectively, the isomorphism

$$\varepsilon^0\colon\ H^0(\tilde{S}, A) = \mathrm{Hom}_{\mathrm{Mod}\, \tilde{S}}(\mathbb{Z}, A) \cong \mathrm{Hom}_{\mathrm{Mod}_0\, S}(\mathbb{Z}, A) = H_0^0(S, A)$$

induces group homomorphisms

$$\varepsilon^n : H^n(\tilde{S}, A) \to H_0^n(S, A)$$

such that $\{\varepsilon^n\}_{n \geqslant 0}$ are morphisms of cohomology functors.

[1] Here we consider A both as a 0-module over S and a module over $\tilde{S}$, which does not lead to misunderstanding in this context.

The homomorphism ε^n is described as follows. If $f \in C^n(\tilde{S}, A)$ then set

$$\big(\eta^n f\big)(x_1, \dots, x_n) = f\big([x_1], \dots, [x_n]\big) \quad \text{for } x_1 \cdot \dots \cdot x_n \neq 0.$$

Then η^n is a homomorphism from $C^n(\tilde{S}, A)$ into $C_0^n(S, A)$ and it induces the homomorphism ε^n.

Direct calculations prove

THEOREM 4. *ε^1 is an isomorphism for any semigroup S.*

Using the corresponding long exact sequence, from Theorem 4 we obtain

COROLLARY 2. *ε^2 is a monomorphism for any semigroup S.*

Generally speaking, the groups $H_0^2(S, A)$ and $H^2(\tilde{S}, A)$ are not isomorphic:

EXAMPLE 3. Let the commutative semigroup S consists of elements $u, v, w, 0$ with the multiplication

$$u^2 = v^2 = uv = w, \qquad uw = vw = 0.$$

One can show that $H^2(\tilde{S}, A) = 0$ for any module A over S. On the other hand, if A (considered now as a 0-module) is not one-element and $a \in A \setminus 0$, then the 0-cocycle f defined by the condition

$$f(x, y) = \begin{cases} a & \text{for } x = y = u, \\ 0 & \text{otherwise,} \end{cases}$$

is not a 0-coboundary and thus $H_0^2(S, A) \neq 0$.

Apropos, this example shows that 0-cohomology is not a derived functor unlike EM-cohomology. Indeed, according to Proposition 1 there are injective objects in the category $\mathrm{Mod}_0\ S$ and a derived functor must vanish on them.

In this situation semigroups categorical at zero are of special interest. We recall that a semigroup S with a zero is called *categorical at zero* if for any $x, y, z \in S$ from $xyz = 0$ it follows that either $xy = 0$ or $yz = 0$. For instance, if we join a new element 0 to the set of all morphisms of a small category and set the product of morphisms equal to 0 when their composition is not defined, then the obtained set becomes a semigroup categorical at zero.

THEOREM 5. (*See* [28].) *If a semigroup S is categorical at zero then ε^n is an isomorphism for any $n \geqslant 0$.*

This theorem is used in two ways. On the one hand, since, for instance, completely 0-simple semigroups are categorical at zero then by Theorem 5 one succeeds to calculate their 0-cohomology in the sense that the problem is reduced to EM-cohomology [30]. On the other hand, in concrete examples usually it is easier to calculate the 0-cohomology of a given semigroup, and then use that for finding EM-cohomology of its gown. We will consider this question in more detail in the following section.

Example 3 shows that in general Theorem 5 does not hold.

3. Calculating EM-cohomology

The free product of semigroups gives a first example of the use of 0-cohomology:

THEOREM 6. *Let S, T be semigroups, $S * T$ their free product. Then*

$$H^n(S * T, A) \cong H^n(S, A) \oplus H^n(T, A)$$

for any $n \geqslant 2$ and for any $(S * T)$-module A.

PROOF. The 0-direct union $S^0 \sqcup_0 T^0$ is a semigroup categorical at zero. Besides, $\widetilde{S^0 \sqcup_0 T^0} \cong S * T$. So

$$\begin{aligned} H^n(S * T, A) &\cong H_0^n\bigl(S^0 \sqcup_0 T^0, A\bigr) \cong H_0^n\bigl(S^0, A\bigr) \oplus H_0^n\bigl(T^0, A\bigr) \\ &\cong H^n(S, A) \oplus H^n(T, A). \end{aligned}$$

Here, the first isomorphism follows from Theorem 5, the second is checked directly, and third follows from Example 2. □

REMARK. There is an analogue of Theorem 6 for groups, but its proof is more complicated.

In Section 1 the notion of cohomological dimension was already mentioned. A counterexample to Mitchell conjecture was obtained by using 0-cohomology. This is the semigroup

$$S = \langle a, b, c, d \mid ab = cd \rangle.$$

Its subset $I = S \setminus \{a, b, c, d, ab\}$ is an ideal and $\widetilde{S/I} \cong S$. By Corollary 2 $H^2(S, A)$ embeds into $H_0^2(S/I, A)$.

On the other hand, it is easy to see that $f = \partial\varphi$ for any 0-cocycle $f \in Z_0^2(S/I, A)$ if we set

$$\varphi(u) = \begin{cases} f(a, b), & \text{if } u = a, \\ f(c, d), & \text{if } u = c, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore $H^2(S, A) = H_0^2(S/I, A) = 0$.

Here is one more example of the calculation of EM-cohomology for a pair of anti-isomorphic semigroups.

Let $p, q \in \mathbb{N}$. Consider the semigroup

$$T = \bigl\langle a, b \mid ab = b, a^p = a^q \bigr\rangle.$$

This is the gown for the semigroup

$$S = \langle x, y \mid xy = y, x^p = x^q, yx = y^2 = 0\rangle.$$

The last is categorical at zero and consists of the elements x^k $(k > 0)$, y and 0. Its 0-cohomology is easily computed directly, and thus we get:

PROPOSITION 3. (*See* [31].) *For any T-module A*:
(1) $H^1(T, A) = H_0^1(S, A) \cong A/\{m \in A \mid am = 2m\}$;
(2) $H^n(T, A) = H_0^n(S, A) = 0$ *for* $n \geqslant 2$.

For the semigroup T^{op} anti-isomorphic to T, i.e.

$$T^{\mathrm{op}} = \langle a, b \mid ba = b, a^p = a^q\rangle,$$

the answer is more complicated:

PROPOSITION 4. (*See* [31].) *Let A be an arbitrary T^{op}-module, A_1 its underlying additive group, considered as a T^{op}-module with trivial multiplication, $\langle a\rangle$ the subsemigroup generated by the element a. Let the homomorphisms ψ^n be induced by the embedding $\langle a\rangle \to T^{\mathrm{op}}$. Then the sequence*

$$0 \to H^0\big(T^{\mathrm{op}}, A\big) \xrightarrow{\psi^0} H^0\big(\langle a\rangle, A\big) \to H^0\big(\langle a\rangle, A_1\big) \to H^1\big(T^{\mathrm{op}}, A\big)$$
$$\xrightarrow{\psi^1} \cdots \to H^n\big(T^{\mathrm{op}}, A\big) \xrightarrow{\psi^n} H^n\big(\langle a\rangle, A\big) \to H^n\big(\langle a\rangle, A_1\big) \to \cdots$$

is exact.

In particular, if A_1 is torsion-free then $H^n(T^{\mathrm{op}}, A) \cong H^n(\langle a\rangle, A)$ for $n \geqslant 1$.

More general results in the calculation of EM-cohomology were obtained by partial cohomology (see Section 8).

Finally let me formulate an unsolved problem. Theorem 2 shows that in some cases the cohomology of a semigroup is defined by the cohomology of an ideal in it. Generally speaking, this is not the case. When I is a two-sided ideal, it is desirable to use for the calculation of $H^n(S, A)$ not only the cohomology of an ideal, but also of the quotient semigroup S/I. However, the EM-cohomology of the latter is always trivial, because S/I contains the zero element. So a question arises: how does the group $H^n(S, A)$ depend on $H^n(I, A)$ and $H_0^n(S/I, A)$ (as well as, maybe, on the cohomology groups of smaller dimension)?

4. Projective representations

In this section we use the following notation. S is an arbitrary semigroup (for simplicity we suppose that it contains an identity), K is a field, $K^\times$ its multiplicative group, $\mathrm{Mat}_n K$ the multiplicative semigroup of all $(n \times n)$-matrices over K (we will often delete the subscript n). Define an equivalence λ on $\mathrm{Mat}_n K$ as follows: for $A, B \in \mathrm{Mat}_n K$ set

$$A \lambda B \quad \Longleftrightarrow \quad \exists c \in K^\times\colon\ A = cB.$$

Evidently λ is a congruence of $\mathrm{Mat}_n K$. The quotient semigroup $\mathrm{PMat}_n K = \mathrm{Mat}_n K/\lambda$ will be called *a semigroup of projective* $(n \times n)$*-matrices*.

Let $\Delta : S \to \mathrm{PMat}_n K$ be a homomorphism and $\alpha : \mathrm{Mat}_n K \to \mathrm{PMat}_n K$ be the canonical homomorphism, corresponding to the congruence λ. Fix a mapping $\beta : \mathrm{PMat}_n K \to \mathrm{Mat}_n K$, by choosing representatives in λ-classes. If we write $\Gamma = \beta\Delta$ then $\Delta = \alpha\beta\Delta = \alpha\Gamma$. Since Δ and α are homomorphisms,

$$\alpha\Gamma(xy) = \Delta(x)\Delta(y) = \alpha\big(\Gamma(x)\Gamma(y)\big)$$

for all $x, y \in S$. Hence $\Gamma(xy)$ and $\Gamma(x)\Gamma(y)$ vanish simultaneously. This provides the motivation for the following definition:

DEFINITION 3. A mapping $\Gamma : S \to \mathrm{Mat}_n K$ is called a projective representation[2] of S over K if it satisfies the following conditions:

(1) $\Gamma(xy) = 0 \Leftrightarrow \Gamma(x)\Gamma(y) = 0$ for all $x, y \in S$;

(2) there is a partially defined mapping $\rho : S \times S \to K^{\times}$ such that

$$\operatorname{dom}\rho = \big\{(x, y) \bigm| \Gamma(xy) \neq 0\big\} \tag{2}$$

and

$$\forall (x, y) \in \operatorname{dom}\rho \quad \Gamma(x)\Gamma(y) = \Gamma(xy)\rho(x, y). \tag{3}$$

The mapping ρ is called a *factor set* of Γ and the number n *the degree* of Γ.

REMARK. It is easy to see that (3) remains valid for all $x, y \in S$ if we extend ρ to a completely defined mapping setting $\rho(x, y) = 0$ for x, y such that $\rho(x, y)$ was not defined. Hereafter we will often suppose this.

As in the case of projective representations of groups, it is desirable to have an independent characterization of the partially defined mappings $\rho : S \times S \to K^{\times}$ which can serve as factor sets for some projective representations of S. Applying (3) to the equality $\Gamma(x)[\Gamma(y)\Gamma(z)] = [\Gamma(x)\Gamma(y)]\Gamma(z)$, we get:

$$\rho(x, y)\rho(xy, z) = \rho(x, yz)\rho(y, z) \tag{4}$$

for all $x, y, z \in S$. However, unlike in the case of projective representations of groups, condition (4) is not sufficient.

THEOREM 7. (*See* [29].) *A mapping* $\rho : S \times S \to K$ *is a factor set for a certain* (*possibly infinite-dimensional*) *projective representation of a monoid* S *if and only if* ρ *satisfies* (4) *and for all* $x, y \in S$

$$\rho(x, y) = 0 \iff \rho(1, xy) = 0. \tag{5}$$

As in the case of groups, the choice of different representatives of the λ-classes leads to an equivalent projective representation. So we call two factor sets ρ and σ *equivalent*

[2] One also calls the homomorphism Δ a projective representation.

($\rho \sim \sigma$) if they vanish simultaneously and there exists a function $\alpha : S \to K^\times$ such that for all $x, y \in S$ we have

$$\rho(x, y) = \alpha(x)\alpha(xy)^{-1}\alpha(y)\sigma(x, y).$$

Define the product of factor sets ρ and σ by pointwise multiplication: $\rho\sigma(x, y) = \rho(x, y)\sigma(x, y)$. It follows immediately from Theorem 7 that $\rho\sigma$ is also a factor set. So the set $m(S)$ of all factor sets is a semigroup and $\sim$ is a congruence on it. The quotient semigroup $M(S) = m(S)/\sim$ is called the *Schur multiplier* of S.

For groups the Schur multiplier is isomorphic to the group $H^2(G, K^\times)$ [8]. In our situation it is a commutative inverse semigroup. Consider the construction of the semigroups $M(S)$ and $m(S)$.

Since $m(S)$ and $M(S)$ are commutative, it follows from the Clifford theorem [7] that they are strong semi-lattices of groups:

$$m(S) = \bigcup_{\alpha \in b(S)} m_\alpha(S), \qquad M(S) = \bigcup_{\alpha \in B(S)} M_\alpha(S),$$

where $b(S)$ and $B(S)$ are semi-lattices, $m_\alpha(S)$ and $M_\alpha(S)$ are groups. We will call $m_\alpha(S)$ and $M_\alpha(S)$ *components* of semigroups $m(S)$ and $M(S)$, respectively.

The first step in our considerations is a description of idempotent factor sets:

LEMMA 1. *There is a bijection $\varepsilon \leftrightarrow I_\varepsilon$ between idempotents $\varepsilon \in m(S)$ and ideals of S such that*

$$\varepsilon(x, y) = \begin{cases} 1, & \text{if } xy \notin I_\varepsilon, \\ 0, & \text{if } xy \in I_\varepsilon, \end{cases}$$

and

$$I_{\varepsilon_1 \varepsilon_2} = I_{\varepsilon_1} \cup I_{\varepsilon_2}. \tag{6}$$

We will denote by $Y(S)$ the semi-lattice of all (two-sided) ideals of S with respect to union. We consider the empty subset as an ideal as well, i.e. $\emptyset \in Y(S)$.

COROLLARY 3. $b(S) \cong B(S) \cong Y(S)$.

It follows that the ideals $I \in Y(S)$ can serve as indices for the components of the semigroups $m(S)$ and $M(S)$; thus

$$m_I(S)m_J(S) \subseteq m_{I\cup J}(S), \qquad M_I(S)M_J(S) \subseteq M_{I\cup J}(S).$$

Let ε_I be the identity of the group $m_I(S)$. Then

$$\varepsilon_I(x, y) = 0 \quad \Longleftrightarrow \quad xy \in I.$$

LEMMA 2. *The group $m_I(S)$ consists of the factor sets ρ for which*

$$\rho(x, y) = 0 \quad \Longleftrightarrow \quad xy \in I.$$

Hence the groups $m_I(S)$ and $m_0(S/I)$ are isomorphic for $I \neq \emptyset$. If $I = \emptyset$ we have $m_\emptyset(S) \cong m_0(S^0)$. Certainly, this also holds for the multiplier:

COROLLARY 4. *$M_I(S) \cong M_0(S/I)$ if $I \neq \emptyset$, and $M_\emptyset(S) \cong M_0(S^0)$.*

Finally, it is easy to see that $M_0(S) \cong H_0^2(S, K^\times)$, and we get the final result:

THEOREM 8. (*See* [29].) *The Schur multiplier $M(S, K)$ of a semigroup S over a field K is isomorphic to the semi-lattice Y of Abelian groups $H_0^2(S/I, K^\times)$, where $I \in Y$, and $K^\times$ is considered as a trivial 0-module over S/I.*

A further analysis of projective representations of semigroups was carried out in [29]; it is similar to the description of linear representations [7].

5. Brauer monoid

In several articles Haile, Larson and Sweedler [14–16,45], see also [17], studied so-called strongly primary algebras. Their definition is rather bulky and we will not need it. Instead of this I cite their description, given in [15].

Let K/L be a finite Galois extension with Galois group G. *A weak 2-cocycle* [45] is defined as a mapping $f : G \times G \to K$ such that for any $\sigma, \tau, \omega \in G$

$$\sigma\big[f(\tau, \omega)\big] f(\sigma\tau, \omega) = f(\sigma, \tau) f(\sigma\tau, \omega),$$
$$f(1, \sigma) = f(\sigma, 1) = 1$$

(hence weak 2-cocycles can take on the value zero unlike usual cocycles).

Let f be a weak 2-cocycle. On the set A of formal sums of the form $\sum_{\sigma \in G} a_\sigma \sigma$, $a_\sigma \in K$, we define a multiplication by the rule:

$$a\sigma \cdot b\tau = a\sigma(b) f(\sigma, \tau)\sigma\tau, \quad \sigma, \tau \in G, \ a, b \in K.$$

Then A becomes an associative algebra. The class of such algebras coincides with the class of strongly primary algebras.

Strongly primary algebras give a generalization of central simple algebras. In accordance with this Haile, Larson and Sweedler introduced a notion of a Brauer monoid as a generalization of the Brauer group. For this purpose an equivalence of weak 2-cocycles is defined: $f \sim g$ if there exists a mapping $p : G \to K^\times$ such that

$$g(\sigma, \tau) = f(\sigma, \tau) p(\sigma) p(\tau) \big(p(\sigma\tau)\big)^{-1}$$

for any $\sigma, \tau \in G$ under condition $f(\sigma, \tau) \neq 0$. After factorization by this equivalence the set of weak 2-cocycles turns into the *Brauer monoid* $Br(G, K)$ which is an inverse semigroup like the Schur multiplier from Section 4. More exactly, denoting by E the semi-lattice of all idempotents from $Br(G, K)$ (i.e. weak cocycles taking only the values 0 and 1), we obtain:

THEOREM 9. (*See* [16].) *$Br(G, K)$ is a semi-lattice E of Abelian groups $Br_e(G, K)$, where $e \in E$ and $Br_e(G, K)$ consists of all weak 2-cocycles which vanish simultaneously with e. In particular, if $e \equiv 1$ then $Br_e(G, K) \cong H^2(G, K^\times)$ is the Brauer group.*

It turns out [20,37] that this construction reduces to 0-cohomology. Let $e \in E$. Add an extra zero 0 to G and define a new operation on G^0:

$$x \circ y = \begin{cases} xy, & \text{if } e(x, y) = 1, \\ 0, & \text{if } e(x, y) = 0 \end{cases}$$

and, moreover, $x \circ 0 = 0 \circ x = 0$. With this operation G^0 is a semigroup which we will denote by G_e. Conversely, a *modification* $G(\circ)$ of the group G is a monoid on G^0 with an operation $\circ$ such that $x \circ y$ is either xy or 0, and moreover $0 \circ x = x \circ 0 = 0$. It is easy to see that there is a bijective correspondence between idempotent weak 2-cocycles and modifications of G. The group $K^\times$ turns into a 0-module over G_e, $Br_e(G, K) \cong H_0^2(G_e, K^\times)$ and Theorem 9 changes into the following statement:

THEOREM 10. *$Br(G, K)$ is a semi-lattice of Abelian groups $H_0^2(G(\circ), K^\times)$, where $G(\circ)$ runs over the set of all modifications of the group G.*

As is shown in [37], in this problem 0-cohomology is used essentially: to describe some properties of the Brauer monoid one has to use 0-cohomology of other (different from modifications) semigroups.

Thus the study of the Brauer monoid is reduced to a description of the modifications of the group and their 0-cohomology. However, it is necessary to note that the study of modifications is a difficult combinatorial problem. In general for a finite group G (only such groups are considered in Haile–Larson–Sweedler theory) we can only confirm that each modification is an union of the subgroup of its invertible elements and a nilpotent ideal. Besides, modifications are 0-cancellative (if $ax = bx \neq 0$ or $xa = xb \neq 0$ then $a = b$). Some examples of modification were considered in [38] and [41].

6. Partial representations of groups

The results of this section were obtained when I worked in São Paulo, Brasil, thanks to the foundation FAPESR. They were announced at the XVIII Brazilian Algebra Meeting [9] and are being prepared for publication.

In connection with studying C^*-algebras so-called partial linear representations of groups appeared [10,43]. It is natural to ask: what do partial *projective* representations of groups look like? It turned out that also here 0-cohomology appears. We start with the needed definitions from [10].

DEFINITION 4. A mapping $\varphi : G \to S$ from a group G into a semigroup S is called a partial homomorphism if for all $x, y \in G$,

$$\begin{aligned}
&\varphi\big(x^{-1}\big)\varphi(x)\varphi(y) = \varphi\big(x^{-1}\big)\varphi(xy),\\
&\varphi(x)\varphi(y)\varphi\big(y^{-1}\big) = \varphi(xy)\varphi\big(y^{-1}\big),\\
&\varphi(x)\varphi(e) = \varphi(x)
\end{aligned}$$

(these equalities imply $\varphi(e)\varphi(x) = \varphi(x)$).

In particular, a partial linear representation (PLR) over a field K is a partial homomorphism into the matrix semigroup, $\varDelta : G \to \mathrm{Mat}_n K$.

R. Exel introduced a monoid $\varSigma(G)$ which plays a special role here. It is generated by symbols $[x]$ $(x \in G)$ with defining relations

$$\begin{aligned}
&\big[x^{-1}\big][x][y] = \big[x^{-1}\big][xy],\\
&[x][y]\big[y^{-1}\big] = [xy]\big[y^{-1}\big],\\
&[x][e] = [x]
\end{aligned}$$

(these equalities imply $[e][x] = [x]$).

$\varSigma(G)$ possesses the following universal property:

(i) The mapping $f : G \to \varSigma(G)$, $f(x) = [x]$, is a partial homomorphism.

(ii) For any semigroup S and any partial homomorphism $\varphi : G \to S$ there exists a unique (usual) homomorphism $\tilde{\varphi} : \varSigma(G) \to S$ such that $\varphi = \tilde{\varphi} f$.

Due to this property the study of PLR's of groups is equivalent to the study of linear representations of its Exel monoid.

It is natural to define (and to study) partial projective representations of groups by means of the usual projective representations of $\varSigma(G)$: we will call the partial homomorphism $\varDelta : G \to \mathrm{PMat}\,K$ a *partial projective representation (PPR)* of G (cf. the footnote on p. 198). We get the diagram

$$\begin{tikzcd}
& \varSigma(G) \arrow[dl, "f"'] \arrow[dr, "\tilde{\varDelta}"] & \\
G \arrow[rr, "\varDelta"'] & & \mathrm{PMat}\,K
\end{tikzcd}$$

where $\tilde{\varDelta}$ is a projective representation of $\varSigma(G)$.

However we will see below that PPR's are not reducible to projective representations of semigroups unlike in the linear case.

The first step in study of PPR's is a translation of their definition into the usual language of matrices:

THEOREM 11. *A mapping $\varGamma : G \to \mathrm{Mat}\,K$ is PPR of G if and only if the following conditions hold*:

(1) *for all $x, y \in G$*

$$\varGamma\big(x^{-1}\big)\varGamma(xy) = 0 \iff \varGamma(x)\varGamma(y) = 0 \iff \varGamma(xy)\varGamma\big(y^{-1}\big) = 0;$$

(2) *there is a mapping $\sigma : G \times G \to K$ such that*

$$\varGamma(x)\varGamma(y) = 0 \iff \sigma(x, y) = 0$$

and

$$\Gamma\big(x^{-1}\big)\Gamma(x)\Gamma(y) = \Gamma\big(x^{-1}\big)\Gamma(xy)\sigma(x, y),$$
$$\Gamma(x)\Gamma(y)\Gamma\big(y^{-1}\big) = \Gamma(xy)\Gamma\big(y^{-1}\big)\sigma(x, y).$$

Note that this theorem gives another definition of the notion of a PPR independent of $\Sigma(G)$.

We call σ a *factor set* of Γ and define a product of factor sets as above. However, it is not evident that this product is also a factor set. To prove this, one uses the Exel monoid again:

PROPOSITION 5. *Let $\sigma : G \times G \to K$ be a mapping for which there is a factor set ρ of the semigroup $\Sigma(G)$, such that*:

(1) $$\forall x, y \in G \quad \sigma(x, y) = 0 \quad \Leftrightarrow \quad \rho\big([x], [y]\big) = 0;$$

(2) $$\sigma(x, y) \neq 0 \quad \Rightarrow \quad \sigma(x, y) = \frac{\rho([x], [y])\rho([x^{-1}], [x][y])}{\rho([x^{-1}], [xy])}.$$

Then σ is a factor set of some PPR of G.

The converse also holds.

Now the desired result about products follows directly. Moreover:

COROLLARY 5. *The factor sets of G form a commutative inverse semigroup $Pm(G)$.*

Again, as in Section 4, we define an equivalence of factor sets and call the respective quotient semigroup *the Schur multiplier $PM(G)$*. The Schur multiplier is also a commutative inverse semigroup. However, the Schur multipliers of G and of $\Sigma(G)$ are different: one can only confirm that $PM(G)$ is an image of $M(\Sigma(G))$ (Proposition 5).

At present the main problem in the description of a Schur multiplier is to find a definition of factor sets which would be independent from both PPR and $\Sigma(G)$. We recall that for the usual projective representations of groups such a definition is the cohomological equation (4). In the case of semigroups (more exactly, monoids) it is necessary to add condition (5). Unfortunately, there is no cohomological equation for PPR's. At most we can assert

PROPOSITION 6. *Let Γ be a PPR with a factor set σ. Then*

$$\forall x, y, z \in G \quad \Gamma(x)\Gamma(y)\Gamma(z) \neq 0$$
$$\Longrightarrow \quad \sigma(x, y)\sigma(xy, z) = \sigma(x, yz)\sigma(y, z).$$

Later I will give an example where the cohomological equation does not hold.

Now we consider the structure of a Schur multiplier. Certainly, it is a semi-lattice of subgroups. So first of all it is necessary to describe its idempotents:

THEOREM 12. *Let $\sigma : G \times G \to K$ be a mapping taking only the values* 0 *and* 1 *and* $\sigma(1,1) = 1$. *Then σ is a factor set if and only if*

$$\begin{aligned}&\forall x, y \in G \quad \sigma(x,y) = 1\\ &\quad \Longrightarrow \quad \sigma\big(xy, y^{-1}\big) = \sigma\big(y^{-1}, x^{-1}\big) = \sigma(x,1) = 1.\end{aligned} \tag{7}$$

EXAMPLE 4. Let $G = \langle a, b, c\rangle$ be an elementary Abelian group of order 8 with generators a, b, c. Let $H = \langle b, c\rangle$, $F = (H \setminus 1) \times (H \setminus 1) \setminus \nabla$ where ∇ is the diagonal of the Cartesian square $H \times H$. Set

$$\sigma(x,y) = \begin{cases} 1 & \text{if } (x,y) \notin F,\\ 0 & \text{if } (x,y) \in F.\end{cases}$$

It is easy to check using Theorem 12 that σ is a factor set. However,

$$\sigma(b,a)\sigma(ba,ac) = 1 \neq 0 = \sigma(b,c)\sigma(a,ac)$$

and so the cohomological equation does not hold for $x = b$, $y = a$, $z = ac$.

One can put Theorem 12 into a more general form. Consider an abstract semigroup $\mathcal{T}$ generated by elements α, β, γ with defining relations

$$\begin{cases} \alpha^2 = \beta^2 = 1, \quad (\alpha\beta)^2 = 1,\\ \gamma^2 = 1, \quad \alpha\gamma = \gamma, \quad \gamma\alpha\beta\gamma = \gamma\beta\alpha\beta, \quad \gamma\beta\gamma = 0.\end{cases}$$

For any group G the semigroup $\mathcal{T}$ acts on $G \times G$ as follows:

$$\begin{aligned}\alpha &: (x,y) \longrightarrow \big(xy, y^{-1}\big),\\ \beta &: (x,y) \longrightarrow \big(y^{-1}, x^{-1}\big),\\ \gamma &: (x,y) \longrightarrow (x,1).\end{aligned}$$

Thus, $G \times G$ turns into $\mathcal{T}$-set and Theorem 5 takes the form:

COROLLARY 6. *An idempotent mapping $\sigma : G \times G \to K$ such that $\sigma(1,1) = 1$ is a factor set if and only if* $\operatorname{supp}\sigma = \{(x,y) \mid \sigma(x,y) \neq 0\}$ *is a $\mathcal{T}$-subset in $G \times G$.*

Now from Corollaries 5 and 6 we get:

THEOREM 13. *The Schur multiplier is a semi-lattice of Abelian groups*

$$Pm(G) = \bigcup_{X \in C(G)} Pm_X(G), \qquad PM(G) = \bigcup_{X \in C(G)} PM_X(G),$$

where $C(G)$ is a semi-lattice of $\mathcal{T}$-subsets in $G \times G$ with respect to intersection.

Let me say a few words about the semigroup $\mathcal{T}$. It plays a remarkable role: for any group G it gives a description of the idempotent factor sets. Since each PLR is a PPR with an idempotent factor set, we obtain, in particular, some sort of classification of all PLR's of a group G. So $\mathcal{T}$ merits to be considered more thoroughly. Here are more details on its structure.

First of all, its order is 25. The elements α and β generate in $\mathcal{T}$ a subgroup $H = \langle\alpha, \beta\rangle$, isomorphic to the symmetric group S_3. The complement $U = \mathcal{T} \setminus H$ is an ideal.

One can prove that U is a completely 0-simple semigroup. In the standard notation of the theory of semigroups [7] it can be written as $U = M^0(D; I, \Lambda; P)$, where $I = \Lambda = \{1, 2, 3\}$, D is the group of order 2 and P is the (3×3)-sandwich-matrix,

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

7. Cohomology of small categories

In the further study of properties of 0-cohomology some difficulties arise because, as I mentioned already, 0-cohomology is not a derived functor in the Abelian category where it is constructed (see Example 3).

So a question appears: is it possible to extend the category of 0-modules so that 0-cohomology becomes a derived functor? One of the useful ways is to pass to bimodules. However, in our situation this does not help as one can see from the following example.

We call an Abelian group A by a 0-*bimodule* over S if A is right and left 0-module, and besides, $(sa)t = s(at)$ for any $s, t \in S \setminus 0$, $a \in A$. The 0-cohomology of S with values in the category of 0-bimodules is defined similarly to 0-cohomology on 0-modules. Denote it by $HH_0^n(S, A)$.

EXAMPLE 5. Let $S = \{u, v, w, 0\}$ be the commutative semigroup with multiplication $u^2 = v^2 = uv = w,\ uw = vw = 0$, M a 0-bimodule over S. Then $HH_0^2(S, M) \neq 0$ for $M \neq 0$.

As in Section 2, this example shows that in the category of 0-bimodules the cohomology functor HH_0^n is not a derived functor. This is the reason why we use the category $\mathcal{N}at\, S$ (which is defined below). Our construction is a generalization of the theory of cohomology for small categories from [2].

As in Section 4, we suppose for simplicity that S is a monoid with a zero. The *category of factorizations in* S is the category $\mathcal{F}ac\, S$ whose objects are all elements from $S \setminus 0$, and the set of morphisms $\mathrm{Mor}(a, b)$ consists of all triples (α, a, β) $(\alpha, \beta \in S)$ such that $\alpha a \beta = b$ (we will denote (α, a, β) by (α, β) if this does not lead to confusion). The composition of morphisms is defined by the rule $(\alpha', \beta')(\alpha, \beta) = (\alpha'\alpha, \beta\beta')$; hence we have $(\alpha, \beta) = (\alpha, 1)(1, \beta) = (1, \beta)(\alpha, 1)$.

A *natural system* on S is a functor $\mathbf{D} : \mathcal{F}ac\, S \to \mathcal{A}b$. The category $\mathcal{N}at\, S = \mathcal{A}b^{\mathcal{F}ac\, S}$ of such functors is an Abelian category with enough projectives and injectives [13]. Denote the value of $\mathbf{D}$ on an object $a \in \mathrm{Ob}\, \mathcal{F}ac\, S$ by $\mathbf{D}_a$. If we write $\alpha_* = \mathbf{D}(\alpha, 1)$ and $\beta^* = \mathbf{D}(1, \beta)$ then $\mathbf{D}(\alpha, \beta) = \alpha_*\beta^*$ for any morphism (α, β).

EXAMPLE 6. Each 0-module A can be considered as a functor $\mathbf{A}$ from $\mathcal{N}at\, S$, defined as follows: $\mathbf{A}_s = A$ for any $s \in S \setminus 0$ and $\alpha_*\beta^* a = \alpha a \beta$ for all $\alpha, \beta \in S$, $a \in A$.

EXAMPLE 7. Consider a functor $\mathbf{Z}$ which assigns to each object $a \in S \setminus 0$ the infinite cyclic group $\mathbf{Z}_a$ generated by a symbol $[a]$; to each morphism $(\alpha, \beta): s \to t$ it assigns a homomorphism of groups $\mathbf{Z}(\alpha, \beta): \mathbf{Z}_a \to \mathbf{Z}_b$ which takes $[a]$ to $[b]$. It is a natural system, which is called trivial.

For a given natural number n denote by $Ner_n\, S$ the set of all n-tuples $(a_1, \dots, a_n)$, $a_i \in S$, such that $a_1 \cdots a_n \neq 0$ (a *nerve* of S). For $n = 0$ we set $Ner_0\, S = \{1\}$. A mapping, defined on the nerves and assigning to each $a = (a_1, \dots, a_n)$ an element from $\mathbf{D}_{a_1 \cdots a_n}$, is called an *$n$-dimensional cochain*. The set of all n-dimensional cochains is an Abelian group $C^n(S, \mathbf{D})$ with respect to pointwise addition. Set $C^0(S, \mathbf{D}) = \mathbf{D}_1$.

Define a *coboundary homomorphism* $\Delta^n : C^n(S, \mathbf{D}) \to C^{n+1}(S, \mathbf{D})$ for $n \geqslant 1$ by the formula

$$\begin{aligned}(\Delta^n f)(a_1, \dots, a_{n+1}) = {} & a_{1*} f(a_2, \dots, a_{n+1}) \\ & + \sum_{i=1}^{n} (-1)^i f(a_1, \dots, a_i a_{i+1}, \dots, a_{n+1}) \\ & + (-1)^{n+1} a^*_{n+1} f(a_1, \dots, a_n).\end{aligned}$$

For $n = 0$ we set $(\Delta^0 f)(x) = x_* f - x^* f$ where $f \in \mathbf{D}_1$, $x \in S \setminus 0$. One can check directly that $\Delta^n \Delta^{n-1} = 0$. The cohomology groups of the complex $\{C^n(S, \mathbf{D}), \Delta^n\}_{n \geqslant 0}$ are denoted by $H^n(S, \mathbf{D})$.

The 0-cohomology of a monoid is a special case of this construction. Indeed, $H^n_0(S, A) \cong H^n(S, \mathbf{A})$, where $\mathbf{A}$ is the functor defined in Example 6.

Since $\mathcal{N}at\, S$ has enough projectives and injectives there exist derived functors $\mathrm{Ext}^n_{\mathcal{N}at\, S}(\mathbf{Z}, \_)$.

THEOREM 14. (*See* [21,22].) *For any monoid S with zero*

$$H^n(S, \_) \cong \mathrm{Ext}^n_{\mathcal{N}at\, S}(\mathbf{Z}, \_).$$

To prove this statement a projective resolution for $\mathbf{Z}$ is built in the following way.

For every $n \geqslant 0$ define a natural system $\mathbf{B}_n : \mathcal{F}ac\, S \to \mathcal{A}b$. Its value on an object $a \in S \setminus 0$ is the free Abelian group $\mathbf{B}_n(a)$ generated by the set of symbols $[a_0, \dots, a_{n+1}]$ such that $a_0 \cdots a_{n+1} = a$. To each morphism (α, β) we assign a homomorphism of groups by the formula

$$\mathbf{B}_n(\alpha, \beta) : [a_0, \dots, a_{n+1}] \to [\alpha a_0, \dots, a_{n+1} \beta].$$

These functors constitute a chain complex $\{\mathbf{B}_n, \partial_n\}_{n \geqslant 0}$, where the natural transformations $\partial_n : \mathbf{B}_n \overset{\cdot}{\to} \mathbf{B}_{n-1}$ $(n \geqslant 1)$ are given by the homomorphisms

$$\begin{aligned}&(\partial_n)_a : \mathbf{B}_n(a) \to \mathbf{B}_{n-1}(a), \\ &(\partial_n)_a [a_0, \dots, a_{n+1}] = \sum_{i=0}^{n} (-1)^i [a_0, \dots, a_i a_{i+1}, \dots, a_{n+1}].\end{aligned}$$

The natural systems $\mathbf{B}_n$ are projective objects in $\mathcal{N}at\, S$ and the complex $\{\mathbf{B}_n, \partial_n\}_{n\geqslant 0}$ is a projective resolution of the natural system $\mathbf{Z}$.

Now one can establish an isomorphism between the complexes

$$\left\{C^n(S, \mathbf{D}), \Delta^n\right\}_{n\geqslant 0} \quad \text{and} \quad \left\{\mathrm{Hom}_{\mathcal{N}at\, S}(\mathbf{B}_n, \mathbf{D}), \partial^n\right\}_{n\geqslant 0}.$$

Our construction differs from the Baues cohomology theory for monoids [2] in the initial stage only. Indeed, in [2] a monoid S is regarded as a category with a single object. At the same time the Baues category of factorizations in S is equal to $\mathcal{F}ac\, S^0$ where S^0 is a semigroup with a zero adjoined. Therefore the Baues cohomology groups of S and the cohomology groups of S^0 in our sense are the same. However if S possesses a zero element then the category $\mathcal{F}ac\, S$ and Baues one are not equivalent and we obtain different cohomology groups. The construction of this section is a generalization simultaneously of both Baues and 0-cohomology.

In conclusion I give an example of using the results obtained.

It is well known that in many algebraic theories the cohomological dimension of free objects is 1. In the category of monoids with zero a free object is a free monoid with a zero adjoined. However in this category the class of objects having cohomological dimension 1 is essentially greater.

Call every quotient monoid of a free monoid by its ideal 0-*free*. Free monoids with an adjoined zero are also considered as 0-free. Let S be a semigroup with zero. The least n such that $H_0^{n+1}(S, A) = 0$ for any 0-module A, is the 0-*cohomological dimension* (0-cd S) of S.

THEOREM 15. 0-cd $M \leqslant 1$ *for any* 0-*free monoid* M.

From this theorem there follows an interesting

COROLLARY 7. *Any projective representation of a* 0-*free monoid is linearizable* (*i.e. is equivalent to a linear one*).

In connection with Theorem 15 a question arises, an answer to which is unknown to me: is a 0-cancellative monoid (Section 6), of 0-cohomological dimension 1, 0-free?

8. Concluding remarks

The study of the general properties of 0-cohomology has only just begun. Here the same difficulties (and even greater) appear as for the EM-cohomology of semigroups. Certain expectancies are raised by Theorem 14, showing that 0-cohomology can be seen as a derived functor. However this interpretation turns out to be too vast. So it remains actual to find a category, smaller than $\mathcal{F}ac$, in which 0-cohomology would be a derived functor.

0-cohomology appears in other problems too. In an article by Clark [6] it was applied to semigroups of matrix units and algebras generated by them. A similar situation often occurs in ring theory. A quotient algebra of a semigroup algebra by the ideal generated by the zero of the semigroup, is called a *contracted* quotient algebra (in other words, the zero of the semigroup is identified with the zero of the algebra). For instance, the well-known theorem of Bautista, Gabriel, Roiter, Salmeron [3] confirms that every algebra of final type is contracted semigroup one.

A natural question arises: how does Hochschild cohomology of contracted algebras relate to the cohomology of semigroups generating them? Since it is supposed that the semigroup contains a zero, then, of course, for study of this question it is necessary to use 0-cohomology. Such an approach could be useful for incidence algebras of simplicial complexes as well (cf. [11]).

In conclusion let me mention a generalization of 0-cohomology. In a semigroup S (not necessary containing 0) fix a certain generating subset $W \subset S$ instead of $S \setminus 0$ and call mappings $W \to A$ *1-dimensional W-coboundaries*. Using this one can construct certain partial n-place mappings $S \times \cdots \times S \to A$ (and call them *n-dimensional W-cochains*) so that the coboundary homomorphisms ∂^n is well defined. I have called the obtained objects *partial cohomology* (they have no relation with the partial representations from Section 6) and considered it in [34] and [35]. Partial cohomology turned out to be useful for calculations of EM-cohomology (as in Section 3), however I have not found other applications of them.

References

[1] W.W. Adams, M.A. Rieffel, Adjoint functors and derived functors with an application to the cohomology of semigroups, J. Algebra 7 (1) (1967) 25–34.

[2] H.-J. Baues, G. Wirsching, Cohomology of small categories, J. Pure Appl. Algebra 38 (2/3) (1985) 187–211.

[3] R. Bautista, P. Gabriel, A.V. Roiter, L. Salmeron, Representation-finite algebras and multiplicative bases, Invent. Math. 81 (2) (1985) 217–286.

[4] K.S. Brown, Cohomology of Groups, Springer, Berlin, 1982.

[5] H. Cartan, S. Eilenberg, Homological Algebra, Princeton Univ. Press, Princeton, NJ, 1956.

[6] W.E. Clark, Cohomology of semigroups via topology with an application to semigroup algebras, Comm. Algebra 4 (1976) 979–997.

[7] A.H. Clifford, G.B. Preston, Algebraic Theory of Semigroups, Amer. Math. Soc., Providence, RI, 1964.

[8] C.W. Curtis, I. Reiner, Representation Theory of Finite Groups and Associative Algebras, Wiley, New York, 1967.

[9] M. Dokuchaev, B. Novikov, Projective representations and Exel's theory, in: XVIII Escola de Algebra, Campinas, July 19–23, 2004, pp. 31–32 (abstracts).

[10] R. Exel, Partial actions of groups and actions of semigroups, Proc. Amer. Math. Soc. 126 (12) (1998) 3481–3494.

[11] M. Gerstenhaber, S.D. Schack, Simplicial cohomology is Hochschild cohomology, J. Pure Appl. Algebra 30 (2) (1983) 143–156.

[12] P.A. Grillet, Commutative semigroup cohomology, Semigroup Forum 43 (2) (1991) 247–252.

[13] A. Grothendieck, Sur quelques points d'algèbre homologique, Tôhoku Math. J. 9 (2) (1957) 119–183, Tôhoku Math. J. 9 (3) (1957) 185–221.

[14] D.E. Haile, On crossed product algebras arising from weak cocycles, J. Algebra 74 (1982) 270–279.

[15] D.E. Haile, The Brauer monoid of a field, J. Algebra 81 (2) (1983) 521–539.

[16] D.E. Haile, R.G. Larson, M.E. Sweedler, A new invariant for **C** over **R**: Almost invertible cohomology theory and the classification of idempotent cohomology classes and algebras by partially ordered sets with Galois group action, Amer. J. Math. 105 (3) (1983) 689–814.

[17] D. Haile, L. Rowen, Weakly Azumaya algebras, J. Algebra 250 (2002) 134–177.

[18] V.R. Hancock, Commutative Schreier semigroup extensions of a group, Acta Sci. Math. 25 (2) (1964) 129–134.

[19] P.J. Hilton, U. Stammbach, A Course in Homological Algebra, Springer, Berlin, 1971.

[20] V.V. Kirichenko, B.V. Novikov, On the Brauer monoid for finite fields, in: Finite Fields and Applications, Augsburg, 1999, Springer, Berlin, 2001, pp. 313–318.

[21] A.A. Kostin, B.V. Novikov, Semigroup cohomology as a derived functor, Filomat 15 (2001) 17–23.

[22] A.A. Kostin, B.V. Novikov, Cohomology of semigroups and small categories, in: Algebraic Structures and Their Applications, Proc. of Ukrain. Math. Congress, Kyiv, 2002, pp. 69–75 (in Russian).

[23] H. Lausch, Cohomology of inverse semigroups, J. Algebra 35 (1–3) (1975) 273–303.

[24] J.E. Leech, Two papers: H-coextensions of monoids and the structure of a band of groups, Mem. Amer. Math. Soc. 157 (1975).

[25] M. Loganathan, Cohomology and extensions of regular semigroups, J. Austral. Math. Soc., Ser. A 35 (2) (1983) 178–193.

[26] B. Mitchell, On the dimension of objects and categories. I. Monoids, J. Algebra 9 (3) (1968) 314–340.

[27] W.R. Nico, On the cohomology of finite semigroups, J. Algebra 11 (4) (1969) 598–612.

[28] B.V. Novikov, 0-cohomology of semigroups, in: Theoretical and Applied Questions of Differential Equations and Algebra, Naukova Dumka, Kiev, 1978, pp. 185–188 (in Russian).

[29] B.V. Novikov, Projective representations of semigroups, Dokl. Akad. Nauk Ukrain. SSR Ser. A 6 (1979) 474–478 (in Russian).

[30] B.V. Novikov, 0-cohomology of completely 0-simple semigroups, Vestnik Kharkov. Gos. Univ. 221 (1981) 80–85 (in Russian).

[31] B.V. Novikov, On computing of 0-cohomology of some semigroups, Vestnik Kharkov. Gos. Univ. 221 (1981) 96 (in Russian).

[32] B.V. Novikov, A counterexample to a conjecture of Mitchell, Tr. Tbiliss. Mat. Inst. Razmadze Akad. Nauk Gruzin. SSR 70 (1982) 52–55 (in Russian).

[33] B.V. Novikov, Defining relations and 0-modules over a semigroup, in: Theory of Semigroups and Its Applications, Saratov. Gos. Univ., Saratov, 1983, pp. 94–99 (in Russian).

[34] B.V. Novikov, On partial cohomologies of semigroups, Semigroup Forum 28 (1–3) (1984) 355–364.

[35] B.V. Novikov, Partial cohomology of semigroups and its applications, Izv. Vyssh. Uchebn. Zaved. Mat. 11 (1988) 25–32 (in Russian); transl. in: Soviet Math. (Izv. VUZ) 32 (11) (1988) 38–48.

[36] B.V. Novikov, Commutative semigroups with cancellation of dimension 1, Mat. Zametki 48 (1) (1990) 148–149 (in Russian).

[37] B.V. Novikov, The Brauer monoid, Mat. Zametki 57 (4) (1995) 633–636 (in Russian); transl. in: Math. Notes 57 (3–4) (1995) 440–442.

[38] B.V. Novikov, On modifications of the Galois group, Filomat 9 (3) (1995) 867–872.

[39] B.V. Novikov, Semigroups of cohomological dimension 1, J. Algebra 204 (1998) 386–393.

[40] B.V. Novikov, The cohomology of semigroups: A survey, Fundam. Prikl. Mat. 7 (1) (2001) 1–18 (in Russian).

[41] B.V. Novikov, Semigroup cohomology and applications, in: K.W. Roggenkamp, M. Ştefănescu (Eds.), Algebra–Representation Theory, Kluwer, Dordrecht, 2001, pp. 219–234.

[42] A.M. Pachkoriya, Cohomology of monoids with coefficients in semimodules, Soobshch. Akad. Nauk Gruzin. SSR 86 (3) (1977) 546–548 (in Russian).

[43] J.C. Quigg, I. Raeburn, Characterizations of crossed products by partial actions, J. Operator Theory 37 (1997) 311–340.

[44] R. Strecker, Über kommutative Schreiersche Halbgruppenerweiterungen, Acta Math. Acad. Sci. Hung. 28 (1–2) (1972) 33–44.

[45] M.E. Sweedler, Weak cohomology, in: Contemp. Math., vol. 13, Amer. Math. Soc., Providence, RI, 1982, pp. 109–119.

[46] Ch. Wells, Extension theories for monoids, Semigroup Forum 16 (1) (1978) 13–35.

This page intentionally left blank

Section 3B
Associative Rings and Algebras

This page intentionally left blank

Finite Rings with Applications

Alexandr A. Nechaev[*]

Center of New Informational Technologies, *Moscow State University*, *Russia*
E-mail: *nechaev@cnit.msu.ru*

Contents

[*] This work is supported by President of Russian Federation and RFFR grants: NSh-8564.2006.10, NSh-5666.2006.1, 05-01-01018, 05-01-01048.

HANDBOOK OF ALGEBRA, VOL. 5
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(07)05005-X

Abstract

The main classes of finite rings (f.r.) and modules interesting for applications are considered: Wedderburn rings, local, chain and Galois rings, (quasi)-Frobenius bimodules. Polynomials, functions, identities, matrices and linear substitutions over commutative chain f.r. (GE-rings) are described. As applications the results about standard bases of polynomial ideals, systems of polynomial equations, periodic properties of polynomial ideals are presented. Properties of matrices, linear sequences and (poly-)linear recurrences over GE-rings and Galois rings are shown. We state also the main results of the general theory of linear codes over finite modules and their representations.

Keywords: nil-radical, Jacobson radical, modular radical, Wedderburn radical, Wedderburn ring, local ring, Hensel lemma, principal ideal ring, chain ring, Galois ring, GE-ring, GEO-ring, quasi-Frobenius ring, symmetric ring, nilpotent ring, character module, quasi-Frobenius

bimodule, Frobenius bimodule, coordinate field, canonical decomposition, trace function, coordinate function, equation with trace, generalized Galois ring, congruence subgroup, polynomial transformation, polynomial identity, Cross variety, complete system of functions over ring, standard base of polynomial ideal, coordinated standard base, canonical generating system, period of the polynomial ideal, distinguished polynomial, polynomial of maximal period, similar matrices, normal matrix, canonically defined matrix, polynomially defined matrix, Fitting invariants, Kurakin invariants, cyclic type of linear substitution, poly-linear recurrence, LRS-family, full-cycle recurrence, k-maximal recurrence, coordinate sequence, rank, linear complexity, recurrence of maximal period, frequency characteristic, pseudo-random sequence, linear code, socle of a linear code, dual code, MacWilliams identity, equivalent codes, extension theorem, egalitarian weight, homogeneous weight, scaled isometry, presentation of a code, generalized Kerdock code, Golay code, MDS-code, linearly optimal code, loop-code, group-code

This page intentionally left blank

Introduction

The class of finite rings is interesting as the first natural class of rings which allows to bring forth problems and conjectures, check validity and demonstrate the efficiency of results in general theory of rings. Moreover some results of the general theory of rings allow interesting revisions for finite rings, which in general are not true. Finally, in the last 20–30 years increased interest in possible application of finite rings, different from the fields, in coding theory and cryptography.

In the first part of this chapter, we present the main results on the structure and properties of finite rings (f.r.), as well as natural closely related questions: linear algebra problems in finite modules, description of identities and polynomial functions on finite rings.

The second part is devoted to applications of the theory of finite rings and modules in the theory of linear codes and (poly-)linear (recurrent) sequences which are important, in particular, in cryptography.

A need to present the available material in a compact form has naturally limited the text to the list of results that are most interesting to the author and important for applications. So, finite fields and non-associative rings were left out of the frame of our presentation. The exceptions are a few short sections devoted to non-associative Galois rings and quasigroup codes, which appear to have good perspectives in applications.

Taking this into account, the bibliography is rather interesting by itself, since it contains practically all publications on finite associative rings that the author is aware of.

I. Finite rings and modules

1. Some properties of finite rings connected with radicals

Here some classical results on Artinian rings are made more precise in the special context of finite rings.

1.1. *Nil-radical, semisimple rings [7,17–21,62,63,136,242,244,246,275,313,386,387]*

Let us recall that a left ideal I of a ring R is called a *nil-ideal* if any element of I is nilpotent. For a f.r. R any nil-ideal is nilpotent and a sum of left nil-ideals is a nil-ideal; moreover the sum $\mathfrak{N}(R)$ of all left nil-ideals of R is a two-sided nilpotent ideal, containing all right nil-ideals of R. This ideal is called *nil-radical* of R, it agrees with the Jacobson radical [184] of the ring R (the intersection of all left (right) maximal ideals of R) and has the following properties:

$$\begin{aligned}&\mathfrak{N}\big(\mathfrak{N}(R)\big) = \mathfrak{N}(R), \qquad \big(\overline{R} = R/\mathfrak{N}(R)\big) \quad \Rightarrow \quad \big(\mathfrak{N}(\overline{R}) = 0\big),\\ &(I \lhd R) \quad \Rightarrow \quad \big(\mathfrak{N}(I) = I \cap \mathfrak{N}(R)\big).\end{aligned} \tag{1.1}$$

A ring R is called *semisimple* if $\mathfrak{N}(R) = 0$. So the *top-factor* $\overline{R} = R/\mathfrak{N}(R)$ of any finite ring R is a semisimple ring. First there is the following classical result [17,275,387].

THEOREM 1.1 *(Molin, Wedderburn, Artin). If R is a f.r. and $\overline{R} \neq 0$ then $\overline{R}$ is a direct sum of full matrix rings over fields*:

$$\overline{R} = M_{n_1}(P_1) \oplus \cdots \oplus M_{n_t}(P_t), \quad P_s = GF(q_s),\ s \in \{1, \ldots, t\}. \tag{1.2}$$

A f.r. R is called *primary* if it contains the identity and in (1.2) $t = 1$; it is called *completely primary* (or *local*) if it is primary and in (1.2) $t = 1, n_1 = 1$.

Related sources: [56,90,115,118,165,172–174,185,186,194,209,216,242,243,338,358, 373,397,402].

1.2. *Quasi-identities and the modular radical [273]*

In general we cannot assert that any f.r. R with $\overline{R} \neq 0$ contains the identity. However there are some main approximations to an identity of R.

Let $\lambda_R(S)$, $\rho_R(S)$ be correspondingly the left and the right annihilators in R of a subset $S \subseteq R$, let $D_{\mathrm{l}}(R)$, $D_{\mathrm{r}}(R)$ be correspondingly the sets of all left and right zero divisors of R including 0, and $D(R) = D_{\mathrm{r}}(R) \cup D_{\mathrm{l}}(R)$.

A *quasi-identity* is an idempotent $e = e^2 \in R$ such that $\lambda_R(e) \cup \rho_R(e)$ does not contain nonzero idempotents.

PROPOSITION 1.2. *Any f.r. R contains a quasi-identity. An idempotent $e \in R$ is a left identity of R exactly if $\rho_R(e) = 0$. The ring R contains a left identity exactly if the set $R \setminus D_{\mathrm{l}}(R)$ of left regular elements of R is nonempty. If the set of left identities of R is nonempty then it is the set of all quasi-identities of R.*

Let $R \neq \mathfrak{N}(R)$ and e be a quasi-identity of R. Then $e \neq 0$, and the image $\bar{e}$ of e under the natural epimorphism $R \to \overline{R}$ is an identity of $\overline{R}$.

PROPOSITION 1.3. *Let R be a nonzero f.r. with a quasi-identity e. Then the following conditions are equivalent.*

(a) *R has identity*;

(b) *e is identity of R*;

(c) $\rho_R(e) = \lambda_R(e) = 0$;

(d) *the set $R \setminus D(R)$ of regular elements of the ring R is nonempty.*

Under these conditions the relations $D_{\mathrm{r}}(R) = D_{\mathrm{l}}(R) \neq R$ hold and the multiplicative group R^* (*the set of invertible elements*) *of the ring R satisfies the equality $R^* = R \setminus D_{\mathrm{r}}(R)$.*

Let $\mathcal{M}(R) \lhd R$ be the two-sided ideal generated by $\lambda_R(e) \cup \rho_R(e)$.

PROPOSITION 1.4. *For any ideal $I \lhd R$ the quotient ring R/I contains an identity if and only if $\mathcal{M}(R) \subseteq I$.*

Hence the ideal $\mathcal{M}(R)$ does not depend on the choice of the quasi-identity e.

PROPOSITION 1.5. *For a f.r. R with quasi-identity e there are equivalent*:

(a) *e is a unique quasi-identity of R*;

(b) $\lambda_R(e) = \rho_R(e)$;
(c) $R = eRe \oplus \mathcal{M}(R)$.

We call $\mathcal{M}(R)$ the *modular radical* of the ring R. The function $\mathcal{M}$ on the class of all finite ring satisfies the conditions

$$\mathcal{M}\big(\mathcal{M}(R)\big) = \mathcal{M}(R), \qquad \mathcal{M}\big(R/\mathcal{M}(R)\big) = 0,$$
$$(I \lhd R) \quad \Rightarrow \quad \big(\mathcal{M}(R/I) \subseteq \big(\mathcal{M}(R) + I\big)/I\big).$$

Related sources: [30,31,94,99,337,338].

1.3. *Wedderburn radical, W-rings [273,313]*

Under condition (1.2) there exists an orthogonal system of idempotents $e_1, \dots, e_t \in R$ such that $\bar{e}_s$ is the identity of $M_{n_k}(P_k)$, $k \in \{1, \dots, t\}$. Then $e = e_1 + \cdots + e_t$ is a quasi-identity of R and the latter is a direct sum of subgroups:

$$R = R_1 \oplus \cdots \oplus R_t \oplus \Delta \quad \text{(Peirce decomposition)},$$

where $R_k = e_k R e_k = M_{n_k}(S_k)$, S_k is a local f.r., $\overline{S_k} = P_k$, for $k \in \{1, \dots, t\}$, and

$$\Delta = \sum_{i \neq j} e_i R e_j \oplus eR(1-e) \oplus (1-e)Re \oplus (1-e)R(1-e) \subseteq \mathfrak{N}(R). \qquad (1.3)$$

Note that $\Delta = 0$ if and only if $e_1, \dots, e_t$ are central idempotents and e is identity of the ring R. A f.r. R is called a *Wedderburn*- or *W-ring*, if $R = 0$, or

$$R = M_{n_1}(S_t) \oplus \cdots \oplus M_{n_t}(S_t), \quad t \geqslant 1, \qquad (1.4)$$

is a direct sum of full matrix rings over local rings. Note that any item in (1.4) is a primary ring. Moreover

THEOREM 1.6. *(See* [187,244].*) A f.r. R is a primary exactly if $R = M_n(S)$, where $n \in \mathbb{N}$ and S is a local ring. Under the lust condition if, in addition, $R = M_k(T)$, where T is a local ring, then $k = n$, $T \cong S$.*

So W-rings are exactly direct sums of primary rings.

THEOREM 1.7. *(See* [273,286].*) For a f.r. R with left identity the following properties equivalent*:

(a) *R is a Wedderburn ring*;
(b) *any two-sided idempotent ideal of R is left principal*;
(c) *for any idempotent $f \in R$ the ideal RfR is left principal.*

COROLLARY 1.1. *Any commutative f.r. R with identity has unique up to permutation decomposition in direct sum of local rings.*

THEOREM 1.8. *Any f.r. R contains an ideal $\mathcal{W} = \mathcal{W}(R)$ such that for any ideal $I \lhd R$ the quotient ring R/I is a W-ring if and only if $\mathcal{W} \subseteq I$. This ideal $\mathcal{W}(R)$ is generated by the set* (1.3).

We call $\mathcal{W}(R)$ *Wedderburn radical* of the ring R. There are the inclusions

$$\mathcal{M}(R) \subseteq \mathcal{W}(R) \subseteq \mathfrak{N}(R) \quad \text{(each can be strict).}$$

Related sources: [207,268,375].

2. Local finite rings

2.1. *Characterizations and parameters of local rings [64,75,119,140,152,204,212,213,280,326,357,365,377–379]*

So the description of finite rings with identity is reduced modulo some nilpotent ideal to the description of local finite rings.

THEOREM 2.1. *For a f.r. S with identity there are equivalent*:

(a) *S is a local f.r.*;
(b) *S does not contain proper idempotents*;
(c) *$S \setminus S^*$ is a subgroup of the group $(S, +)$*;
(d) *$\mathfrak{N}(S) = S \setminus S^*$.*

THEOREM 2.2. *Let S be a local f.r. Then $\bar{S} = GF(q)$, $q = p^r$, for some prime p and $r \in \mathbb{N}$. Let n be the nilpotency index of the ideal $\mathfrak{N}(S)$. Then S contains a strictly descending chain of the ideals*

$$S \rhd \mathfrak{N}(S) \rhd \cdots \rhd \mathfrak{N}(S)^t \rhd \cdots \rhd \mathfrak{N}(S)^{n-1} \rhd 0, \tag{2.1}$$

and satisfies the conditions

$$\operatorname{char} S = p^d, \qquad |S| = q^c, \qquad \left|\mathfrak{N}(S)\right| = q^{c-1}, \quad d \leqslant n \leqslant c. \tag{2.2}$$

The quotients $\mathfrak{N}_t = \mathfrak{N}(S)^t/\mathfrak{N}(S)^{t+1}$, $t \in \{0, 1, \dots, n-1\}$, are left and right spaces over the field $\bar{S} = GF(q)$ and the finiteness of $\bar{S}$ implies

$$\dim{}_{\bar{S}}\mathfrak{N}_t = \dim \mathfrak{N}_{t\,\bar{S}} = m_t, \quad t \in \{1, \dots, n-1\}.$$

The parameters $m_0, \dots, m_{n-1}$ are called the *Loewy invariants* of a local f.r. S. We have:

$$\left|\mathfrak{N}(S)^t\right| = q^{m_t + \cdots + m_{n-1}}, \quad t \in \{1, \dots, n-1\}. \tag{2.3}$$

Commutative local rings were first studied by Krull [212–215].

Examples of local f.r. are: $GF(q)$; $\mathbb{Z}_{p^n}$; $GF(q)[x]/(f(x))$, where $f(x) \in GF(q)[x]$ is a primary polynomial (a power of an irreducible polynomial); and, more general, finite chain rings (see below).

Up to now (2007) there is no full description of all isomorphism classes of local finite rings.

2.2. *Some properties of the multiplicative group of a local f.r. [25,26,296]*

Let S be a ring with parameters as described in the last theorem. Then $S^* = S \setminus \mathfrak{N}(S)$ and $|S^*| = (q-1)q^{c-1}$. Consider the natural epimorphism $\psi : S^* \to \bar{S}^* = GF(q)^*$. Then

$\operatorname{Ker}\psi = e + \mathfrak{N}$ is a normal subgroup of the group S^* of order $|e + \mathfrak{N}| = |\mathfrak{N}| = q^{c-1} = p^{r(c-1)}$. It is Sylow p-subgroup of the group S^*, also called a *congruenc-subgroup*.

THEOREM 2.3. *The group S^* is a semidirect product*:

$$S^* = \bigl(e + \mathfrak{N}(S)\bigr) \leftthreetimes G, \tag{2.4}$$

where G is a cyclic subgroup of S^* of order $q-1$ mapping onto $\bar{S}^*$ under the epimorphism ψ. There are the equalities

$$\exp S^* = (q-1)p^{\mu(S)}, \quad \textit{where } p^{\mu(S)} = \exp(e + \mathfrak{N}) \tag{2.5}$$

for the exponents of the groups S^* and $e + \mathfrak{N}$.

The structure of a congruence subgroup and even its exponent are not defined uniquely by the numerical parameters of the ring S that were discussed above. There is only the following general upper estimate.

LEMMA 2.4. *Let S be any ring with identity e of characteristic p^d and I be a nilpotent ideal of S of nilpotency index n. Then $e + I$ is a subgroup of S^* and* $\exp(e+I) \mid p^{\omega}$, *where ω is the minimal $t \in \mathbb{N}$ with the property $I^{p^t} = pI^{p^{t-1}} = \cdots = p^t I = 0$.*

Here if

$$\nu =]\log_p n[, \quad \text{and} \quad \operatorname{char} I^{p^s} = p^{\omega_s}, \quad \text{for } s \in \{0, \dots, \nu\}, \tag{2.6}$$

then

$$\omega = \max\bigl\{s + \omega_s\colon s \in \{0, \dots, \nu\}\bigr\} \leqslant w_0 + \nu - 1. \tag{2.7}$$

Here $]x[$ is a minimal $m \in \mathbb{N}$ with property $x \leqslant m$.

PROPOSITION 2.5. *Let S be a local f.r. with parameters as fixed in Theorem 2.1 and let $\nu =]\log_p n[$, $\operatorname{char} \mathfrak{N}^{p^s} = p^{d_s}$ for $s \in \{0, \dots, \nu\}$, $\omega = \max\{s + d_s\colon s \in \{0, \dots, \nu\}\}$. Then*

$$\exp(e + \mathfrak{N}) = p^{\mu(S)}, \quad \mu(S) \leqslant \omega \leqslant d_0 + \nu - 1. \tag{2.8}$$

In general case the estimate $p^{d_0+\nu-1}$ for $\exp(e + \mathfrak{N})$ is precise. For example if $S = \mathbb{Z}_{p^2}[x]/(x^2)$, $p \geqslant 3$, then $\mathfrak{N} = (p, x)$, $n - 3$, $\nu = 1$, $d_0 = 2$ and $d_0 + \nu - 1 = 2$. Moreover, the order of the element $e + x \in e + \mathfrak{N}$ equals p^2 and therefore $\exp(e + \mathfrak{N}) = p^2$.

The value of the parameter ω in the last proposition can be made more precise under some additional condition on the ring S. We call a local f.r. *balanced* if for some $\varepsilon \in \mathbb{N}$ the following equality holds

$$\mathfrak{N}^{\varepsilon} = pS \quad (p = \operatorname{char} \bar{S}). \tag{2.9}$$

The least ε satisfying (2.9) is called the *ramification index* of the ring S. The class of balanced ring is a big enough. In particular, such are all rings of the characteristic p (then $\varepsilon = n$) and all local principal ideal (chain) f.r. (see below). Note that under condition (2.9) the parameters n and d are connected by the relations $n \geqslant \varepsilon$, $d =]\frac{n}{\varepsilon}[$.

THEOREM 2.6. *Let S be a balanced local f.r. with ramification index ε. Then*

$$\exp(e + \mathfrak{N}) = p^{\mu(S)}, \quad \mu(S) \leqslant \omega, \quad \textit{where}$$
$$\omega = \left]\frac{n - p^b}{\varepsilon}\right[+ b, \quad b = \max\left\{0, \left]\log_p \frac{\varepsilon}{p-1}\right[\right\}. \tag{2.10}$$

2.3. *Polynomials over a commutative local f.r. [53,170,212,213,262,280,284,357,403]*

Here R is a commutative local f.r. with nilradical $\mathfrak{N} = \mathfrak{N}(R)$ of nilpotency index n; and $\mathcal{R} = R[x]$ is the ring of polynomials in one variable over R.

2.3.1. *Invertible polynomials, Krull theorem* The canonical epimorphism $\nu : R \to \overline{R}$ implies an epimorphism $\sigma : \mathcal{R} \to \overline{\mathcal{R}}$, such that image of any polynomial $F(x) = \sum f_i x^i \in \mathcal{R}$ is $\sigma(F(x)) = \overline{F}(x) = \sum \overline{f}_i x^i \in \overline{\mathcal{R}}$. It is evident that $\operatorname{Ker}\sigma = \mathfrak{N}[x]$ consists of all polynomials with coefficients from $\mathfrak{N}$ and $\operatorname{Ker}\sigma = \mathfrak{N}(\mathcal{R})$ the nilradical of $\mathcal{R}$. The reduction from $\mathcal{R}$ to the ring $\overline{\mathcal{R}} = \overline{R}[x]$ of polynomials over the field $\overline{R}$ is one of the main approaches to the investigation of the ring $\mathcal{R}$.

PROPOSITION 2.7. *A polynomial $U(x) = \sum u_s x^s \in \mathcal{R}$ is invertible in the ring $\mathcal{R}$ if and only if* $\deg \overline{U}(x) = 0$, *i.e.* $u_0 \in R^*$, $u_s \in \mathfrak{N}$ *for* $s \geqslant 1$.

The following result generalizes the well-known fact that any nonzero polynomial over a field has a unique monic polynomial associated to it.

THEOREM 2.8 *(Krull). Let $H(x) = \sum h_s x^s \in \mathcal{R}$ and $\overline{H}(x) \neq 0$. Then there exists a unique invertible polynomial $U(x) \in \mathcal{R}^*$ and monic polynomial $F(x) \in \mathcal{R}$ such that*

$$U(x)H(x) = F(x). \tag{2.11}$$

Moreover $\deg F(x) = k$, *where* $k = \max\{l \in \mathbb{N}_0\colon\ h_l \in R^*\}$.

2.3.2. *Coprime polynomials* Let $(F_1, \ldots, F_t)_{\mathcal{R}}$ be the ideal generated in $\mathcal{R}$ by polynomials $F_1, \ldots, F_t \in \mathcal{R}$, and let $(\overline{F}_1, \ldots, \overline{F}_t)$ be the gcd of the system of the polynomials $\overline{F}_1, \ldots, \overline{F}_t$ over the field $\overline{R}$. The polynomials $F_1, \ldots, F_t \in \mathcal{R}$ are said to be *coprime* if $(F_1, \ldots, F_t)_{\mathcal{R}} = (e)_{\mathcal{R}} = \mathcal{R}$. The last condition is equivalent to the existence of polynomials $A_1, \ldots, A_t \in \mathcal{R}$ such that

$$A_1(x)F_1(x) + \cdots + A_t(x)F_t(x) = e. \tag{2.12}$$

PROPOSITION 2.9. *For any polynomials $F_1, \ldots, F_t \in \mathcal{R}$ there is the equivalence*

$$(F_1, \ldots, F_t)_{\mathcal{R}} = (e)_{\mathcal{R}} \quad \Leftrightarrow \quad (\overline{F}_1, \ldots, \overline{F}_t) = \bar{e}.$$

Some of the properties of the coprime polynomials over a field still hold for coprime polynomials over any commutative ring.

PROPOSITION 2.10. *Let R be an arbitrary commutative ring with identity. Then for any polynomials $A(x), B(x), C(x) \in \mathcal{R}$ the following conditions are true.*

(a) *If $(A(x), B(x))_{\mathcal{R}} = (e)_{\mathcal{R}}$ and $(A(x), C(x))_{\mathcal{R}} = (e)_{\mathcal{R}}$, then $(A(x), B(x)C(x))_{\mathcal{R}} = (e)_{\mathcal{R}}$.*

(b) *If $(A(x), B(x))_{\mathcal{R}} = (e)_{\mathcal{R}}$ and $A(x) \mid B(x)C(x)$, then $A(x) \mid C(x)$.*

(c) *If $(A(x), B(x))_{\mathcal{R}} = (e)_{\mathcal{R}}$, $A(x) \mid C(x)$ and $B(x) \mid C(x)$, then $A(x)B(x) \mid C(x)$.*

2.3.3. *Hensel lemma and canonical decomposition [53,262,280,284,403]* The following important result gives an analog of the canonical decomposition of polynomials over a field to polynomials over a local commutative f.r..

THEOREM 2.11 *(Hensel lemma). Let $F(x) \in \mathcal{R}$ be a monic polynomial and let $g(x), h(x) \in \overline{R}[x]$ be monic polynomials such that $\overline{F}(x) = g(x)h(x)$, $(g(x), h(x)) = \bar{e}$. Then there exists a unique pair of monic polynomials $G(x), H(x) \in \mathcal{R}$ with the properties*

$$F(x) = G(x)H(x), \qquad \overline{G}(x) = g(x), \qquad \overline{H}(x) = h(x). \tag{2.13}$$

A monic polynomial $F(x) \in \mathcal{R}$ is called *primary* (*with base* $g(x) \in \overline{\mathcal{R}}$) if $\overline{F}(x) = g(x)^k$, where $g(x)$ is an irreducible polynomial over $\overline{R}$.

THEOREM 2.12. *Any polynomial $F(x) \in \mathcal{R}$ such that $\overline{F}(x) \neq \bar{0}$ has a unique (up to a permutation of the factors) decomposition into a product*

$$F(x) = U(x)F_1(x) \cdots F_t(x), \ t \geqslant 0, \tag{2.14}$$

where $U(x)$ is invertible and (for $t > 1$) $F_1(x), \ldots, F_t(x)$ are primary pairwise coprime monic polynomials.

We call the decomposition (2.14) the *canonical decomposition* of $F(x)$.

Of course any monic polynomial $F(x)$ can be presented as product of monic irreducible polynomials: $F(x) = G_1(x) \cdots G_r(x)$. However this decomposition is not only nonunique, but even can consists of a different number of factors. For example if $R = \mathbb{Z}_4$ then

$$x^4 = x \cdot x \cdot x \cdot x = \bigl(x^2 + 2\bigr)\bigl(x^2 + 2\bigr) = \bigl(x^2 + 2x + 2\bigr)\bigl(x^2 + 2x + 2\bigr).$$

2.3.4. *Lifting of divisors and roots modulo the radical* Let us say that a *polynomial $G(x) \in \mathcal{R}$ lies over the polynomial $g(x) \in \overline{\mathcal{R}}$* or *$G(x)$ is a lifting of the polynomial $g(x)$* from $\overline{\mathcal{R}}$ to $\mathcal{R}$, if $\overline{G}(x) = g(x)$. We say that a polynomial $G(x)$ is a *full divisor* of the polynomial $F(x)$ if $F(x) = G(x)H(x)$ and $(G(x), H(x))_{\mathcal{R}} = \mathcal{R}$.

PROPOSITION 2.13. *Any monic full divisor of a polynomial $\overline{F}(x)$ can be uniquely lifted to a monic full divisor of the polynomial $F(x)$ in the ring $\mathcal{R}$.*

PROPOSITION 2.14. *Let $F(x) \in \mathcal{R}$ and $a \in \overline{R}$ be such that $\overline{F}(a) = 0$ and $\overline{F}'(a) \neq 0$. Then element a uniquely lifts up to a root α of the polynomial $F(x)$ in the ring R.*

PROPOSITION 2.15. *Let $F(x) \in \mathcal{R}$ be a monic polynomial of degree m such that $\overline{F}(x)$ has m different roots in the field $\overline{R}$. Then the polynomial $F(x)$ has a unique* (*up to permutation*) *decomposition in linear pairwise coprime factors over R.*

2.4. *Coordinate field of a commutative local f.r.*

Let $\overline{R} = GF(q)$. Consider the polynomial $F(x) = x^q - x \in \mathcal{R}$. It is well known that

$$\overline{F}(x) = \prod_{a \in \overline{R}} (x - a).$$

Together with Proposition 2.15 this implies that the set $\Gamma(R) = \{\alpha \in R\colon \alpha^q = \alpha\}$ of all roots in R of the polynomial $x^q - x$ consists of q elements, all different modulo $\mathfrak{N}$, and

$$x^q - x = \prod_{\alpha \in \Gamma(R)} (x - \alpha).$$

Evidently $\Gamma(R)$ is a subsemigroup of the semigroup $(R, \cdot)$ containing 0. The set $\Gamma(R)^* = \Gamma(R) \setminus 0$ is the set of all roots of the polynomial $x^{q-1} - e$ in the ring R and a subgroup of the group R^* of order $q - 1$.

The map $\mu : \Gamma(R) \to \overline{R}$ given by the rule $\mu(\alpha) = \bar{\alpha}$ is an isomorphism of multiplicative semigroups, which induces an isomorphism of groups $\mu : \Gamma(R)^* \to \overline{R}^*$. So we have

PROPOSITION 2.16. *For any element $\alpha \in R$ there exists a unique element $\gamma(\alpha) \in \Gamma(R)$ with the property $\bar{\alpha} = \overline{\gamma(\alpha)}$. The function $\gamma(x)$ on R is given by a polynomial*: $\gamma(x) = x^{p^{d-1}}$, *where* $p^d = \operatorname{char} R$.

The set $\Gamma(R)$ is called the (*Teichmüller*) *coordinate set* of the ring R.

Since $\overline{R}^* = \langle a \rangle$ is a cyclic group the decomposition of the polynomial $x^{q-1} - \bar{e}$ over the field $\overline{R}$ is

$$x^{q-1} - \bar{e} = \prod_{i=0}^{q-2} (x - a^i).$$

If $\alpha \in \Gamma(R)^*$ is a lift of a then α is a cyclic generator of the group $\Gamma(R)^*$ and

$$x^{q-1} - e = \prod_{i=0}^{q-2} (x - \alpha^i).$$

Let us consider a binary operation $\oplus$ on the semigroup $(\Gamma(R), \cdot)$ given by

$$\alpha \oplus \beta = \gamma(\alpha + \beta), \quad \alpha, \beta \in \Gamma(R). \tag{2.15}$$

PROPOSITION 2.17. *The algebra $(\Gamma(R), \oplus, \cdot)$ is a field of q elements. The map μ is an isomorphism of the algebra $(\Gamma(R), \oplus, \cdot)$ to the field $(\overline{R}, +, \cdot)$.*

3. Principal ideal rings [111,181,190,191,284,317,375]

A f.r. R with identity is called *left principal ideal ring* (*left PIR*) if any left ideal $I \leqslant {}_R R$ is left principal: $I = Ra$; it is called *principal ideal ring* (*PIR*) if it is a left PIR and a right PIR.

3.1. *General construction*

THEOREM 3.1. *(See [284].) For a f.r. R with identity the following are equivalent*:

(a) *R is a left PIR*;
(b) *any two-sided ideal of R is left principal*;
(c) *R is a PIR*;
(d) *R is a W-ring such that the rings $S_1, \dots, S_t$ in (1.4) are local PIR.*

The equivalence of (a) and (c), together with Theorem 1.7, give interesting pair of conditions of type

$$\text{LEFT} \quad \Rightarrow \quad \text{RIGHT},$$

which are specific for finite rings. Note that if R is an Artinian ring then conditions (a) and (c) are not equivalent, [190].

About finite principal ideal rings without identity see [265].

3.2. *Chain rings [40,72,93,124–127,129,137,188,197,282,284,319,353–355]*

A ring S is called a *left chain ring* if the lattice of all its left ideals is a chain; it is called *chain ring* if S is a left and right chain ring.

THEOREM 3.2. *For a f.r. S with identity the following conditions are equivalent*:

(a) *S is a local PIR*;
(b) *S is a chain ring*;
(c) *S is a left chain ring*;
(d) *S is a local f.r. with Loewy invariants $m_1 = \cdots = m_n = 1$*;
(e) *S is a local f.r. with $m_1 = 1$*;
(f) *S is a local f.r. and $\mathfrak{N}(S)$ is a left principal ideal*;
(g) *the lattice of all one-sided ideals of S is a chain $S \triangleright \mathfrak{N}(S) \triangleright \cdots \triangleright \mathfrak{N}(S)^{n-1} \triangleright 0$, for some $n \in \mathbb{N}$*;

Under condition (g), *if $n > 1$ then $\mathfrak{N}^i = S\pi^i = \pi^i S$, for any $\pi \in \mathfrak{N} \setminus \mathfrak{N}^2$, $i \in \{0, \dots, n\}$.*

3.3. *π-Adic decomposition in a chain ring*

In accordance with Theorem 2.3, the multiplicative group S^* of the finite chain ring S has a decomposition (2.4). Using the subgroup G of this decomposition we can state that the set $\Gamma = G \cup \{0\}$ is a subsemigroup of $(S, \cdot)$, and that the canonical homomorphism $\varphi : S \to \overline{S}$ induces an isomorphism $\Gamma \to \overline{S}$.

PROPOSITION 3.3. *Let S be a finite chain ring with $n > 1$ and $\pi \in \mathfrak{N} \setminus \mathfrak{N}^2$. Then any element $a \in S$ has unique representation in the form*

$$a = a_0 + a_1\pi + \cdots + a_{n-1}\pi^{n-1}, \quad a_0, \ldots, a_{n-1} \in \Gamma. \tag{3.1}$$

We call (3.1) a *π-adic decomposition* of the element a (*relative to the coordinate set Γ and uniformizing element π*).

Note that if S is a commutative chain f.r. then the decomposition (2.4) is a direct product, G is the unique subgroup of S^* of order $q - 1$ and $\Gamma = \Gamma(S)$ is the Teichmüller coordinate set of the ring S (see Section 2.4).

4. Galois rings [111,189,213,217,218,226,239,240,262,284,294,320,362]

These rings are the simplest, most investigated, and the most important in the theory and applications of finite chain rings.

4.1. *Equivalent definitions and constructions of Galois rings*

If R is a f.r. with identity then $R \setminus R^*$ is the set of all two-sided zero divisors of R.

A f.r. R with identity e is called a *Galois ring* (*GR*) if $R \setminus R^* = \lambda R$ for some $\lambda \in \mathbb{N}$. This is the short definition of a GR. More useful for applications is the definition given by the following theorem.

THEOREM 4.1. *A f.r. R is a* GR *if and only if R is a commutative local* PIR *such that $\mathfrak{N}(R) = pR$ for some prime p. The lattice of all ideals of R is a chain*

$$R \triangleright \mathfrak{N} = pR \triangleright \cdots \triangleright \mathfrak{N}^{n-1} = p^{n-1}R \triangleright \mathfrak{N}^n = p^n R = 0,$$

for some $n \in \mathbb{N}$; the top factor of R is $\overline{R} = GF(q)$, where $q = p^r$, for some $r \in \mathbb{N}$, and

$$\begin{aligned} &\mathrm{char} R = p^n, \qquad |R| = q^n, \\ &|R^*| = q^{n-1}(q-1), \qquad \left|p^i R\right| = q^{n-i}, \quad i \in \{0, 1, \ldots, n\}. \end{aligned}$$

The simplest examples of GR' are the $GF(q)$, $\mathbb{Z}_{p^n}$. As for Galois fields we have

THEOREM 4.2. *For every prime $p \in \mathbb{N}$ and for any $n, r \in \mathbb{N}$ there exists a unique up to isomorphism* GR *R of characteristic p^n consisting of q^n elements, where $q = p^r$.*

The main part of the proof of the last theorem is formed by the following *construction of a GR*. Let R be a GR. A monic polynomial $F(x) \in \mathcal{R}$ is called a *Galois polynomial* (*GP*) if its image $\overline{F}(x)$ under the natural epimorphism $R \to \overline{R} = R/pR = GF(q)$ is irreducible in $\overline{R}[x]$. By a standard way we can consider R as a subring of the ring $S = \mathcal{R}/F(x)\mathcal{R}$.

THEOREM 4.3. *Let R be a* GR *of the characteristic p^n consisting of q^n elements, and let $F(x) \in \mathcal{R}$ be a* GP *of the degree m. Then $S = \mathcal{R}/F(x)\mathcal{R}$ is a* GR *of characteristic p^n consisting of q^{mn} elements. This ring contains all roots of any* GP *$G(x) \in \mathcal{R}$ satisfying the condition* $\deg G \mid m$. *For $\xi \in S$ the equality $S = R[\xi]$ holds iff ξ is a root of some* GP *$G(x) \in \mathcal{R}$ of degree m. The ring S does not contains a root of* GP *$G(x) \in \mathcal{R}$ such that* $\deg G \nmid m$.

We call S a *Galois extension of the GR R of degree m.*

Now it is correct to use notation: $R = GR(q^n, p^n)$ (in some articles $R = GR(r, p^n)$) for a Galois ring R of characteristic p^n consisting of q^n elements. So we have: $GF(q) = GR(q, p)$, $\mathbb{Z}_{p^n} = GR(p^n, p^n)$ and for any $r \in \mathbb{N}$ the ring $R = GR(q^n, p^n)$ with $q = p^r$ can be constructed in the form:

$$R = \mathbb{Z}_{p^n}[x]/F(x)\mathbb{Z}_{p^n}[x],$$

where $F(x)$ is a GP of the degree r over $\mathbb{Z}_{p^n}$. So any GR $R = GR(q^n, p^n)$, $q = p^r$, is a Galois extension of the degree r of its subring $\mathbb{Z}_{p^n}$ generated by the identity.

4.2. *p-Adic decomposition of elements*

In accordance with the results of Section 3.3 for any $a \in R$ there exists a unique element $\gamma(a) \in \Gamma(R) = \{\alpha \in R\colon \alpha^q = \alpha\}$ such that $a \equiv \gamma(a) \pmod{pR}$. This implies that any $a \in R$ has unique *p-adic decomposition*:

$$a = a_0 + a_1 p + \cdots + a_{n-1}p^{n-1}, \quad a_s = \gamma_s(a) \in \Gamma(R),\ s \in \{0, 1, \ldots, n-1\}. \tag{4.1}$$

We call $\gamma_s : R \to \Gamma(R)$ *s-th p-adic coordinate function* ($\gamma_0(x) = \gamma(x)$).

Note that according to Proposition 2.16 the function $\gamma_0(x) = \gamma(x)$ is the polynomial function: $\gamma_0(x) = x^{p^{n-1}}$. However for $s \geqslant 1$ the function $\gamma_s(x)$ cannot be a polynomial function on R.

The set $\Gamma(R)$ is closed relative to multiplication on R, but not relative to addition. Using the operations of the field $(\Gamma(R), \oplus, \cdot)$ we can prove the following result.

THEOREM 4.4. *(See [239].) For any $\alpha, \beta \in \Gamma(R)$ there holds the equality*

$$\gamma_1(\alpha + \beta) = \bigoplus_{i \in \overline{1, p-1}} \bigl((-1)^i/i\bigr)\alpha^{p^{r-1}i}\beta^{p^{r-1}(p-i)}.$$

If $R = \mathbb{Z}_{p^n}$ then $\Gamma(R)$ is not equal to the usual *p-ary coordinate set* $\{0, \ldots, p-1\}$. We have instead

$$\Gamma(R) = \bigl\{0, 1^{p^{n-1}} = 1, 2^{p^{n-1}}, \ldots, (p-1)^{p^{n-1}}\bigr\},$$
$$\Gamma(R) = \{0, 1, \ldots, p-1\} \iff (p = 2 \text{ or } n = 1).$$

Formulae for the first p-ary coordinate of the sum $a + b$, where $a, b \in \mathbb{Z}_{p^n}$, are given in [226,228].

4.3. *Group of automorphisms of a GR and the Galois theorem [284,320]*

THEOREM 4.5. *The group* $\mathrm{Aut}(R)$ *of automorphisms of the ring* $R = GR(q^n, p^n)$, $q = p^r$, *is a cyclic group of order* r *generated by the* Frobenius automorphism σ, *acting on an element* $a = \sum_{i=0}^{n-1} a_i p^i$, $a_i = \gamma_i(a)$, *by the rule* $\sigma(a) = \sum_{i=0}^{n-1} a_i^p p^i$. *There is a canonical isomorphism* $\varphi: \mathrm{Aut}(R) \to \mathrm{Aut}(\overline{R})$ *such that*

$$\forall \tau \in \mathrm{Aut}(R)\colon\ \varphi(\tau) = \overline{\tau}, \quad \text{where } \overline{\tau}(\overline{\alpha}) = \overline{\tau(\alpha)} \text{ for all } \overline{\alpha} \in \overline{R}.$$

The correspondence $R \to \overline{R}$ *gives an equivalence between the category of Galois rings of characteristic* p^n *and the category of Galois fields of characteristic* p.

THEOREM 4.6. *A subring* $K < R$ *is a GR iff* $K = R_\tau = \{a \in R\colon \tau(a) = a\}$ *for some* $\tau \in \mathrm{Aut}(R)$. *In such a case* $K = GR(p^{tn}, p^n)$, *where* $t = m/\mathrm{ord}\,\tau$. *So if* $K = GR(p^{tn}, p^n) \subseteq R$ *then* $t \mid r$ *and if* $t \mid r$ *then* R *contains a unique subring* $K = GR(p^{tn}, p^n) \leqslant R$.

In this connection it is important to note that the result of [262,320], stating that any subring with the identity of the ring R is a GR, is not true for the case $r > 1, n > 1$. For example: under the last conditions the subring $K = e\mathbb{Z}_{p^n} + \mathfrak{N}$ of the ring R is not a GR.

THEOREM 4.7. *Let* $R \subset S = GR(q^{mn}, p^n)$ *be a Galois extension of the ring* R *of degree* m. *Then the group* $\mathrm{Aut}(S/R)$ *of automorphisms of* S *over* R *is a cyclic group* $\mathrm{Aut}(S/R) = \langle\sigma\rangle$ *of order* m *generated by an automorphism* σ, *acting on an element* $\alpha \in S$ *with* p*-adic decomposition* $\alpha = \sum_{i\in\{0,\dots,n-1\}} \alpha_s p^s$ *as* $\sigma(\alpha) = \sum_{i\in\{0,\dots,n-1\}} \alpha_s^q p^s$.

There exists a one to one correspondence between subgroups $H = \langle\tau\rangle \subseteq \mathrm{Aut}(S/R)$ *and Galois extensions* K *of the subring* R *in the ring* S, *given by the map*

$$H \mapsto S_H = \bigl\{\alpha \in S\colon \forall h \in H\ h(\alpha) = \alpha\bigr\} = S_\tau.$$

Related sources: [8,279].

4.4. *Trace function*

Let $S = GR(q^{mn}, p^n)$ be a Galois extension of the ring $R = GR(q^n, p^n)$. The trace from S onto R is defined as the function $\mathrm{Tr}_R^S(x) = \sum_{\tau\in\mathrm{Aut}(S/R)} \tau(x)$.

PROPOSITION 4.8. *(See [284].) The function* Tr_R^S *is an epimorphism of modules* $\mathrm{Tr}_R^S : {}_RS \to {}_RR$.

4.4.1. *Coordinate functions of trace* Using the p-adic decomposition (Section 4.2) we have

$$\mathrm{Tr}_R^S(x) = \gamma_0\bigl(\mathrm{Tr}_R^S(x)\bigr) + p\gamma_1\bigl(\mathrm{Tr}_R^S(x)\bigr) + \cdots + p^{n-1}\gamma_{n-1}\bigl(\mathrm{Tr}_R^S(x)\bigr).$$

Different applications of this function in coding theory and cryptography are connected with properties of the *coordinate functions* $\gamma_s(\mathrm{Tr}_R^S(x))$. Some expressions for these functions are known. Let us denote by $\mathrm{tr}_{\Gamma(R)}^{\Gamma(S)}(z)$ the trace from the coordinate field $(\Gamma(S), \oplus, \cdot)$

onto the subfield $\Gamma(R)$:

$$\operatorname{tr}^{\Gamma(S)}_{\Gamma(R)}(z) = z \oplus z^q \oplus \cdots \oplus z^{q^{m-1}}.$$

THEOREM 4.9. *(See [226,239].) The zero and first coordinate functions of the trace function have the form*

$$\gamma_0\bigl(\operatorname{Tr}^S_R(x)\bigr) = \operatorname{tr}^{\Gamma(S)}_{\Gamma(R)}\bigl(\gamma_0(x)\bigr) = \gamma_0(x) \oplus \gamma_0(x)^q \oplus \cdots \oplus \gamma_0(x)^{q^{n-1}};$$

$$\gamma_1\bigl(\operatorname{Tr}^S_R(x)\bigr) = \Psi\bigl(\gamma_0(x)\bigr) \oplus \operatorname{tr}^{\Gamma(S)}_{\Gamma(R)}\bigl(\gamma_1(x)\bigr), \quad \textit{where}$$

$$\Psi(x) = \bigoplus_{k_0+\cdots+k_{m-1}=p,\, k_i \in \{0,\ldots,p-1\}} \frac{1}{k_0! \cdots k_{m-1}!} x^{p^{r-1}(k_0+qk_1+\cdots+q^{m-1}k_{m-1})}.$$

If $p = 2$ then

$$\Psi(x) = \varkappa(x)^{2^{r-1}}, \quad \textit{where } \varkappa(x) = \bigoplus_{0 \leqslant i < j \leqslant m-1} x^{q^i+q^j}$$

is a quadratic function on the field $\Gamma(S) = GF(q^m)$ over $\Gamma(R) = GF(q)$.

Indeed this fact allows one to find a linear presentation of the binary Kerdock code and to describe a generalization of this code over any finite field of characteristic 2 [240] (see Section 14.3.3). Formulas for $\gamma_i(\operatorname{Tr}(x))$, $i \geqslant 2$, are given in [226,239].

4.4.2. *Equations with trace* In the study of linear recurrences over Galois rings in connection with their applications to algebraic coding theory and cryptography, a special role is played by equations of type

$$\operatorname{Tr}^S_R(ax) = c, \tag{4.2}$$

where $c \in R, a \in S^*$. Here the number $M_a(c)$ of solutions of this equation in the coordinate field $\Gamma(S)$ is of special interest.

Exact formulas for these numbers are known only for Galois rings of characteristic 4, i.e. for the case $p = 2$, $q = 2^r$, $n = 2$. In this case the last statement of the Theorem 4.9, allows to use the theory of quadrics over a field of characteristic 2.

THEOREM 4.10. *(See [240].) Let $a \in S^*$, $a_0 = \gamma_0(a)$, $a_1 = \gamma_1(a)$, $\alpha = a_0^{-1}a_1$ and $c(\alpha) = \operatorname{tr}^{\Gamma(S)}_{\mathbb{Z}_2}(\alpha) \oplus \lambda e$. Then the number $M_a(c)$ satisfies the following conditions.*

(a) *If $m = 2\lambda + 1 \geqslant 3$, then*

$$M_a(c) = q^{m-2} + v\bigl(c_1 \oplus c_0 c(\alpha)\bigr)\delta_\alpha q^{\lambda-1}, \tag{4.3}$$

where $v(0) = q - 1$, $v(b) = -1$ for $b \neq 0$;

$$\delta_\alpha = (-1)^{\varepsilon(\alpha)}, \quad \varepsilon(\alpha) = \operatorname{tr}^{\Gamma(R)}_{\mathbb{Z}_2}\bigl(\varkappa(\alpha)\bigr) \oplus \lambda \operatorname{tr}^{\Gamma(S)}_{\mathbb{Z}_2}(\alpha) \oplus \binom{\lambda+1}{2} le.$$

(b) *If $m = 2\lambda > 2$, then the values of $M_a(c)$ are given by the following table*:

No.	Conditions on α and c	Values of $M_a(c)$
1	$(c_0, c(\alpha)) = (0, 0)$	$q^{m-2} + v(c_1)\delta_\alpha q^{\lambda-1}$
2	$(c_0, c(\alpha)) \neq (0, 0)$, $c_0 c(\alpha) = 0$	q^{m-2}
3	$c_0 c(\alpha) \neq 0$, $\mathrm{tr}^{\Gamma(R)}_{\mathbb{Z}_2}(c_0^{-1} c_1 c(\alpha)) = \sigma$	$q^{m-2} + (-1)^{\sigma}\delta_\alpha q^{\lambda-1}$

where $\delta_\alpha = (-1)^{\varepsilon(\alpha)}$ *and*

$$\varepsilon(\alpha) = \mathrm{tr}^{\Gamma(R)}_{\mathbb{Z}_2}\big(\varkappa(\alpha)\big) \oplus (\lambda+1)\mathrm{tr}^{\Gamma(S)}_{\mathbb{Z}_2}(\alpha) \oplus \left(\binom{\lambda}{2} l + 1\right)e.$$

(c) *For a given $\delta \in \{-1, 1\}$ the number $D_q(m, \delta)$ of elements $a \in S^*$ such that in* (a), (b) *the equality $\delta(\alpha) = \delta$ holds, has the value*

$$D_q(m, \delta) = \frac{q^m - 1}{2}\big(q^m + \delta q^{[\frac{m+1}{2}]}\big). \tag{4.4}$$

4.5. *Multiplicative group of a Galois ring [226,262,320]*

Group R^* of $R = GR(q^n, p^n)$ is a direct product $R^* = \Gamma(R)^* \times (e + pR)$ of the cyclic group $\Gamma(R)^* = \Gamma(R) \setminus 0$ of order $q - 1$ and the p-group $e + pR$ (congruence-subgroup). For the latter a decomposition in a direct product of cyclic subgroups is known.

In many applications the following description of the possible orders of elements in the group $e + pR$ elements is useful.

PROPOSITION 4.11. *Let $u = e + p^k v \in e + pR$, $1 \leqslant k \leqslant n - 1$, $\bar{v} \neq 0$.*

(a) *If $p^k > 2$ or $p^k = 2$ and $\bar{v} + \bar{v}^2 \neq 0$, then* $\operatorname{ord} u = p^{n-k}$;

(b) *If $p^k = 2$, $\bar{v} + \bar{v}^2 = 0$, then $u^2 = e + p^l w_1$, where $\overline{w}_1 \neq 0$, $3 \leqslant l \leqslant n$,*

$$\operatorname{ord} u = 2^{n-l+1} < 2^{n-k} \leqslant 2^{n-1}.$$

In particular

$$\exp(e + pR) = \begin{cases} 2^{n-2}, & \text{if } R = \mathbb{Z}_{2^n},\ n \geqslant 3, \\ p^{n-1} & \text{in all other cases.} \end{cases}$$

In general case canonical decomposition of the group $e + pR$ for $q = p^r$, $r > 1$, $n > 1$ described by the following way. Let us choose elements $\omega_1, \dots, \omega_r \in R$ such that $\overline{\omega}_1, \dots, \overline{\omega}_r$ is a basis of the field $\overline{R}$ over simple subfield $GF(p)$. In the case $p = 2$, $n > 2$ we can also suppose that $\overline{\omega}_r = \bar{e}$, and can choose element $\omega \in R$ such that $\mathrm{tr}^{\overline{R}}_{GF(2)}(\overline{\omega}) = \bar{e}$. Let finally $g_i = e + p\omega_i$ for $i \in \overline{1, r}$ and $g = e + p\omega$.

THEOREM 4.12. *Under the above denotations*

(a) *if $p > 2$ or $p^n = 4$, then group $e + pR$ has canonical decomposition*

$$e + pR = \langle g_1 \rangle \dot{\times} \cdots \dot{\times} \langle g_r \rangle, \qquad \operatorname{typ}(e + pR) = \left(p^{n-1}, \dots, p^{n-1}\right);$$

(b) *if $p = 2$, $n \geqslant 3$, then group $e + pR$ has canonical decomposition*

$$e + pR = \langle g_1 \rangle \dot{\times} \cdots \dot{\times} \langle g_{r-1} \rangle \dot{\times} \langle g \rangle \dot{\times} \langle -e \rangle,$$
$$\operatorname{typ}(e + pR) = \left(2^{n-1}, \dots, 2^{n-1}, 2^{n-2}, 2\right).$$

4.6. *Bimodules over Galois rings*

Any local f.r. S with $\overline{S} = GF(q)$, $\operatorname{char} S = p^d$, contains a subring $R = GR(q^d, p^d)$ such that $\overline{S} = \overline{R}$. It allows the investigation of S as an (R, R)-bimodule. Here is useful theorem.

THEOREM 4.13. *(See [284].) For any finite (R, R)-bimodule ${}_RM_R$ there exist a generator system $\mu_1, \dots, \mu_k \in M$ and a system of automorphisms $\sigma_1, \dots, \sigma_k \in \operatorname{Aut}(R)$, such that*

$$\forall a \in R,\ l \in \{1, \dots, k\}\colon\ \mu_l a = \sigma_l(a)\mu_l, \quad \text{and} \quad M = R\mu_1 \oplus \cdots \oplus R\mu_k$$

is a direct sum of cyclic (R, R)-bimodules.

A first variant of this theorem for the case $R = \overline{R} = GF(q)$ was proved in [320].

This *theorem about distinguished bases* allows one to prove structure theorems for various different classes of finite rings considered as algebras over a GR. In particular it makes possible a description of finite chain rings [284] (see below) and to present any finite indecomposable ring as a ring of matrices of special form over a Galois ring [71,121,370, 393–395].

4.7. *Generalized (nonassociative) Galois ring*

4.7.1. *Semifields (division algebras) [2–5,85,95,96,171,261,332,388,389]* A ring $(\mathcal{C}, +, *)$ (not necessary associative) is called a *semifield* or *division algebra* if $\mathcal{C}\backslash 0$ is a subloop of the groupoid $(\mathcal{C}, *)$. Such a ring has prime characteristic p, and its (*associative-commutative*) *center*

$$Z(\mathcal{C}) = \left\{a \in \mathcal{C}\colon\ \forall x, y \in \mathcal{C}\colon\ ax = xa,\ (xy)a = x(ya),\ a(xy) = (ax)y\right\} \tag{4.5}$$

is a field: $Z(\mathcal{C}) = GF(p^r)$ and $\mathcal{C}$ is $Z(\mathcal{C})$-space of some dimension m, so $|\mathcal{C}| = p^s$, $s = rm$. For a given $g \in \mathcal{C}$, $k \in \mathbb{N}$ let us define the *right principal k-power* of g by

$$g^{[k]} = \left(\dots\left((g * g) * g\right)^{\dots k \text{ times} \dots}\right).$$

A semifield $\mathcal{C}$ is called *right cyclic* or *primitive* if there exists $g \in \mathcal{C}$ such that any $a \in \mathcal{C}$ has the form $a = g^{[k]}$ for some $k \in \mathbb{N}$.

The conjecture that any finite division ring is primitive [388,389] is not true: there exist at least two non-primitive division rings of orders 32 (see [332]) and 64 (see [171]).

4.7.2. *Main properties of generalized Galois rings [153–156]* For any (possibly nonassociative) ring S let $\mathrm{D}(S)$ be the set of all one-sided zero divisors together with 0. A finite ring S with the identity e is called a *generalized Galois ring* (*GGR*) if $\mathrm{D}(S) = \lambda S$ for some $\lambda \in \mathbb{N}$ (see Section 4.1). The standard Galois ring $S = GR(q^n, p^n)$ is a GGR with $\mathrm{D}(S) = pS$.

THEOREM 4.14. *Let $(S, +, *)$ be a finite GGR. Then*

(a) *the minimal $\lambda \in \mathbb{N}$ with $\mathrm{D}(S) = \lambda S$ is a prime number p, $\mathrm{D} = \mathrm{D}(S) = pS$ is the unique maximal two-sided ideal of S and $\overline{S} = S/\mathrm{D}$ is a semifield of $q = p^s$ elements for some $s \in \mathbb{N}$*;

(b) *the set $S^* = S \setminus \mathrm{D}$ is closed relative to the operation $*$ and $(S^*, *)$ is a loop*;

(c) *the lattice of one-sided ideals of S is a strong chain of two-sided ideals*:

$$S = \mathrm{D}^0 \rhd \mathrm{D}^1 = pS \rhd \cdots \rhd \mathrm{D}^{n-1} = p^{n-1}S \rhd \mathrm{D}^n = 0, \quad \textit{for some } n \in \mathbb{N};$$

char$S = p^n$; *and $S\alpha = \alpha S = \mathrm{D}^t$ for any $\alpha \in \mathrm{D}^t \setminus \mathrm{D}^{t+1}$, $t \in \{0, 1, \dots, n-1\}$.*

(d) *the equalities $|S^*| = q^n - q^{n-1}$, $|\mathrm{D}^t| = q^{n-t}$, for $t \in \{0, 1, \dots, n\}$ take place.*

For a finite semifield $\mathcal{C}$ we shall call any GGR S with *top-factor* $\overline{S} \cong \mathcal{C}$ — a *lift of the semifield $\mathcal{C}$ to GGR S*.

PROPOSITION 4.15. *For any division algebra $(\mathcal{C}, +, *)$ of dimension m over the center $Z(\mathcal{C}) = P = GF(p^r)$ and for any $n \in \mathbb{N}$ there exists a lift to a GGR $(S, +, *)$ of the characteristic p^n with center $Z(S) = R = GR(p^{rn}, p^n)$.*

This proposition is based on the following general construction of a GGR S with a fixed Galois subring $R = GR(p^{rn}, p^n)$ in the center $Z(S)$. Let $S = GR(q^{mn}, p^n)$ be a Galois extension of the degree m of the ring R where $q = p^r$. Let us fix some basis $\mathbf{e} = (e_0 = e, \dots, e_{m-1})$ of the module ${}_RS$ and denote by $\alpha_{\mathbf{e}}$ the column of coordinates of the element $\alpha \in S$ in the basis $\mathbf{e}$. So $\alpha = \mathbf{e}\alpha_{\mathbf{e}}$. Now let us fix some $m \times m$-matrix $G = (g_{ij})_{i,j \in \{0,\dots,m-1\}}$ over the module S and define an operation $*$ on S by the condition

$$\forall \alpha, \beta \in S\colon \quad \alpha * \beta = \alpha_{\mathbf{e}}^T G \beta_{\mathbf{e}}. \tag{4.6}$$

We shall call $G = G(\mathbf{e})$ *the Gramm matrix of the operation $*$ with respect to basis* $\mathbf{e}$; G is said to be *unitary* if all elements of its first row and first column are equal to e; and it is called a *linear Latin square over R* if any linear combinations of the rows (of the columns) of the matrix G with coefficients in R, not all of them belonging to pR, is a basis of ${}_RS$.

PROPOSITION 4.16. *The algebra $(S, +, *)$ is a* (*in general nonassociative*) *ring with $R \subseteq Z(S)$*; *an element e is the identity of this ring if and only if G is an unitary matrix*; $(S, +, *)$ *is a GGR if and only if G is an unitary linear Latin square over R.*

Now it is not difficult to see that in using the notation above we can suppose that the top-factor $\overline{S} = S/pS$ is the set $\mathcal{C}$ in Proposition 4.15 and $\overline{R} = P$. Then $\bar{\mathbf{e}} = (\bar{e}_0, \dots, \bar{e}_{m-1})$ is a basis of ${}_P\mathcal{C}$ and we can choose the matrix G in Proposition 4.16 such that $\overline{G} = G(\bar{\mathbf{e}})$ is a Gramm matrix of the operation $*$ on $\mathcal{C}$ in the basis $\bar{\mathbf{e}}$. It imply that GGR $(S, +, *)$ in Proposition 4.16 is a lift of $\mathcal{C}$ with center R.

5. Structure and properties of chain rings [22,23,72,111,120,184,212–214,262,264,280,283–285,357,403]

As was noted above, a finite chain ring S is a local f.r. with $\mathfrak{N}(S) = S\pi$ for some $\pi \in \mathfrak{N}(S)$. The lattice of all left (right) ideals of S is a chain of two-sided ideals $\mathfrak{N}(S)^i = S\pi^I = \pi^i S$:

$$S \rhd \mathfrak{N}(S) \rhd \cdots \rhd \mathfrak{N}(S)^{n-1} \rhd \mathfrak{N}(S)^n = 0, \quad n = \operatorname{ind}\mathfrak{N}(S) \in \mathbb{N}. \tag{5.1}$$

If $\bar{S} = S/\mathfrak{N}(S) = GF(q)$, $q = p^r$, then $|S| = q^n$, $\operatorname{char} S = p^d$, $1 \leqslant d \leqslant n$; and the ring S contains a Galois subring $K = GR(q^d, p^d)$ called the *coefficient ring* of S. Any such subring satisfies the relations:

$$\overline{K} = \bar{S}, \qquad S = K[\pi], \quad \text{for any } \pi \in \mathfrak{N}(S) \setminus \mathfrak{N}(S)^2. \tag{5.2}$$

A coefficient ring of the ring S is uniquely defined if and only if S is a commutative ring. In this case it has the form

$$K = \varGamma(S) + \varGamma(S)p + \cdots + \varGamma(S)p^{d-1}, \tag{5.3}$$

where $\varGamma(S)$ is the coordinate field of the ring S (see Section 2.4)

Since every ideal of the ring S is a power of $\mathfrak{N}(S)$ there exists a parameter $\varepsilon \in \mathbb{N}$, called the *ramification index* of S, such that $pS = \mathfrak{N}(S)^\varepsilon$.

5.1. *Commutative chain rings (GE-rings)*

In the following we fix parameters

$$p, r, q = p^r, d, n, \varepsilon \tag{5.4}$$

of the ring S and consider only the nontrivial case when $n > 1$ (i.e. S is not a field) and $\varepsilon > 1$ (i.e. S is not a Galois ring).

5.1.1. *Description of commutative chain rings [16,73,283,357,381]*

THEOREM 5.1. *(See* [357].*) Let S be a commutative chain ring with coefficient ring $K = GR(q^d, p^d)$. Then π is a root of an* Eisenstein polynomial $E(x) = x^\varepsilon - c_{\varepsilon-1}x^{\varepsilon-1} - \cdots - c_0 \in K[x]$, *where $c_0, \dots, c_{\varepsilon-1} \in pK$, $c_0 \notin p^2K$, if $d > 1$, and*

$$S \cong K[x]/\big(E(x), p^{d-1}x^\rho\big), \qquad n = (d-1)\varepsilon + \rho, \quad 1 \leqslant \rho \leqslant \varepsilon. \tag{5.5}$$

In view of (5.5) a commutative chain f.r. is also called a *Galois–Eisenstein ring* or *GE-ring*.

Any GE-ring is a quotient ring of some commutative local principal ideal domain. Let $\mathbb{Q}_p$ be the field of p-adic numbers and $\mathcal{E}$ let be some finite extension of $\mathbb{Q}_p$. Then the ring $\mathbb{Z}_\mathcal{E}$ of integer numbers of the field $\mathcal{E}$ [245] is a local principal ideal domain and if $J(\mathbb{Z}_\mathcal{E})$ is its Jacobson radical, then $\mathbb{Z}_\mathcal{E}/J(\mathbb{Z}_\mathcal{E}) = GF(p^r)$, $p\mathbb{Z}_\mathcal{E} = J(\mathbb{Z}_\mathcal{E})^\varepsilon$ for some $r, \varepsilon \in \mathbb{N}$. In this situation any quotient ring $S = \mathbb{Z}_\mathcal{E}/J(\mathbb{Z}_\mathcal{E})^n, n \geqslant \varepsilon$, is a GE-ring with ramification index ε and top-factor $\bar{S} = GF(p^r)$ [245].

PROPOSITION 5.2. *(See [283].) Any GE-ring S is of the form pointed out above*: $S \cong \mathbb{Z}_{\mathcal{E}}/J(\mathbb{Z}_{\mathcal{E}})^n$ *for a suitable finite extension* $\mathcal{E}$ *of the field* $\mathbb{Q}_p$.

A GE-ring is a simple extension of any chain subring $A < S$. Let $\mathfrak{N}(A) = \pi_A A$, and let n_A be the nilpotency index of the ring A; let $\delta = \varepsilon(S/A)$ be the ramification index of S over A, defined by $\mathfrak{N}(A)S = \mathfrak{N}(S)^\delta$; finally let $m = [\bar{S} : \bar{A}]$ be the degree of the extension $\bar{S}$ of the field $\bar{A}$. Then $n = (n_A - 1)\delta + \nu$ for some $\nu \in \{1, \dots, \delta\}$.

PROPOSITION 5.3. *(See [283].) For any Galois polynomial* $G(x) \in A[x]$ *of degree* m *there exists a polynomial* $F(x) = G(x)^\delta + F_{\delta-1}(x)G(x)^{\delta-1} + \cdots + F_0(x)$ *such that* $\deg F_i(x) < m$, $F_i(x) \in \mathfrak{N}(A)[x]$ *for* $i \in \{0, \dots, \delta - 1\}$, $F_0(x) \notin \mathfrak{N}(A)^2[x]$ *if* $n_A > 1$, *and*

$$S \cong A[x]/\big(F(x), \pi_A^{n_A-1} G(x)^\nu\big).$$

In particular $S = A[\alpha]$ *for some* $\alpha \in S$.

Under the conditions of the last proposition the polynomial $F(x)$ is called a *Galois–Eisenstein* polynomial over the ring A.

COROLLARY 5.1. *For any Galois polynomial* $G(x) \in \mathbb{Z}_{p^d}[x]$ *of degree* r *there exists an isomorphism*

$$S \cong \mathbb{Z}_{p^d}[x]/\big(F(x), p^{d-1} G(x)^\rho\big), \tag{5.6}$$

where $F(x) = G(x)^\varepsilon + F_{\varepsilon-1}(x)G(x)^{\varepsilon-1} + \cdots + F_0(x) \in \mathbb{Z}_{p^d}[x]$ *is such that* $\deg F_i(x) < r$, $F_i(x) \in p\mathbb{Z}_{p^d}[x]$ *for* $i \in \{0, \dots, \varepsilon - 1\}$, $F_0(x) \notin p^2\mathbb{Z}_{p^d}[x]$ *if* $d > 1$. *Any ring of the form* (5.6) *is a GE-ring.*

A GE-ring S is called a *weakly ramified* if $(\varepsilon, p) = 1$. In this case there exists an element $\pi \in \mathfrak{N} \setminus \mathfrak{N}^2$ such that in (5.5) $E(x) = x^\varepsilon - \pi$.

THEOREM 5.4. *(See [16].) A GE-ring S is uniquely defined by the parameters* q, n, ε *up to isomorphism in exactly the following cases*:

(a) $n = \varepsilon$, *i.e.* $d = 1$ (*then* $S \cong GF(q)[x]/(x^n)$);

(b) $n = \varepsilon + 1$, $(\varepsilon, q - 1) = 1$ (*then* $S \cong GR(q^2, p^2)[x]/(x^2 - pe, px)$);

(c) $(\varepsilon, q - 1) = 1$, $(\varepsilon, p) = 1$ (*then* $S \cong GR(q^d, p^d)[x]/(x^\varepsilon - pe, p^{d-1}x^\rho)$, $\rho = n - (d - 1)\varepsilon$).

Some estimates of the number of classes of isomorphic GE-rings for a fixed parameters q, n, ε are given in [73]. In general the *isomorphism problem* for GE-rings with given parameters q, n, ε is still open.

The pair of GE-rings $\mathbb{Z}_9[x]/(x^2 - 3, 3x)$ and $\mathbb{Z}_9[x]/(x^2 - 6, 3x)$ gives the simplest example of finite non-isomorphic rings with isomorphic additive and multiplicative groups [141]. Another example: $\mathbb{Z}_4[x]/(x^2 - 2)$ and $\mathbb{Z}_4[x]/(x^2 - 2x - 2)$.

5.1.2. *Systems of linear equations over a GE-ring* Here S is a GE-ring with radical $\mathfrak{N} = \pi S$, nilpotency index $n \geqslant 1$ and residue field $\overline{S} = GF(q)$ of characteristic p. Let us consider a system

$$A_{k\times m}x^{\downarrow} = b^{\downarrow}_{k\times 1} \tag{5.7}$$

of k linear equations in m variables over the ring S. A solvability criterion of this system, results on the number of solutions and a description of the algebraic structure of the set of solutions are based on the following notions.

We say that a matrix $B_{k\times m}$ over S is (*row*) *equivalent* to A and write $A \sim B$ (resp. $A \overset{\text{rw}}{\sim} B$) if B can be obtained from A by a finite series of elementary transformations (resp. of rows). Condition $A \sim B$ ($A \overset{\text{rw}}{\sim} B$) is equivalent to the equality $B = UAV$ (resp. $B = UA$), where U, V are invertible matrices.

PROPOSITION 5.5. *Any matrix $A_{k\times m}$ over the ring S is equivalent to a unique diagonal matrix of the form*

$$\mathcal{K}(A) = \operatorname{diag}\bigl(\pi^{s_1}, \ldots, \pi^{s_t}\bigr), \quad 0 \leqslant s_1 \leqslant s_2 \leqslant \cdots \leqslant s_t \leqslant n,\ t = \min\{k, m\}. \tag{5.8}$$

The matrix $\mathcal{K}(A)$ is called the *canonical form* of the matrix A. We define the *signature* and (*right*) *defect* of A respectively by

$$\operatorname{sign} A = [s_1, \ldots, s_t], \qquad \operatorname{def} A = s_1 + \cdots + s_t + (m-t)n. \tag{5.9}$$

PROPOSITION 5.6. *Two matrices $A_{k\times m}$, $B_{k\times m}$ over the ring S are equivalent if and only if* $\operatorname{sign} A = \operatorname{sign} B$. *A square matrix A is invertible if and only if* $\operatorname{def} A = 0$.

The following result is a generalization of well-known criterion of Kronecker–Capelli.

THEOREM 5.7. *A system of linear equations* (5.7) *is solvable if and only if*

$$\operatorname{sign}\bigl(A, b^{\downarrow}\bigr) = \operatorname{sign}\bigl(A, 0^{\downarrow}\bigr). \tag{5.10}$$

Under this condition the number of solutions of the system (5.7) *is equal to* $q^{\operatorname{def} A}$.

In order to describe the set of all solutions of the system (5.7) let us consider first the associated system of homogeneous linear equations:

$$A_{k\times m}x^{\downarrow} = 0^{\downarrow}. \tag{5.11}$$

Under the condition (5.8) let

$$\mathcal{K}(A) = UAV, \quad U \in S^*_{k,k},\ V \in S^*_{m,m},\ V = \bigl(V_1^{\downarrow} \cdots V_t^{\downarrow} V_{t+1}^{\downarrow} \cdots V_m^{\downarrow}\bigr). \tag{5.12}$$

THEOREM 5.8. *The set $\mathfrak{R}(A)$ of solutions of the system* (5.11) *is a submodule of ${}_S S^{(m)}$ with direct decomposition*

$$\mathfrak{R}(A) = V\mathfrak{R}\big(K(A)\big) = S\cdot\big(\pi^{n-s_1}V_1^{\downarrow}\big) \dotplus \cdots \dotplus S\cdot\big(\pi^{n-s_t}V_t^{\downarrow}\big) \dotplus S\cdot V_{t+1}^{\downarrow} \dotplus \cdots \dotplus S\cdot V_m^{\downarrow}. \tag{5.13}$$

The system (5.7) *is solvable exactly if the column $Ub^{\downarrow}$ of length k has the form*

$$Ub^{\downarrow} = \big(\pi^{s_1}d_1, \ldots, \pi^{s_t}d_t,\ 0, \ldots, 0\big)^T, \quad \textit{for some } d_1, \ldots, d_t \in R, \tag{5.14}$$

and then the set of all solutions of this system is

$$c^{\downarrow} + \mathfrak{R}(A), \quad \textit{where } c^{\downarrow} = Ud^{\downarrow},\ d^{\downarrow} = (d_1, \ldots, d_t,\ 0, \ldots, 0)^T. \tag{5.15}$$

Some others approaches to the solution systems of linear equation over GE-rings and other finite rings are in [105–110,112].

5.1.3. *Exponent of congruence-subgroup of a GE-ring* For a GE-ring S the parameter $\exp(e+\mathfrak{N})$ can be calculated exactly. In this case parameter $\mu(S)$ in (2.8) equals ω in (2.10) for "almost" all values of parameters q, n, ε corresponding to S.

Under the conditions of Theorem 5.1 we can state that

$$\pi^{\varepsilon} = c_0 + c_1\pi + \cdots + c_{\varepsilon-1}\pi^{\varepsilon-1} = pv, \quad \text{where } v \in S^*. \tag{5.16}$$

A GE-ring (5.5) is called *non-stable* if

$$\begin{aligned} &q = p; \qquad \varepsilon = p^b(p-1), \quad \text{for some } b \in \mathbb{N}_0;\\ &n > p^{b+1}; \qquad c_0 = pv, \quad \bar{v} = \bar{e}. \end{aligned} \tag{5.17}$$

Otherwise the GE-ring S is said to be *stable*.

The simplest example of a non-stable ring is $\mathbb{Z}_{2^d}$ for $d \geqslant 3$. Some series of other examples are given in [224].

THEOREM 5.9. *(See* [296].*) If S is a stable GE-ring, then* $\exp(e+\mathfrak{N}) = p^{\omega}$.

For a non-stable GE-ring (5.5) with a root π of the polynomial $E(x)$ let us define a parameter δ by the conditions:

$$\delta = \begin{cases} \min\{\|v-e\|,\ p^b\}, & \text{if } p \geqslant 3;\\ 1, & \text{if } p = 2,\ b = 0;\\ \min\{\|v-e-\pi+\pi^2\|,\ 3\}, & \text{if } p = 2,\ b = 1;\\ \min\{\|v-e-\pi^{2^{b-1}}+\pi^{2^b}+\pi^{2^b+2^{b-2}}\|,\ 2^b+2^{b-1}\}, & \text{if } p = 2,\ b \geqslant 2. \end{cases} \tag{5.18}$$

THEOREM 5.10. *(See* [224].*) Let S be a non-stable GE-ring. Then*

$$\exp(e+\mathfrak{N}) = p^{\varkappa}, \quad \textit{where } \varkappa = \max\left\{\left]\frac{n-p^b-\delta}{\varepsilon}\right[+ b,\ b+1\right\}. \tag{5.19}$$

In some cases it is possible to present $\exp(e+\mathfrak{N})$ in a more explicit form.

THEOREM 5.11. *(See [224].) Let S be a nonstable ring. If $p \geqslant 3$, then*

$$\exp(e + \mathfrak{N}) = \begin{cases} p^{\omega-1}, & \text{if } p^b < \rho \leqslant p^b + \delta; \\ p^{\omega}, & \text{otherwise}; \end{cases} \tag{5.20}$$

if $p = 2$, then

$$\exp(e + \mathfrak{N}) = \begin{cases} 2^{\omega-2}, & \text{if } \rho + \varepsilon \leqslant \delta,\ d \geqslant 4; \\ 2^{\omega-1}, & \text{if } \rho \leqslant \delta < \rho + \varepsilon,\ \text{or } \rho + \varepsilon \leqslant \delta,\ d = 3; \\ 2^{\omega}, & \text{if } \delta < \rho. \end{cases} \tag{5.21}$$

5.1.4. *A ring of polynomial transformations of a GE-ring [287]* A function $f : S \to S$ is called a *polynomial transformation* of the ring S if there exists a polynomial $F(x) \in S[x]$ such that $f(a) = F(a)$ for all $a \in S$. We shall write that $f = F_S$ and say also that $F(x)$ is a polynomial presentation of the transformation f.

The set $\mathcal{P}(S)$ of all polynomial transformations of the ring S is a ring relative to addition and (pointwise) multiplication of functions. Consider the epimorphism $\psi : S[x] \to \mathcal{P}(S)$ defined by the rule $\psi(F(x)) = F_S$. Then

$$\mathcal{P}(S) \cong S[x]/\operatorname{Ker}\psi. \tag{5.22}$$

In order to describe $\mathcal{P}(S)$ we need to describe $\operatorname{Ker}\psi$.

Note that polynomial presentations of transformations of the finite field $\bar{S}$ are connected with its characteristic polynomial $x^q - x \in \bar{S}[x]$. Properties of the polynomial $\Phi(x) = x^q - x \in \mathcal{R}$ are important also in our investigations. It is evident that $\Phi(a) \in \pi S$ for every $a \in S$.

PROPOSITION 5.12. *The map Φ_S: $S \to \pi S$ is surjective. For every element $b \in \pi S$ there exist exactly q different elements $a \in S$ with the property $\Phi(a) = b$ and any two of them are different modulo πS.*

Let us consider the system of polynomials

$$\Phi_0(x) = \Phi(x) = x^q - x, \qquad \Phi_t(x) = \Phi_{t-1}(x)^q - \pi^{q^t-1}\Phi_{t-1}(x), \quad t \in \mathbb{N}; \tag{5.23}$$

and put

$$\delta_t = \frac{q^{t+1} - 1}{q - 1}. \tag{5.24}$$

PROPOSITION 5.13.

(a) *For $t \geqslant 0$ there are equalities*

$$\deg \Phi_t(x) = q^{t+1}, \qquad \Phi_t(S) = \pi^{\delta_t} S. \tag{5.25}$$

(b) *For every $t \geqslant 0$ and any elements $\rho, \rho_0, \ldots, \rho_{t-1} \in \bar{S}, b \in S$ there exists an element $a \in S$ satisfying the conditions*:

$$\bar{a} = \rho, \quad \Phi_0(a) = \pi^{\delta_0} r_0, \quad \bar{r}_0 = \rho_0; \qquad \ldots \quad ;$$
$$\Phi_{t-1}(a) = \pi^{\delta_{t-1}} r_{t-1}, \quad \bar{r}_{t-1} = \rho_{t-1}; \qquad \Phi_t(a) = \pi^{\delta_t} b.$$

For any $i \in \mathbb{N}_0$ let us consider its p-adic decomposition $i = i_0 + qi_1 + \cdots + q^h i_h$, $0 \leqslant i_t \leqslant q-1$, $t \in \{0, \ldots, h\}$, and define a polynomial $F_i(x) \in S[x]$ by the rule

$$F_i(x) = \Phi_0(x)^{i_0}\Phi_1(x)^{i_1} \cdots \Phi_h(x)^{i_h}.$$

Then $F_0 = e$ and $\deg F_i(x) = qi$ for $i \in \mathbb{N}$. Introduce also the following functions on $\mathbb{N}_0$:

$$\alpha_q(t) = \left[\frac{t}{q}\right] + \left[\frac{t}{q^2}\right] + \cdots, \qquad \beta_q(k) = \min\bigl\{t\colon \alpha_q(t) \geqslant k\bigr\},$$

and note that $q|\beta_q(k)$ for every $k \in \mathbb{N}_0$. Below $m = (1/q)\beta_q(n)$.

THEOREM 5.14.

(a) *The ideal* Ker ψ *in* (5.22) *is generated by the system of polynomials*

$$F_m(x),\ \pi^{\varepsilon_{m-1}} F_{m-1}(x), \ldots, \pi^{\varepsilon_1} F_1(x),$$
$$\text{where } \varepsilon_i = n - \alpha_q(qi),\ i \in \{1, \ldots, m-1\}.$$

(b) *Any polynomial transformation $g \in \mathcal{P}(S)$ has a polynomial presentation of the form*

$$\begin{aligned} &g = G_S, \quad \textit{where} \\ &G(x) = G_0(x) + G_1(x)F_1(x) + \cdots + G_{m-1}(x)F_{m-1}(x), \\ &\deg G_i(x) < q. \end{aligned} \tag{5.26}$$

(c) *A transformation g of the form* (5.26) *is a permutation on S exactly if $\overline{G_0(x)}$ presents a permutation on the field $\overline{S} = GF(q)$ and polynomial $\overline{G_1}(x) - \overline{G'}_0(x)$ has no roots in this field.*

(d)

$$\bigl|\mathcal{P}(S)\bigr| = q^{q(nm - \sum_{i=1}^{m-1} \alpha_q(qi))} = q^{\sum_{i=1}^{n} \beta_q(i)},$$
$$\bigl|\mathcal{P}(S)^*\bigr| = \frac{q!}{q^q}\left(\frac{q-1}{q}\right)^q \bigl|\mathcal{P}(S)\bigr|.$$

In [330] some results about $|\mathcal{P}(S)|$ for any commutative f.r. S with identity are given.

5.1.5. *Polynomial functions over integer residue rings* Let $S = \mathbb{Z}/(N)$ be the ring of residues modulo N. For any function $f : S \to S$ and $t \in \mathbb{N}$ *the t-th difference* $\Delta^t f(x)$ is defined recursively by

$$\Delta f(x) = f(x+1) - f(x), \qquad \Delta^t f(x) = \Delta\Delta^{t-1} f(x).$$

Using the notations of the previous section set $\gamma(t) = \min\{n, \alpha_p(t)\}$.

THEOREM 5.15. *(See* [61].*) If $S = \mathbb{Z}/(N)$, $N = p^n$, p is a prime, then the following conditions are equivalent*:

(a) $f(x) \in \mathcal{P}(S)$;

(b) $\forall t \in \mathbb{N}$; $\varDelta^t f(x) \equiv 0 \pmod{p^{\gamma(t)}}$;

(c) $\forall t \in \{1, \ldots, p^n - 1\}$; $\varDelta^t f(x) \equiv 0 \pmod{p^{\gamma(t)}}$;

(d) $\forall t \in \{1, \ldots, p^n - 1\}$; $\sum_{s\in\{0,\ldots,t\}}(-1)^{t-s}\binom{t}{s} f(x+ps) \equiv 0 \pmod{p^{\gamma(t)}}$;
(e) $\exists f_0(x), f_1(x), \ldots, f_{n-1}(x) \in \mathcal{P}(S)$: $f(x+py) = f_0(x) + pyf_1(x) + \cdots + (py)^{n-1} f_{n-1}(x)$.

Any function $f \in \mathcal{P}(S)$ can be presented in the form

$$f(x) = \sum_{t\in\{0,\ldots,\beta_p(n-1)\}} c_t x(x-1)\cdots(x-t+1),$$

$$\text{where } t!c_t \equiv \Delta^t f(0) \pmod{p^n}.$$

For any $N \in \mathbb{N}$ let $\beta(N) = \min\{t \in \mathbb{N}\colon N \mid t!\}$ (then $\beta(p^n) = \beta_p(n)$).

THEOREM 5.16. *(See [195,352].) If $S = \mathbb{Z}/(N)$, $N \in \mathbb{N}$, then any function $f \in \mathcal{P}(S)$ has unique representation in a form*

$$f(x) = \sum_{s\in\{0,\ldots,\beta(N)-1\}} b_s x^s, \quad \text{where } 0 \leqslant b_s < \frac{N}{(s!, N)}.$$

See also [44].

5.2. *Noncommutative chain rings (GEO-rings) [284,285]*

Let now S be a noncommutative finite chain ring with its parameters given as in the beginning of the Section 5.

Let $R = GR(q^d, p^d)$, $(q = p^r)$, be a Galois subring of S with the property $\overline{R} = \overline{S}$ (coefficient subring of S). Theorem 4.13 (about distinguished bases of an (R, R)-bimodule) gives

PROPOSITION 5.17. *There exists a generator π of $\mathfrak{N}(S)$ and an automorphism $\tau \in \operatorname{Aut}(R)$ such that*

$$\forall a \in R\colon \quad \pi a = \tau(a)\pi.$$

The automorphism τ is completely defined by a parameter $\lambda \in \{0, \ldots, r-1\}$ satisfying the condition

$$\forall a \in R\colon \quad \overline{\tau(a)} = \bar{a}^{p^\lambda}.$$

Under the last condition we have $\operatorname{ord}\tau = t = \frac{r}{(r,\lambda)}$ *and $t|(\varepsilon, r)$. If $t = 1$, in particular, if $(r, \varepsilon) = 1$, then S is commutative, i.e. a GE-ring.*

Under the condition of Proposition 5.17 we call π a *distinguished generator* of the radical and τ an *associated automorphism* of the ring S.

Let $R[x, \tau]$ be an *Ore polynomial ring*, i.e. a polynomial ring with the usual addition but multiplication (on the right) defined by $xa = \tau(a)x$, $a \in R$.

THEOREM 5.18.

(a) *A distinguishing generator π of $\mathfrak{N}(S)$ is a root of a* special Eisenstein polynomial:

$$E(x) = x^{tm} - c_{m-1}x^{t(m-1)} - \cdots - c_1x^t - c_0 \in R[x;\tau], \quad \textit{where}$$
$$c_i \in pR,\ i \in \{0,\dots,m-1\}; \quad c_0 \notin p^2R,\ \textit{if } d > 1; \tag{5.27}$$

and either

(a1) $c_i \in R_\tau = \{a \in R\colon\ \tau(a) = a\}$, $i \in \{0,\dots,m-1\}$; *or*

(a2) $n - (d-1)\varepsilon = tk + 1$, *for some* $k \in \{1,\dots,m-1\}$, $\tau(c_k) - c_k \in p^{d-1}R \setminus 0$, *and* $c_i \in R_\tau$, *for* $i \in \{0,\dots,m-1\}$, $i \neq k$.

(b) *Let* $\rho = n - (d-1)\varepsilon$. *Then* $\rho \in \{1,\dots,\varepsilon\}$ *and there exists an isomorphism*

$$S \cong R[x;\tau]/I,$$
$$\textit{where } I = E(x)R[x;\tau] + p^{d-1}x^\rho R[x;\tau] = \bigl(E(x), p^{d-1}x^\rho\bigr). \tag{5.28}$$

(c) *For any special Eisenstein polynomial* (5.27) *the ideal* I *of the form* (5.28) *is a two-sided ideal and the ring* (5.28) *is a chain ring with its parameters as described.*

(d) *The center* Z *of the ring* S *is* $Z = C = R_\tau + R_\tau\pi^t + \cdots + R_\tau\pi^{tk}$ *in case* (a1), *and* $Z = C + p^{d-1}R\pi^{tk}$ *in case* (a2).

A weaker version of this theorem was obtained in [72].

In view of this result we call a noncommutative chain f.r. also a *Galois–Eisenstein–Ore-* or *GEO-ring*. In case (a1) the GEO-ring S can be represented as a quotient ring of some prime principal ideal ring (a generalization of Proposition 5.2), in case (a2) it cannot be represented in such a form [285].

THEOREM 5.19. *(See* [333].*) A GEO-ring* S *is uniquely defined by its parameters* n, d, q, λ *up to isomorphism exactly in the following cases.*

(a) $n = \varepsilon$ *(then* $S = R[x,\tau]/(x^\varepsilon)$*)*;

(b) $n = \varepsilon + 1$, $(q-1, m(p^{\lambda t}-1)/(p^\lambda - 1)) = 1$ *(then* $S = R[x,\tau]/(x^\varepsilon - p, x^n)$*))*;

(c) $n > \varepsilon + 1$, $(q-1, m(p^{\lambda t}-1)/(p^\lambda - 1)) = (q-1)/(p^{(\lambda,r)} - 1)$ *(then* $S = R[x,\tau]/(x^\varepsilon - p, x^n)$*)*.

Related sources: [9,11,12,97,98,175].

6. Finite (quasi-)Frobenius rings and bimodules

Let A, B be f.r. with identity. Recall that an Abelian group M is called an (A,B)-bimodule $({}_AM_B)$ if it is left A-module, a right B-module and

$$\forall a \in A, b \in B, \alpha \in M\colon \quad a(\alpha b) = (a\alpha)b.$$

For any subsets $I \subseteq A$, $J \subseteq B$, $N \subseteq M$ define the *right annihilator of* I *in* M by $\rho_M(I) = \{\beta \in M\colon\ I\beta = 0\}$, and by analogy define $\lambda_M(J)$, the left *annihilator of* J *in* M, $\lambda_A(N)$, the *left annihilator of* N *in* A, $\rho_B(N)$, the *right annihilator of* N *in* B.

A bimodule ${}_AM_B$ is called *quasi-Frobenius* (QF-bimodule) [27], or a *duality context* [116], if it is faithful from the left ($\lambda_A(M) = 0$) and from the right ($\rho_B(M) = 0$); for every maximal left ideal $I <_{\max} {}_AA$ its (right) annihilator $\rho_M(I)$ is an irreducible B-module, and for every maximal right ideal $J <_{\max} B_B$ its (left) annihilator $\lambda_M(J)$ is an irreducible A-module. A ring R is called *quasi-Frobenius (QF-ring)* if the regular bimodule ${}_RR_R$ is QF.

EXAMPLE 6.1. Any principal ideal f.r. R in particular $R = \mathbb{Z}_m$, is a QF-ring. Indeed, if $R = \mathbb{Z}_m$ and $I <_{\max} {}_RR$ then $I = pR$, where $p \mid m$ and p is a prime. Therefore $\rho_R(I) = dR$, where $d = m/p$, and dR is an irreducible (minimal) ideal of R of cardinality p.

EXAMPLE 6.2. The bimodule ${}_AM_B$, where $A = M_m(P)$, $B = M_n(P)$ are matrix rings over $P = GF(q)$, and $M = M_{m,n}(P)$ is the space of $m \times n$-matrix, is a QF-bimodule.

THEOREM 6.3. *(See* [193,244]*.) For a finite field P and a finite group G the group ring PG is a QF-ring.*

6.1. *Characterizations of QF-bimodules*

Let $\mathrm{End}(M_B)$ and $\mathrm{End}({}_AM)$ be the rings of endomorphisms of M_B and ${}_AM$, where the elements of $\mathrm{End}({}_AM)$ act on elements of M from the right, and the elements of $\mathrm{End}(M_B)$ act on elements of M from the left. Then we can consider the ring B as a subring of the ring $\mathrm{End}({}_AM)$ identifying $b \in B$ with the map $\check{b} : M \to M$, $\check{b}(x) = xb$. Symmetrically we can consider the ring A as a subring of the ring $\mathrm{End}(M_B)$, identifying $a \in A$ and the map $\hat{a} : M \to M$, $\hat{a}(x) = ax$ [26,65,116].

6.1.1. *Characterizations in terms of categories and annihilators* Recall that a module ${}_AM$ is called *injective* (a *cogenerator* in the category $A-\mathrm{Mod}$) if for any module ${}_AN$, any submodule $K \leqslant {}_AN$, and any homomorphism $\varphi : {}_AK \to {}_AM$ there exists an extension of φ to a homomorphism $\psi : {}_AN \to {}_AM$ (and so ${}_AM$ contains isomorphic images of all irreducible A-modules [116,396]).

THEOREM 6.4. *(See* [27,116,163]*.) For a bimodule ${}_AM_B$ the following conditions are equivalent.*

(a) *${}_AM_B$ is a (finite) QF-bimodule.*

(b) *${}_AM$ is a finite injective cogenerator and $B = \mathrm{End}({}_AM)$.*

(c) *M_B is a finite injective cogenerator and $A = \mathrm{End}(M_B)$.*

(d) *$A = \mathrm{End}(M_B)$, $B = \mathrm{End}({}_AM)$ and $\lambda_M(\rho_B(L)) = L$, $\rho_M(\lambda_A(N)) = N$ for all submodules $L \leqslant {}_AM$ and $N \leqslant M_B$.*

If ${}_AM_B$ is a QF-bimodule then also $\lambda_A(\rho_M(I)) = I$, $\rho_B(\lambda_M(J)) = J$, for every left ideal $I \leqslant {}_AA$ and right ideal $J \leqslant B_B$.

6.1.2. *Socle characterization of QF-bimodules* Recall that the nil-radical $\mathfrak{N} = \mathfrak{N}(R)$ of a f.r. R coincides with the Jacobson radical of R and can be defined as the intersection of all right maximal ideals of R.

The *socle* of ${}_AM$ is a notion dual to the notion of Jacobson radical, it is the sum $\mathfrak{S}({}_AM)$ of all left minimal (irreducible) submodules of ${}_AM$. We can represent it also as the right annihilator of $\mathfrak{N} = \mathfrak{N}(A)$ in M: $\mathfrak{S}({}_AM) = \rho_M(\mathfrak{N}) = \{\alpha \in M\colon \mathfrak{N}\alpha = 0\}$. In fact the A-module $\mathfrak{S}({}_AM)$ is a left module over the top-factor $\overline{A} = A/\mathfrak{N}$ where the multiplication of $\alpha \in \mathfrak{S}({}_AM)$ by $\bar{a} = a + \mathfrak{N} \in \overline{A}$ is defined as $\bar{a}\alpha = a\alpha$. There is the following useful addition to Theorem 6.4

THEOREM 6.5. *(See [163,306].) A faithful bimodule ${}_AM_B$ is QF iff*
(e) $\mathfrak{S}({}_AM) = \mathfrak{S}(M_B) = \mathfrak{S}$ *and* ${}_{\overline{A}}\mathfrak{S}_{\overline{B}}$ *is a QF-bimodule.*

EXAMPLE 6.6. The ring R of all upper-triangle 2×2-matrices over a field P is not a QF-ring. In fact, $\mathfrak{N}(R)$ is the subset of all matrices from R with zero diagonal; $\mathfrak{S}({}_RR) = \lambda_R(\mathfrak{N}(R))$ is the subset of all matrices with zero first column; $\mathfrak{S}(R_R) = \rho_R(\mathfrak{N}(R))$ is the subset of all matrices with zero second row. So $\mathfrak{S}({}_RR) \neq \mathfrak{S}(R_R)$.

The equivalence of statements (a, b, c) of Theorem 6.4 permits the introduction of the following definition: a module ${}_AM$ is called a *QF-module* if the natural bimodule ${}_AM_B$, where $B = \mathrm{End}({}_AM)$ is a QF-bimodule.

6.2. *Existence and construction of finite QF-modules [163]*

There is natural question: for a given f.r. A does there exist a QF-module ${}_AM$?

6.2.1. *Character module* Let ${}_AM$ be a finite module and let $M^\flat = \mathrm{Hom}_{\mathbb{Z}}(M, \mathbb{Q}/\mathbb{Z})$ be the *group of rational characters* of the group $(M, +)$ with the natural addition. Then [246]

$$(M^\flat, +) \cong (M, +), \qquad M^{\flat\flat} = M. \tag{6.1}$$

The last equality is a natural identification by understanding the element $x \in M$ as a character of the group $(M^\flat, +)$, where the action is given by $x(\omega) := \omega(x)$ for all $\omega \in M^\flat$.

Since M is a left A-module we can consider $M^\flat$ as a right A-module by defining

$$\forall \omega \in M^\flat,\ a \in A,\ x \in M\colon \quad (\omega a)(x) := \omega(ax) \tag{6.2}$$

Symmetrically, if M_B is a right B-module then $M^\flat$ is a left B-module with $(b\omega)(x) := \omega(xb)$ for all $\omega \in M^\flat$, $x \in M$ and $b \in B$.

For subgroups $N \leqslant (M, +)$, $W \leqslant (M^\flat, +)$ define their *map-annihilators*:

$$N^\perp := \{\omega \in M^\flat\colon \forall x \in N\ \omega(x) = 0\},$$
$$W^\perp := \{x \in M\colon \forall \omega \in W\ \omega(x) = 0\}.$$

Then

$$W^\perp \leqslant (M, +), \qquad N^\perp \leqslant (M^\flat, +), \qquad N^{\perp\perp} = N, \qquad W^{\perp\perp} = W. \tag{6.3}$$

PROPOSITION 6.7. *Let ${}_AM_B$ be a bimodule and $N \leqslant (M, +)$, $W \leqslant (M^\flat, +)$. Then*

$$N \leqslant {}_AM \quad \Leftrightarrow \quad N^\perp \leqslant M^\flat_A, \qquad W \leqslant {}_BM^\flat \quad \Leftrightarrow \quad W^\perp \leqslant M_B.$$

So we have a 1–1 Galois correspondence between submodules of ${}_AM$ (resp. of M_B) and submodules of $M^\flat_A$ (resp. of ${}_BM^\flat$).

6.2.2. *Construction of QF-modules* Let $M = A$ be a f.r. (with identity). Then there are the bimodules ${}_AA_A$ and ${}_AA^\flat_A$. and hence the annihilator of a left (right) ideal of A is a right (left) submodule of $A^\flat$. But even more is true.

PROPOSITION 6.8. *Let $I \leqslant {}_AA$, $J \leqslant A_A$, $L \leqslant {}_AA^\flat$ and $N \leqslant A^\flat_A$ then*

$$\rho_{A^\flat}(I) = I^\perp, \qquad \lambda_{A^\flat}(J) = J^\perp. \qquad \rho_A(L) = L^\perp, \qquad \lambda_A(N) = N^\perp. \tag{6.4}$$

Now we have the following existence theorem for QF-modules.

THEOREM 6.9. *For every f.r. A with identity the module ${}_AA^\flat$ is a QF-module* (*i.e. ${}_AA^\flat_A$ is a QF-bimodule*).

See also [167,198,380].

6.3. *Frobenius rings and bimodules*

6.3.1. *Definitions, characterizations and examples [163,179,193,281,398]* A QF-ring R is called *Frobenius* if ${}_{\overline{R}}\overline{R} \cong {}_{\overline{R}}\mathfrak{S}(R)$, and $\overline{R}_{\overline{R}} \cong \mathfrak{S}(R)_{\overline{R}}$. Any principal ideal f.r. is a Frobenius ring. We have the following characterizations of Frobenius f.r.'s.

A character $\varepsilon \in R^\flat$ is called *left generating* or *left admissible* [74] or *left distinguished* [300] if $R^\flat = R\varepsilon$. The last equality is equivalent to the equality $\lambda_R(\varepsilon) = 0$ which means that the kernel $\operatorname{Ker}\varepsilon = \varepsilon^\perp$ of the homomorphism $\varepsilon : R \to \mathbb{Q}/\mathbb{Z}$ contains no nonzero left ideals. A character that is left and right generating is called a *generating character*.

THEOREM 6.10. *(See* [398].*) For a f.r. R every left* (*or right*) *generating character is generating*, *and the following statements are equivalent*:

(a) *R is a Frobenius ring.*

(b) *R has a* (*left*) *generating character ε.*

(c) *There exists an isomorphism $\varphi : {}_RR \to {}_RR^\flat$.*

(d) *There exists an isomorphism $\psi : R_R \to R^\flat_R$.*

There exist finite non-Frobenius algebras that are, however, QF-algebras [398].

For a f.r. there exists more simple criterion for it to be Frobenius.

THEOREM 6.11. *(See* [179].*) A f.r. R is a Frobenius ring iff $\mathfrak{S}({}_RR)$ is a left principal ideal.*

REMARK. Theorems 6.10, 6.11 give an interesting continuation of the series of

(LEFT) $\Rightarrow$ (RIGHT) theorems for f.r.'s.

1. If R is a f.r. with identity and any two-sided idempotent ideal of R is *left* principal then R is a Wedderburn ring and any idempotent ideal is *right* principal (Theorem 1.7).

2. If R is a f.r. and any two-sided ideal of R is *left* principal then R is a PIR and any two sided ideal of R is *right* principal (Theorem 3.1).

3. If R has a *left* generating character, then R has a *right* generating character (Theorem 6.10).

4. If $\mathfrak{S}({}_RR)$ is a *left* principal ideal then $\mathfrak{S}({}_RR) = \mathfrak{S}(R_R)$ is *right* principal ideal (Theorem 6.11).

In analogy with the definition of a Frobenius ring we call a finite QF-bimodule a ${}_AQ_B$ *Frobenius bimodule*, if

$$_{\overline{A}}\overline{A} \cong {}_{\overline{A}}\mathfrak{S}(Q), \quad \text{and} \quad \overline{B}_{\overline{B}} \cong \mathfrak{S}(Q)_{\overline{B}}; \tag{6.5}$$

The condition (6.5) gives some hard restriction on rings A, B.

PROPOSITION 6.12. *(See [305].) Let ${}_AQ_B$ be a Frobenius bimodule. Then $\overline{A} \cong \overline{B}$.*

Whether in this situation that $A \cong B$ is an *open question*?

The following result proves in particular the existence of Frobenius bimodules.

THEOREM 6.13. *For every f.r. A with identity there exists an isomorphism of bimodules*

$$_{\overline{A}}\overline{A}_{\overline{A}} \cong {}_{\overline{A}}\mathfrak{S}(A^\flat)_{\overline{A}}. \tag{6.6}$$

The submodules ${}_{\overline{A}}\mathfrak{S}(A^\flat)$ and $\mathfrak{S}(A^\flat)_{\overline{A}}$ are cyclic and there exists a common generator ω of these submodules such that $\bar{r}\omega = \omega\bar{r}$ for all $\bar{r} \in \overline{A}$. In particular ${}_AA^\flat_A$ is a Frobenius bimodule.

COROLLARY 6.1. *For every f.r. A with identity there exists a Frobenius (A, A)-bimodule.*

6.3.2. *Reduction from bimodules to rings [182,305]* In order to generalize Theorem 6.11 on finite (A, A)-bimodules we introduce the following generalization to the noncommutative case of a construction from [193, Chapter 12, Ex. 10].

For an arbitrary f.r. A with identity and a finite bimodule ${}_AM_A$ define the *idealizer* $A \ltimes M$ of M by A as the ring $A \times M$ with operations of addition and multiplication defined in the following way: for every $b, b' \in A$, $\beta, \beta' \in M$

$$\begin{aligned} &(b, \beta) + (b', \beta') = (b + b', \beta + \beta'); \\ &(b, \beta) \cdot (b', \beta') = (bb',\ b\beta' + \beta b'). \end{aligned} \tag{6.7}$$

This construction is called also *triangle extension* of A by M. The algebra $(A \ltimes M, +, \cdot)$ is a f.r. with identity. If $I \leqslant {}_AA$ is a left ideal of the ring A, then $I \ltimes M$ is a left ideal of the ring $A \ltimes M$. If J is a left ideal of the ring $A \ltimes M$, then its projection on A: $I = \mathrm{pr}_A(J) = \{b \in A\colon \exists\alpha \in M(b, \alpha) \in J\}$ is a left ideal of the ring A, moreover J is a nilpotent ideal exactly if I is a nilpotent ideal.

PROPOSITION 6.14. *Let A be a f.r. with identity, let M be a finite (A, A)-bimodule and let $R = A \ltimes M$. Then*

(a) *the multiplicative group of the ring R is $R^* = A^* \times M$;*

(b) *the radical of R is $\mathfrak{N}(R) = \mathfrak{N}(A) \ltimes M$;*

(c) $R/\mathfrak{N}(R) \cong A/\mathfrak{N}(A)$;

(d) *the left (right) socle of the ring R is equal to $\mathfrak{S}({}_RR) = 0 \ltimes \mathfrak{S}({}_AM)$ $(\mathfrak{S}(R_R) = 0 \ltimes \mathfrak{S}(M_A))$.*

Now the following characterization of QF-bimodules of type ${}_AM_A$ can be given.

THEOREM 6.15. *A faithful bimodule ${}_AM_A$ over a f.r. A with identity is a (quasi-)Frobenius bimodule if and only if its idealizer $R = A \ltimes M$ is a (quasi-)Frobenius ring.*

6.3.3. *Description of finite Frobenius (A, A)-bimodules [305,308]* The full description of Frobenius bimodules of the form ${}_AQ_A$ with a given finite coefficient ring A is the following. For any fixed $\vartheta \in \operatorname{Aut}(A)$ define a structure of (A, A)-bimodule on the group $(A^\flat, +)$ by the conditions:

$$\forall a \in A,\ \omega \in A^\flat,\ \text{let } a\omega \in A^\flat \text{ be such that } \forall x \in A\colon\ (a\omega)(x) = \omega(xa),$$

$$\text{and let } \omega a \in A^\flat \text{ be such that } \forall x \in A\colon\ (\omega a)(x) = \omega(\vartheta(a)x).$$

We denote this bimodule by ${}_AA^\vartheta_A$. Note that for $\vartheta = 1$ we have ${}_AA^1_A = {}_AA^\flat_A$.

THEOREM 6.16. *For a faithful bimodule ${}_AQ_A$ the following conditions are equivalent:*

(a) *${}_AQ_A$ is a Frobenius bimodule;*

(b) *$\mathfrak{S}({}_AQ) = \mathfrak{S}(Q_A) = \mathfrak{S}$ and ${}_{\bar{A}}\mathfrak{S}_{\bar{A}}$ is a Frobenius bimodule;*

(c) *$\mathfrak{S}({}_AQ)$ is a left cyclic A-module;*

(d) *${}_AQ \cong {}_AA^\flat$;*

(e) *${}_AQ_A \cong {}_AA^\vartheta_A$ for some $\vartheta \in \operatorname{Aut}(A)$.*

The equivalency of statements (a) and (c) is a generalization of Theorem 6.11. Again there is a

$$\text{(LEFT)} \quad \Rightarrow \quad \text{(RIGHT) theorem} \quad \text{(see Section 6.3.1)!}$$

THEOREM 6.17. *Let $\operatorname{Inn}(A)$ be the group of inner automorphisms of the ring A, then*

$$\forall \vartheta, \tau \in \operatorname{Aut}(A)\ \left({}_AA^\vartheta_A \cong {}_AA^\tau_A\right) \quad \Leftrightarrow \quad \left(\vartheta \equiv \tau \left(\operatorname{mod} \operatorname{Inn}(A)\right)\right).$$

The number of classes of isomorphic Frobenius (A, A)-bimodules equals to $|\operatorname{Aut}(A)/\operatorname{Inn}(A)|$.

6.4. *Frobenius bimodules over a commutative f.r.*

Let R be a commutative f.r. with identity. In this case for any faithful module ${}_RM$ there is the evident inclusion $R \leqslant \operatorname{End}({}_RM))$ and we can consider M as a natural bimodule ${}_RM_R$ such that

$$\forall \alpha \in M,\ r \in R\colon \quad \alpha r = r\alpha.$$

Moreover, in this case ${}_R M$ is a QF-module if and only if ${}_R M_R$ is a QF-bimodule. The ring R is a direct sum of local rings (Corollary 1.1); this decomposition implies a decomposition of the (bi-)module M into direct sum of (bi-)modules over local rings. Then ${}_R M$ is a (quasi-) Frobenius module exactly if this is so for every component in the last decomposition.

THEOREM 6.18. *(See [298,308].) For a faithful module ${}_R Q$ over a commutative local f.r. R the following conditions are equivalent*:

(a) *${}_R Q$ is a QF-module*;
(b) ${}_R Q \cong {}_R R^{\flat}$;
(c) *$\mathfrak{S}({}_R Q)$ is an irreducible module*;
(d) *$\mathfrak{S}({}_R Q)$ is a cyclic module*;
(e) *${}_R Q$ is a Frobenius module.*

COROLLARY 6.2. *A faithful module ${}_R Q$ over a commutative f.r. R is quasi-Frobenius if and only if it is a Frobenius module.*

For any (not necessary free) module ${}_R M$ the cardinality of smallest generating system is denoted by $\operatorname{rank}({}_R M)$.

THEOREM 6.19. *(See [298].) Let ${}_R Q$ be a QF-module over a local commutative f.r. R, and $\dim_{\overline{R}} \mathfrak{S}(R) = t$. Then $\operatorname{rank}({}_R Q) = t$. Let $a_1, \ldots, a_t$ be a basis of the space ${}_{\overline{R}}\mathfrak{S}(R)$ and ω be a generator of the socle $\mathfrak{S}({}_R Q)$ of ${}_R Q$, then there exists a system of elements $\alpha_1, \ldots, \alpha_t \in \mathfrak{S}(Q)$ with the property*

$$a_i \alpha_j = \delta_{ij} \omega, \quad i, j \in \{1, \ldots, t\},$$

where δ_{ij} is the Kronecker delta. Any such system generates the module ${}_R Q$.

6.5. *Frobenius rings and symmetric rings [163,298,308,398]*

Note that in general if R is a noncommutative Frobenius ring it is not necessarily the case that there exists isomorphism of Frobenius bimodules ${}_R R_R$ and ${}_R R^{\flat}_R$. Under condition (b) of Theorem 6.10 the isomorphisms φ, ψ in statements (c), (d) can be chosen to be of the form

$$\varphi(a) = a\varepsilon, \qquad \psi(a) = \varepsilon a. \tag{6.8}$$

However in this case we cannot state that bimodules ${}_R R_R$ and ${}_R R^{\flat}_R$ are isomorphic because the isomorphisms (6.8) can be different. A finite ring R is called *symmetric*, if

$${}_R R_R \cong {}_R R^{\flat}_R. \tag{6.9}$$

Of course any symmetric ring is a Frobenius one. We have the following characterization of symmetric rings. Let $K(R) = {}_{\mathbb{Z}}\langle ab - ba \mid a, b \in R\rangle$ be the subgroup of $(R, +)$, generated by all differences $ab - ba$.

THEOREM 6.20. *A f.r. R is symmetric iff it has a generating character $\varepsilon \in R^{\flat}$ such that $\varepsilon(K(R)) = 0$.*

COROLLARY 6.1. *If R is a symmetric ring then $K(R)$ does not contain any nonzero left or right ideals of R.*

The converse of the latter statement is an *open question*.

The class of finite symmetric rings is large enough.

PROPOSITION 6.21. *The following f.r. with identity are symmetric*:

(a) *all finite commutative Frobenius rings* (*in particular all finite commutative PIR*);

(b) *all finite Frobenius rings R with* $\mathrm{Aut}(R) = \mathrm{Inn}(R)$;

(c) *every ring-direct sum of symmetric rings*;

(d) *full matrix rings over symmetric rings*;

(e) *every finite group ring over a symmetric ring.*

COROLLARY 6.2. *Every finite semisimple ring is symmetric.*

Finally note that there exist finite Frobenius nonsymmetric rings.

EXAMPLE 6.22. Let $P = GF(q)$ be a finite field with a nontrivial automorphism σ and let $P[x;\sigma]$ be an Ore polynomial ring with multiplication defined for $a \in P$ by $xa = \sigma(a)x$. Then $R = P[x;\sigma]/(x^2)$ is a GEO-ring (Section 5.2), and hence a Frobenius ring, consisting of elements $\alpha = a_0 + a_1 z$, $a_0, a_1 \in P$, $z = x + (x^2)$ [320]. The unique proper ideal of R is $\mathfrak{N}(R) = Rz = Pz$. For a pair of elements $\alpha \in R$ and $\beta = b_0 + b_1 z \in R$ we have

$$\alpha\beta - \beta\alpha = \big(a_1\big(\sigma(b_0) - b_0\big) + b_1\big(\sigma(a_0) - a_0\big)\big)z.$$

It is evident that the set of all such differences is $Pz = Rz$ and $K(R) = \mathfrak{N}(R)$ is a nonzero ideal. So R is not a symmetric ring.

7. Finite rings with specific conditions [111,262]

7.1. *Rings of fixed order*

A f.r. R is called *decomposable* if it is a direct sum of two nonzero two-sided ideals, and *indecomposable* otherwise. A f.r. with identity is decomposable exactly if it contains a proper central idempotent.

THEOREM 7.1. *(See [111].) Any f.r. is a direct sum of indecomposable ideals*:

$$R = A_1 \oplus \cdots \oplus A_s, \quad s \geqslant 1.$$

Such a decomposition is defined uniquely up to a permutation and isomorphism of summands, and for a ring with identity it is unique up to a permutation of summands.

Here each summand A_i is a p_i- ring for some prime p_i, i.e. a ring of order $p_i^{n_i}$, $n_i \in \mathbb{N}$.

Let $\mathcal{N}(k)$ (respectively $\mathcal{N}_e(k), \mathcal{N}_0(k)$) be the set of classes of isomorphic indecomposable rings (respectively indecomposable rings with identity, indecomposable nilpotent rings) of order k; let $[R]$ be the class of rings isomorphic to R; and let $(G)_0$ be the ring with zero multiplication on Abelian group G.

THEOREM 7.2. *(See [15,104,111].) For any prime p the following equalities hold*:

$$\begin{aligned}
&\left|\mathcal{N}(p)\right| = 2;\\
&\mathcal{N}_e(p) = \left\{\left[GF(p)\right]\right\}; \qquad \mathcal{N}_0(p) = \left\{\left[(\mathbb{Z}_p, +)_0\right]\right\}.\\
&\left|\mathcal{N}\left(p^2\right)\right| = 9;\\
&\mathcal{N}_e\left(p^2\right) = \left\{\left[GF\left(p^2\right)\right],\ \left[GF(p)[x]/\left(x^2\right)\right],\ [\mathbb{Z}_{p^2}]\right\};\\
&\mathcal{N}_0\left(p^2\right) = \left\{\left[(\mathbb{Z}_{p^2}, +)_0\right],\ [p\mathbb{Z}_{p^3}],\ \left[x\, GF(p)[x]/\left(x^3\right)\right]\right\}.\\
&\left|\mathcal{N}\left(2^3\right)\right| = 32, \qquad \left|\mathcal{N}_e\left(2^3\right)\right| = 7, \qquad \left|\mathcal{N}_0\left(2^3\right)\right| = 18;\\
&\quad \textit{for } p \geqslant 3:\\
&\left|\mathcal{N}\left(p^3\right)\right| = 3p + 30, \qquad \left|\mathcal{N}_e\left(p^3\right)\right| = 8, \qquad \left|\mathcal{N}_0\left(p^3\right)\right| = 3p + 15.
\end{aligned}$$

The parameters $|\mathcal{N}_x(p^4)|$ already have different values for the cases: $p = 2, 3$, $p \equiv 1 \pmod 3$, $p \equiv 2 \pmod 3$ [111].

Now there exist descriptions of all:

- rings R of the order p^5 [69,81,82];
- local rings R of characteristic p and order p^6 with $\mathfrak{N}(R)^4 = 0$ [164].

If $|R| = p^k$, where p is a prime and $k \leqslant 4$, then R is a subring of a matrix ring over some commutative ring, but this is false for $k \geqslant 5$ [41,367].

If $n \to \infty$ then the number of semisimple rings of the order p^n is $\exp((\pi^2/9 + O(1))\sqrt{n})$ [201].

Related sources about f.r. with the conditions on the order: [32,45,48,49,70,88,92,101,103,113,128,134,135,151,202,327,328,331,342–344,351,371,383].

7.2. *Nilpotent rings and rings with other conditions*

For $n \to \infty$ we have an asymptotic equivalence of functions [203,222]:

$$\left|\mathcal{N}\left(p^n\right)\right| \sim \left|\mathcal{N}_0\left(p^n\right)\right| \sim p^{(4/27)n^3}.$$

So “almost all” f.r. of a given order are nilpotent.

However nilpotent rings are difficult to investigate. The results of [111,221] give a chance to classify nilpotent rings of order p^4. If R is an indecomposable right principal ideal ring and $R^n = 0$, $R^{n-1} \neq 0$, then R is a commutative chain ring and $|R| = p^n$ for a prime p.

A nilpotent f.r. R of characteristic m can be considered as an ideal of the f.r. $R' = R \times \mathbb{Z}_m$ with identity (0, 1), with componentwise addition and the multiplication defined by the formula

$$(r,k)(s,l) = (rs + rl + ks, kl).$$

Some other results about nilpotent f.r. see in [54,55,60,131,144,161,208,220,346,347, 361,363,372,374].

Let us introduce on a f.r. R an operation called *composition* and denoted $\circ$ by the condition:

$$\forall x, y \in R\colon \quad x \circ y = x + y - xy.$$

Then $(R, \circ)$ is a semigroup with the identity 0. The set R° of all invertible elements of the semigroup $(R, \circ)$ is a subgroup, called the *group of quasi-regular elements* of the ring R. If R contains an identity e, then there exists an isomorphism $\sigma : R^{\circ} \to R^{*}$ by the rule $\sigma(x) = e - x$ [187].

There exists a description of finite rings with any of the following properties:

(a) the group $(R, +)$ or R^0 is cyclic [262];
(b) the group R^* is cyclic (*Gilmer ring* [102,148,321]);
(c) $|D_{\mathrm{r}}(R)|^2 = |R|$ (*Corbas ring* [79]);
(d) the product of any two zero divisors is zero [10,80];
(e) $\operatorname{char} R = p^2$, $\mathfrak{N}(R)^2 = 0$ [10,80];
(f) R is non-nilpotent and $(R, +)$ is a group of type (p^a, p^b), or (p^a, p^b, p) [33,35,36, 104,111,391,392];
(g) $D(R)$ is an ideal of R and $D(R)^3 = 0$ [67,68].

See additionally the results about:

– f.r. with the conditions on $(R, +)$: [15,34,38,39,114,130,312,376,385,390];
– f.r. with the conditions on R^0, R^*, $(R, \cdot)$: [13,14,100,117,311,324,325,349,350,356];
– f.r. with the conditions on $D(R)$, $\mathfrak{N}(R)$: [6,66,132,133,138,139,150,205,250,252,322, 323,382,401];
– rings with the conditions on ideals and subrings: [43,86,149,206,210,248,249,251,339–341,369].

8. Identities and varieties of finite rings

8.1. *Finite rings and Cross varieties*

Let $\mathcal{F} = \mathbb{Z}\langle x_1, x_2, \dots, x_i, \dots\rangle$ be the free algebra on a countable system of independent variables over $\mathbb{Z}$, and let $f(x_1, \dots, x_n) \in \mathcal{F} \setminus \{0\}$. We say that a ring R *satisfies an identity* $f(x_1, \dots, x_n) = 0$ if

$$\forall r_1, \dots, r_n \in R\colon \quad f(r_1, \dots, r_n) = 0.$$

A ring R is called a *polynomial identity ring* (*PI-ring*) if it satisfies some polynomial identity with a coefficient 1. Every f.r. R is PI-ring because for $n > |R|$ it satisfies the *standard identity of degree n*:

$$\sum_{\sigma \in S_n} (-1)^{\sigma} x_{\sigma(1)} \cdots x_{\sigma(n)} = 0.$$

Let $\mathcal{B}$ be the system of left parts of the identities of the ring R and $\mathcal{T}(\mathcal{B})$ be the ideal of $\mathcal{F}$ generated by all polynomials of the form $f(g_1, \dots, g_n)$, where $f \in \mathcal{B}$, $g_1, \dots, g_n \in \mathcal{F}$.

A system $\mathcal{B}$ is called a *basis of identities* of the ring R, if $\mathcal{T}(\mathcal{B}) = T(R)$ is the set of all identities of R.

The class $Var(R)$ of all rings satisfying identities of the set $T(R)$ is called the *variety generated by* R.

A variety V is called *locally finite* if any finitely generated ring $S \in V$ is finite.

A ring is called *critical* if it does not belong to the variety generated by its proper factors.

THEOREM 8.1. *(See [219,254].) A variety $Var(R)$, generated by some f.r. R is a* Cross variety, *i.e. it is a locally finite variety with a finite basis of identities, containing only a finite set of pairwise nonisomorphic critical rings, each of which is finite. Vice versa, any Cross variety in the class of associative rings is generated by some f.r.*

THEOREM 8.2. *(See [219,254].) A variety V of rings is a Cross one if and only if the ideal $T(V)$ of its identities in $\mathcal{F}$ contains polynomials of the following form: mx, for some $m \in \mathbb{N}$, and $x_1 \cdots x_n - f(x_1, \ldots, x_n)$, where $f \in F$ is a polynomial with monomials of degree not less than $n + 1$. In this case every nilpotent ring A of the variety V satisfies the equality $A^n = 0$ and the ideal $T(V)$ contains some polynomial $x^k - x^l$, $1 \leqslant k < l$. Such a variety is generated by its critical rings.*

THEOREM 8.3. *(See [269,271].) Critical rings are subdirect indecomposable. A f.r. R with identity is critical exactly if the matrix ring $R_{n,n}$ is critical.*

THEOREM 8.4. *(See [292].) A finite principal ideal ring is critical if and only if it is primary, i.e. is a full matrix ring over a chain ring.*

One of the main ways to prove that a given f.r. R is critical is to build a *critical polynomial* for this ring, i.e. a polynomial which is an identity for any proper subring and quotient ring of R but does not belong to the ideal $T(R)$ of all identities of R. For example a critical polynomial of the field $R = GF(q)$, $q = p^r$, is

$$f(x) = \prod_{t \mid r, t \neq r} \left(x^{p^t} - x\right),$$

THEOREM 8.5. *(See [292].) Let S be a chain ring with residue field $\bar{S} = GF(q)$ and nilpotency index of its radical equal to n. Then $f(x)(y^q - y)^{n-1}$ is a critical polynomial of S.*

Let $\mathcal{N}_{\text{crit}}C(k)$ be the set of classes of isomorphic critical rings of order k. Using the description of f.r.'s of orders p, p^2, p^3 (Theorem 7.2) it is possible to describe the critical rings of these orders.

THEOREM 8.6. *For a prime p*

$$\mathcal{N}_{\text{crit}}(p) = \left\{\left[GF(p)\right], \left[(\mathbb{Z}_p, +)_0\right]\right\},$$

$$\left|\mathcal{N}_{\text{crit}}\left(p^2\right)\right| = 8, \qquad \left|\mathcal{N}_{\text{crit}}(8)\right| = 18, \qquad \left|\mathcal{N}_{\text{crit}}\left(p^3\right)\right| = 3p + 14, \quad \textit{for } p \geqslant 3.$$

Related sources: [28,29,42,255,274,314–316,384].

8.2. *Bases of identities of some finite rings*

The bases of identities (BI) are described only for a few classes of rings. For a field $GF(q)$ a BI in the class of associative commutative rings consists of the polynomials $x^q - x$, px.

THEOREM 8.7. *(See [270].) A basis of the identities of the matrix ring $M_2(GF(q))$ is described by polynomials*

$$(xy)z - x(yz), \qquad (x^q - x)(y^{q^2} - y)([x, y]^{q-1} - 1),$$
$$(x^q - x) \circ (y^q - y))^q - (x^q - x) \circ (y^q - y),$$

where $[x, y] = xy - yx$, $x \circ y = xy + yx$.

There are also known BI's of the rings $M_n(GF(q))$, for $n = 3, 4$ [145–147]; and $M_2(GR(q^2, p^2))$ [309,310].

See also [257,260,368].

8.3. *Identities of Galois and GE-rings*

Let $R = GR(q^n, p^n)$ be a Galois ring. Consider the polynomials $\Psi_j(x) \in \mathcal{F}$ that are defined recursively by the equalities

$$\Psi_0(x) = x^q - x, \qquad \Psi_t(x) = \Psi_{t-1}(x)^q - p^{q^t - 1}\Psi_{t-1}(x), \quad t \geqslant 1.$$

THEOREM 8.8. *(See [286].) The set of polynomials*

$$p^j \Psi_0(x_1) \cdots \Psi_0(x_u)\Psi_1(y_1) \cdots \Psi_1(y_v) \cdots \Psi_t(z_1) \cdots \Psi_t(z_w),$$

satisfying the conditions

$$j + \frac{1}{q-1}\big(u(q-1) + v(q^2 - 1) + \cdots + w(q^t - 1)\big) = n,$$

is a basis of the identities of a ring $R = GR(q^n, p^n)$ in the class of associative–commutative rings of the characteristic p^n.

A description of bases of identities of GE-rings with parameters (5.4) in the case $\varepsilon > 1$ is an *open question.*

We can only state that the system of parameters (5.4) does not define a basis of GE-ring identities uniquely. Indeed, let $R = GR(4^2, 2^2)$ and let $\alpha \in R$ be an element such that $\alpha^2 - \alpha \notin 2R$. Then

$$S_1 = \mathcal{R}/(x^3 - 2, x^5), \qquad S_2 = \mathcal{R}/(x^3 - 2\alpha, x^5)$$

are two GE-rings with the same parameters (5.4): $q = 4$, $n = 5$, $d = 2$, $\varepsilon = 3$. However, the polynomial $f(x) = \Psi_0(x)^4 - 2\Psi_0(x)$ is an identity for S_1, but not an identity for S_2.

In view of this remark it is interesting to solve the following problems.

PROBLEMS.

1. To describe systems of parameters q, n, d such that any two GE-rings with these parameters have the same identities;

2. To describe identities of the variety $\mathcal{V}(GE(q,n,d))$ generated by all GE-rings with parameters q,n,d;
3. To estimate the number of different varieties $\mathcal{V}(R)$, where R is a GE-ring with the parameters q,n,d (weak isomorphism problem).

8.4. *Generalized identities of GE-rings*

Let S be GE-ring with the same parameters (5.4), and let $S_0[X] = S_0[x_1,x_2,\ldots]$ be the associative-commutative ring of polynomials in a countably set of variables $X = \{x_1,x_2,\ldots,y_1,y_2,\ldots,z_1,z_2,\ldots\}$ with zero constant term.

A polynomial $f(X) = f(x_1,\ldots,x_n) \in S_0[X]$ is called *generalized identity* of the ring S if

$$\forall a_1,\ldots,a_n \in S{:}\quad f(a_1,\ldots,a_n) = 0.$$

Let B be a system of generalized identities of the ring S and let $T(B)$ be the ideal of $S_0[X]$ generated by all polynomials of the form $g(f_1(X),\ldots,f_m(X))$, where $g(x_1,\ldots,x_m) \in B$, $f_1(X),\ldots,f_m(X) \in S_0[X]$.

A system B is called a *basis of generalized identities* of the ring S (in the variety of associative-commutative rings), if $T(B)$ is the set of all generalized identities of the ring S.

THEOREM 8.9. *(See [287].) Under the conditions (5.23), (5.24) the system of polynomials*

$$\pi^j\Phi_0(x_1)\cdots\Phi_0(x_u)\Phi_1(y_1)\cdots\Phi_1(y_v)\cdots\Phi_t(z_1)\cdots\Phi_t(z_w), \tag{8.1}$$

where $j+u\delta_0+v\delta_1+\cdots+w\delta_t = n$, is a basis of generalized identities of the ring S in the variety of associative-commutative rings.

9. Complete system of functions over finite rings

Let A be a set of cardinality $|A| = k$, and let P_A be the set of all functions $f: A^n \to A$, $n \in \mathbb{N}$. A *selector* or *projection* is any function $\varepsilon_i^n : A^n \to A$ of the form $\varepsilon_i^n(x_1\cdots x_n) = x_i$. Denote by $\mathcal{E}$ the set of all selectors. Say that x_i is a *fictitious* variable of a function $f(x_1,\ldots,x_n)$ if $f(a_1,\ldots,a_{i-1},x_i,a_{i+1},\ldots,a_n) =$ constant for every $a_1,\ldots,a_{i-1},a_{i+1},\ldots,a_n \in A$.

The *closure* of a subset $G \subseteq P_A$ is defined as a minimal subset $[G] \subseteq P_A$ with the properties:

1. $G \cup \mathcal{E} \subseteq [G]$;
2. if $f(x_1,\ldots,x_n), g_1,\ldots,g_n \in [G]$, then $f(g_1,\ldots,g_n) \in [G]$;
3. if $f(x_1,\ldots,x_n) \in [G]$, then any function obtained from f by addition or deleting of a fictitious variable belongs to $[G]$.

A set of functions $G \subseteq P_A$ is called *complete*, if $[G] = P_A$.

We shall say that a system G of functions *preserves an equivalence relation* ρ on A, if for every function $f(x_1,\ldots,x_n) \in G$ and for any $(a_1,\ldots,a_n), (b_1,\ldots,b_n) \in A^n$

$$\bigl(a_i\rho b_i,\ i \in \{1,\ldots,n\}\bigr) \quad\Rightarrow\quad \bigl(f(a_1,\ldots,a_n)\rho f(b_1,\ldots,b_n)\bigr).$$

If G is a full system, then it does not preserve any nontrivial equivalence relation on A.

Let now $(A, +, \cdot)$ be a finite, not necessary associative, ring. Then any complete system G of functions on A satisfy the following conditions.

F1. System G does not preserve nontrivial congruences on A.

F2. Class $[G]$ contains functions $x + y$ and the set const of all constant functions.

F3. Class $[G]$ contains all functions $ax, xa, a \in A$.

F4. Class $[G]$ contains function xy.

THEOREM 9.1. *A system G of functions on a finite, not necessarily associative, ring A with nonzero multiplication is complete iff it satisfies the conditions* F1, F2, F4.

A ring is called *polynomially complete* if the system $G = \{xy, x + y, \text{const}\}$ is complete.

COROLLARY 9.1. *(See [290].) A finite, not necessarily associative, ring A with nonzero multiplication is polynomially complete iff it is simple.*

For associative rings see [57,183,247] (firstly this result was first obtained by A.V. Kuznetsov in 1955 [258, p. 106]).

We call a ring A with nonzero multiplication *linearized* if for any $G \subseteq P_A$

$$(F1, F2, F3) \Rightarrow F4,$$

i.e. if completeness of G is equivalent to F1–F3.

THEOREM 9.2. *(See [289].) Let A be a ring of order p^n, $n > 1$, with cyclic additive group and $G \subset P_A$. Then*

$$[G] = P_A \quad \Leftrightarrow \quad (F1, F2).$$

THEOREM 9.3. *(See [290].) A matrix ring $M_n(R)$ is linearized iff R is linearized.*

THEOREM 9.4. *(See [290].) Let R_1, R_2 be finite linearized rings and each of them contains an identity or $(|R_1|, |R_2|) = 1$. Then $R = R_1 \oplus R_2$ is a linearized ring.*

THEOREM 9.5. *(See [290].) Let $S_1, \dots, S_t$ be local quasi-Frobenius rings, such that each of them has a non-simple characteristic. Then the ring (1.4) is linearized.*

See also [288,329].

II. Applications of finite rings

10. Standard bases of polynomial ideals over a commutative chain ring

In this section R is a commutative chain f.r. (GE-ring, Section 5.1). Since R is a commutative local ring all results of Section 2.3 for polynomials in $R[x]$ are valid. Here we discuss "deeper" aspects of polynomials and ideals in $R[x]$ and $R[\mathbf{x}] = R[x_1, \dots, x_k]$ allowing the calculation certain combinatorial parameters connected with some applications of polynomial rings.

10.1. *Canonical generating systems of a polynomial ideals of one variable*

Suppose that the nilradical of the ring R has the form $\mathfrak{N} = \mathfrak{N}(R) = \pi R$ and has nilpotency index $n > 1$. Then the lattice of the ideals of the ring R is a chain

$$R \rhd \pi R \rhd \cdots \rhd \pi^{n-1} R \rhd \pi^n R = 0. \tag{10.1}$$

Let's also fix the parameters

$$p = \operatorname{char} \overline{R}, \qquad |\overline{R}| = |R/\pi R| = q = p^r. \tag{10.2}$$

10.1.1. *Main properties of canonical generating systems (CGS)* The well-known Hilbert theorem state that any ideal $I \lhd R[x]$ has a finite generating system. Moreover under condition (10.1) the ideal I has such a system of cardinality not more than n (see, e.g. [116, Theorem 2, Section 17.11]). Below there is a revised version of this result based on results of [211], [357, Section 8].

Let us define the *norm* of an element $r \in R$, of a polynomial $G(x) = g_m x^m + \cdots + g_1 x + g_0 \in \mathcal{R}$ and of a subset $M \subseteq \mathcal{R}$ respectively by the equalities

$$\|r\| = \max\{i \in \{0, \dots, n\}\colon r \in \pi^i R\}, \qquad \|G\| = \min\{\|g_s\|\colon s \in \{0, \dots, m\}\},$$
$$\|M\| = \min\{\|G\|\colon G \in M\}. \tag{10.3}$$

We call a polynomial $G(x)$ of degree m a *right* polynomial, if $\|g_m\| \leqslant \|g_s\|$ (i.e. $g_s \in g_m R$) for $s \in \{0, \dots, m\}$.

PROPOSITION 10.1. *A right polynomial $G(x) \in \mathcal{R}$ of degree m divides a polynomial $F(x) = \sum f_i x^i \in \mathcal{R}$ with remainder exactly if $\|f_i\| \geqslant \|g_m\|$ for $i \geqslant m$. Moreover, the remainder* $\operatorname{Res}(F/G)$ *of the division of $F(x)$ by $G(x)$ is uniquely defined.*

We say that $G(x)$ *divides* $F(x)$ *modulo* π^d and write $G(x) \mid F(x) \bmod \pi^d$, if $F = QG + \pi^d H$, where $Q, H \in \mathcal{R}$.

THEOREM 10.2. *(See [226,295].) Let I be a nonzero ideal in $\mathcal{R}$ and $\|I\| = a_0$. Then I contains a system of $t + 1 \leqslant n - a_0$ right polynomials*

$$G_0(x), \ \dots, \ G_t(x) \in I, \quad \|G_s\| = a_s, \qquad \deg G_s(x) = m_s, \quad s \in \{0, \dots, t\}, \tag{10.4}$$

with the properties:

(C1) $\|I\| = a_0 < a_1 < \cdots < a_t < n$;
(C2) $m_0 > m_1 > \cdots > m_t \geqslant 0$;
(C3) *if $F(x) \in I$ and* $\deg F(x) < m_s$, $s \geqslant 0$, *then $F(x) = 0$ if $s = t$ and $\|F(x)\| \geqslant a_{s+1}$ if $s < t$.*

Any such system of polynomials has also the following additional properties:

(C4) *If $F(x) \in I$ and $\|F\| \geqslant a_s$, then*

$$G_s(x) \mid F(x) \bmod \pi^{a_{s+1}} \tag{10.5}$$

(*here we suppose that* $a_{t+1} = n$) *and*

$$F(x) \in \big(G_s(x),\ G_{s+1}(x), \dots, G_t(x)\big). \tag{10.6}$$

(C5) *There is the equality*

$$I = \big(G_0(x), G_1(x), \dots, G_t(x)\big), \tag{10.7}$$

and every polynomial $F(x) \in I$ *can be presented as a sum*

$$F(x) = Q_0(x)G_0(x) + \cdots + Q_t(x)G_t(x), \tag{10.8}$$

such that $\deg Q_s G_s < m_{s-1}$ *for* $s \in \{1, \dots, t\}$, *and the summands in this representation are defined uniquely.*

(C6) *If* $s \in \{0, \dots, t\}$ *and* $a_s \leqslant a < a_{s+1}$, *then*

$$I \cap \pi^a \mathcal{R} = \big(\pi^{a-a_s} G_s(x), G_{s+1}(x), \dots, G_t(x)\big), \tag{10.9}$$

$$\big(I : \pi^a\big) = \big(F_s(x), \pi^{a_{s+1}-a} F_{s+1}(x), \dots, \pi^{a_t - a} F_t(x), \pi^{n-a}\big), \tag{10.10}$$

where $F_0(x), \dots, F_t(x)$ *are polynomials with invertible highest coefficients, satisfying the conditions*

$$G_s(x) = \pi^{a_s} F_s(x), \quad s \in \{0, \dots, t\}. \tag{10.11}$$

(C7) *Under the condition* (10.11) *for any polynomial* $H(x) \in \mathcal{R}$ *there exists a unique representation in the form*

$$H(x) = H_0(x)F_0(x) + \cdots + H_t(x)F_t(x) + H_{t+1}(x), \tag{10.12}$$

where $\deg H_s(x)F_s(x) < m_{s-1}$ *for* $s \in \{1, \dots, t\}$, $\deg H_{t+1}(x) < m_t$, *and* $H(x) \in I$ *is equivalent to the system of inequalities*

$$\big\| H_s(x) \big\| \geqslant a_s, \quad s \in \{0, \dots, t+1\}, \ \textit{where } a_{t+1} = n. \tag{10.13}$$

Any system of polynomials (10.4) with the properties (C1)–(C3) is called a *canonical generating system* (CGS) of the ideal I. The proof of this theorem in [295, Proposition 13] contains an algorithm for building such a CGS.

10.1.2. *The cardinality of a quotient ring of a polynomial ring* An ideal $I \lhd \mathcal{R}$ is called *monic*, if it contains some monic polynomial. Evidently, this condition is equivalent to the condition $\|I\| = 0$ and to the condition $|\ \mathcal{R}/I\ | < \infty$.

THEOREM 10.3. *(See* [226, *Corollary* 16.7].*) A monic ideal* $I \lhd \mathcal{R}$ *has a CGS of the form*

$$F_0(x), \quad \pi^{a_1} F_1(x), \quad \dots, \quad \pi^{a_t} F_t(x), \tag{10.14}$$

were $F_s(x)$ *is a monic polynomial of degree* m_s, $s \in \{0, \dots, t\}$,

$$m_0 > m_1 > \cdots > m_t \geqslant 0, \quad 0 < a_1 < a_2 < \cdots < a_t < n.$$

Under this condition the cardinality of the quotient ring $\mathcal{R}/I$ *is*

$$|\mathcal{R}/I| = q^{(m_0-m_1)a_1 + \cdots + (m_{t-1}-m_t)a_t + m_t n}. \tag{10.15}$$

10.1.3. *Decomposition of a polynomial ideal into its product of primary ideals* The CGS of a polynomial ideal $I \lhd \mathcal{R}$ makes it possible to construct the primary components of I.

Let us recall that an ideal I is called *primary* if for any $K(x), H(x) \in \mathcal{R}$ the condition $K(x)H(x) \in I$, $K(x) \notin I$ implies that $H(x)^r \in I$ for some $r \in \mathbb{N}$. A polynomial $G(x) \in \mathcal{R}$ is called primary if $G(x)\mathcal{R}$ is a primary ideal. The last condition is equivalent to that $\overline{G}$ is a power of some irreducible polynomial over the field $\overline{R}$ (see Section 2.3.3) or $G(x) = \pi^s U(x)$ for some $s \in \{0, \ldots, n\}$, $U(x) \in \mathcal{R}^*$.

PROPOSITION 10.4. *An ideal I with CGS (10.4) is a primary exactly if either $a_0 = 0$ and $G_0(x)$ is a primary monic polynomial or* $\deg G_0(x) = 0$ (*i.e.* $I = (\pi^{a_0})$).

It is well known (see e.g. [24,403]), that any ideal $I \lhd \mathcal{R}$ is the intersection (product) of a finite set of two by two coprime primary ideals called *primary components of the ideal I*. If I has a CGS

$$\pi^{a_0} F_0(x), \quad \pi^{a_1} F_1(x), \quad \ldots, \quad \pi^{a_t} F_t(x), \tag{10.16}$$

where the polynomials $F_s(x)$ are monic, the canonical generating systems of its primary components can be constructed based on the following results.

In accordance with property (C6), the equality

$$I = \left(\pi^{a_0}\right) \cap \left(F_0(x), \pi^{a_1} F_1(x), \ldots, \pi^{a_t} F_t(x)\right). \tag{10.17}$$

is true.

PROPOSITION 10.5. *Let an ideal $I \lhd \mathcal{R}$ have a CGS (10.14) and let $F_0(x)$ be a product of coprime polynomials*

$$F_0(x) = K_0(x)H_0(x), \quad \left(K_0(x), H_0(x)\right) = e. \tag{10.18}$$

Then for every $s \in \{1, \ldots, t\}$ the polynomial $F_s(x)$ has a unique representation in the form

$$F_s(x) = K_s(x)H_s(x), \tag{10.19}$$

where $K_s(x)$, $H_s(x)$ are monic polynomials with the properties

$$\overline{K}_s(x) \mid \overline{K}_0(x), \qquad \overline{H}_s(x) \mid \overline{H}_0(x), \tag{10.20}$$

Moreover, $(K_s(x), H_s(x)) = e$ *and the ideal I is the intersection of two coprime ideals*

$$\begin{aligned} &I = \mathcal{K} \cap \mathcal{H}, \quad \mathcal{K} + \mathcal{H} = (e), \quad \textit{where} \\ &\quad \mathcal{K} = \left(K_0(x), \pi^{a_1} K_1(x), \ldots, \pi^{a_t} K_t(x)\right), \\ &\quad \mathcal{H} = \left(H_0(x), \pi^{a_1} H_1(x), \ldots, \pi^{a_t} H_t(x)\right). \end{aligned} \tag{10.21}$$

A CGS of the ideal $\mathcal{K}$ obtained from the system of polynomials

$$K_0(x), \quad \pi^{a_1} K_1(x), \quad \ldots, \quad \pi^{a_t} K_t(x) \tag{10.22}$$

by deleting of every polynomial $\pi^{a_s} K_s(x)$ with $\deg K_s(x) = \deg K_{s-1}(x)$.

So if the ideal I of the ring $\mathcal{R}$ has a CGS (10.14), and the polynomial $F_0(x)$ has canonical decomposition (see Section 2.3.3) of the form $F_0(x) = F_0^{(1)}(x) \cdots F_0^{(r)}(x)$, then every one of the polynomials $F_s(x)$ has the form $F_s(x) = F_s^{(1)}(x) \cdots F_s^{(r)}(x)$, where $\overline{F}_s^{(i)}(x) \mid \overline{F}_0^{(j)}(x)$ for $j \in \{1, \dots, r\}$, and the ideal I has the following representation as an intersection (product) of two by two comaximal primary ideals:

$$I = I^{(1)} \cap \cdots \cap I^{(r)} = I^{(1)} \cdot \cdots \cdot I^{(r)}, \quad \text{where}$$
$$I^{(j)} = \left(F_0^{(j)}(x),\ \pi^{a_1} F_1^{(j)}(x),\ \dots,\ \pi^{a_t} F_t^{(j)}(x)\right), \quad j \in \{1, \dots, r\}.$$

10.2. *Standard bases coordinated with the norm (multivariate case)*

From the application point of view one of the most important problems in the theory of ideals in a commutative ring $R[X] = R[x_1, \dots, x_k]$ is the membership problem: for an ideal $I = (\chi) \lhd R[X]$, given by some generating system χ, to verify whether $F(X) \in I$, where $F(X) \in R[X]$. This whether problem was solved for ideals in Lee algebras [348] and for polynomial ideals over fields [59] by the construction of some standard bases, called also Gröbner bases of the ideal. The same Shirshov–Buchberger algorithm allows to find standard bases in any polynomial ideal over a Noetherian ring (see, e.g, [1]). We call this algorithm a *formal generalization* of the Shirshov–Buchberger algorithm. In this section a modification of the algorithm mentioned is presented, giving standard bases of polynomial ideals over a commutative chain f.r. coordinated with the norm on the ring R (see Section 10.1). These bases, named *coordinated standard bases* (CSB's), can be viewed as a generalization of Gröbner bases for fields and a generalization of the CGS from Section 10.1. Using CSB's one can solve not only membership problem, but also some other classical computational problems: listing of representatives of cosets of the ideal I, building of the set of generators of syzygy module, decision of the systems of the polynomial equations, etc. ([158,159,277,307]) (see Sections 10.2.3, 10.2.4).

10.2.1. *Definition and characterization of coordinated standard bases* Recall that a reflexive transitive and antisymmetric binary relation $\preccurlyeq$ on a set M is called an *order* on M. An order $\preccurlyeq$ is called *linear* if any two elements of M are comparable. A linear order $\preccurlyeq$ is called a *well-ordering* if it satisfies the descending chain condition.

Fix the set of variables $X = \{x_1, \dots, x_k\}$, $k \geqslant 1$. Let $[X\rangle = [x_1, \dots, x_k\rangle$ be the semigroup of commuting monomials over X. A linear order on $[X\rangle$ is called *admissible* if it is multiplicative ($\forall u, v, w \in [X\rangle$: $u \preccurlyeq v \Rightarrow uw \preccurlyeq vw$) and it is a well-ordering (equivalently $\forall u \in [X\rangle$: $1 \preccurlyeq u$, see, e.g., [89]). The most used admissible orders are the following:

- *The lexicographic order* (*lex*): $x_1^{i_1} \cdots x_k^{i_k} \preccurlyeq_{\text{lex}} x_1^{j_1} \cdots x_k^{j_k}$ if and only if the first non-zero number in the row $j_1 - i_1, j_2 - i_2, \dots, j_k - i_k$ is positive.
- *The degree-lexicographic order* (*deglex*): $x_1^{i_1} \cdots x_k^{i_k} \preccurlyeq_{\text{deglex}} x_1^{j_1} \cdots x_k^{j_k}$ if and only if the first non-zero number in the row $(j_1 + \cdots + j_k) - (i_1 + \cdots + i_k), j_1 - i_1, j_2 - i_2, \dots, j_k - i_k$ is positive.
- *The degree-reversed-lexicographic order* (*degrevlex*): $x_1^{i_1} \cdots x_k^{i_k} \preccurlyeq_{\text{degrevlex}} x_1^{j_1} \cdots x_k^{j_k}$ if and only if the first non-zero number in the row $(j_1 + \cdots + j_k) - (i_1 + \cdots + i_k), i_k - j_k, i_{k-1} - j_{k-1}, \dots, i_1 - j_1$ is positive.

Let $R[X] = R[x_1, \dots, x_k]$ be a polynomial algebra over a commutative chain f.r. R with parameters (10.1), (10.2). The subsemigroup

$$[R, X\rangle = \{au \mid a \in R,\ u \in [X\rangle\}$$

of the semigroup $(R[X], \cdot)$ is called a *semigroup of terms*. Any admissible order on $[X\rangle$ can be extended to $[R, X\rangle$ as follows: for $a, b \in R$ and $u, v \in [X\rangle$:

$$(au \preccurlyeq bv) \quad \Leftrightarrow \quad \big(\big(\|a\| > \|b\|\big),\ \text{or}\ \big(\|a\| = \|b\|,\ \text{and}\ u \preccurlyeq v\big)\big), \tag{10.23}$$

where $\|a\|$ is defined in (10.3). Any non-zero polynomial $F \in R[X]$ can be represented as

$$F = a_1u_1 + a_2u_2 + \cdots + a_nu_n, \quad a_1u_1 \succ a_2u_2 \succ \cdots \succ a_nu_n, \tag{10.24}$$

where $a_i \in R \setminus 0$ and the $u_i \in [X\rangle$ are pairwise distinct monomials. The *leading term*, the *leading monomial* and the *leading coefficient* of the polynomial F are defined to be

$$Lt(F) = a_1u_1, \qquad Lm(F) = u_1, \quad \text{and} \quad Lc(F) = a_1.$$

It is assumed that $Lt(0) = Lc(0) = 0$. Note that for any $F \in R[X]$ and $U \in [R, X\rangle$: $Lt(UF) = U\,Lt(F)$ [159]. (This important property does not hold if we consider a formal generalization of Shirshov–Buchberger algorithm. Such a generalization does not take into account the norms of the coefficients of terms and instead of (10.23) the order used on $[R, X\rangle$ is of the following form: $(au \preccurlyeq bv) \Leftrightarrow ((a = 0)$, or $(b \neq 0$, and $u \preccurlyeq v)).)$

The coefficient of a monomial $v \in [X\rangle$ in a polynomial $F \in R[X]$ is denoted by $Cf(F, v)$. Let $F, G, H \in R[X]$ with $F \neq 0$, $G \neq 0$, $Lt(G) = au$ and let $\chi \subset R[X] \setminus 0$. It is said that:

(a) *F reduces to H modulo G by eliminating $v \in [X\rangle$* (notation $F \xrightarrow{G} H[v]$) if $Cf(F, v) = c \neq 0$, and there exist $b \in R$ and $w \in [X\rangle$ such that $c = ba$, $v = wu$ and $H = F - bwG$,

(b) *F reduces to H modulo G* (notation $F \xrightarrow{G} H$) if $F \xrightarrow{G} H[v]$ for some $v \in [X\rangle$,

(c) *F reduces to H modulo χ* (notation $F \xrightarrow{\chi} H$) if $F \xrightarrow{G} H$ for some $G \in \chi$.

F is called *normal modulo χ* if F cannot be reduced modulo χ.

Let now $* \xrightarrow{\chi}$ be the reflexive-transitive closure of the relation $\xrightarrow{\chi}$. A polynomial H is called a *normal form* of *F modulo χ* if it is normal modulo χ and $F* \xrightarrow{\chi} H$. The set of all normal forms of F modulo χ denoted by $\mathrm{NF}(F, \chi)$. Let $F, G \in R[X] \setminus 0$ and

$$Lt(F) = \alpha\pi^a u, \qquad Lt(G) = \beta\pi^b v, \quad \text{where}$$
$$\alpha, \beta \in R^*,\ a, b \in \{0, \dots, n-1\},\ u, v \in [X\rangle.$$

Let $w = g.c.d.(u, v) \in [X\rangle$ and $c = \max\{a, b\}$. Then there exist monomials $u', v' \in [X\rangle$, such that $u = wu'$ and $v = wv'$. The polynomial

$$S(F, G) = \pi^{c-a}v'\alpha^{-1}F - \pi^{c-b}u'\beta^{-1}G \tag{10.25}$$

is called the *S-polynomial* of F and G.

The following theorem is an analog of the composition lemma from the theory of Gröbner bases over fields, called also diamond lemma (see, e.g., [1,89]).

THEOREM 10.6. *(See [158,159].) Let χ be a nonempty subset of an ideal $I \lhd R[X]$. Then the following assertions are equivalent*:

(a) $F* \xrightarrow{\chi} 0$ *for any* $F \in I$.

(b) *For any* $F \in I$ *there exists* $G \in \chi$ *such that* $Lt(G)$ *divides* $Lt(F)$.

(c) $I = (\chi)$ *and* $S(G_1, G_2) * \xrightarrow{\chi} 0$ *for all* $G_1, G_2 \in \chi$.

(d) *Any* $F \in I$ *can be represented as* $F = \sum_{\alpha=1}^{m}(\sum_{k=1}^{m_\alpha} a_\alpha^k u_\alpha^k)G_\alpha$ *where* $a_\alpha^k \in R$, $u_\alpha^k \in [X\rangle$, $G_\alpha \in \chi$, *and* $Lt(a_\alpha^k u_\alpha^k G_\alpha) \preccurlyeq Lt(F)$.

We call the last representation of F a *homogeneous representation* or H-representation. In the notation of the Theorem 10.6 the set χ is called a *standard basis* of the ideal I coordinated with the norm (*coordinated standard basis* (CSB in brief)).

Condition (c) from Theorem 10.6 gives the possibility to construct an effective procedure determining whether a finite polynomial set $\chi \subset R[X]$ is a CSB of the ideal (χ). Moreover, given a finite set of polynomials ψ, the Algorithm 10.7 will construct a CSB of the ideal (χ) (see [158,159]). This algorithm is analogous to the familiar algorithm for fields [1,89].

ALGORITHM 10.7 *(Construction of a CSB).*

```
**input:** A finite set of non-zero polynomials $\psi = \{F_1, F_2, \ldots, F_s\}$.
**output:** A CSB $\chi$ of the ideal $(\psi)$.
$\chi := \psi, \mathcal{G} := \{(F_i, F_j) \mid 1 \leqslant i < j \leqslant s\}$
**while** $\mathcal{G} \neq \emptyset$ **do**
    Choose any $(F, G) \in \mathcal{G}, \mathcal{G} = \mathcal{G} \setminus \{(F, G)\}$.
    Find a normal form $H$ of $S(F, G)$ modulo $\chi$.
    **if** $H \neq 0$ **then**
        $\mathcal{G} := \mathcal{G} \cup \{(U, H) \mid U \in \chi\}$
        $\chi := \chi \cup \{H\}$
    **end**
**end**
**return** $\chi$
```

The generating system of the ideal (ψ) constructed by this algorithm is in general excessive. For example it contains in any case the initial generating system ψ. A polynomial system $\chi \subseteq R[X]$ is called *minimal* if for any polynomial $G \in \chi$ the ideal $(\chi \setminus \{G\})$ is not equal to (χ). Characterizations of minimal coordinated standard bases of a given ideal are connected with the following notions.

Let us define the *leading π-monomial* of $F(X) \in R[X]$ by $LM(F) = \pi^{\|Lc(F)\|} Lm(F)$ and the *set of obstructions* of an ideal $I \lhd R[X]$ as the set $\mathcal{O}(I)$ of all minimal elements in the partially ordered set $(LM(I), \,|)$.

THEOREM 10.8. *A system $\chi \subseteq I$ is a CSB of I precisely when $\mathcal{O}(I) \subseteq LM(\chi)$.*

A polynomial $G \in R[X]$ is called *self-normal* if G can be reduced modulo G only to 0. A polynomial system $\chi \subseteq R[X]$ is called *reduced* if any polynomial $G \in \chi$ is self-normal and normal modulo $\chi \setminus \{G\}$;

THEOREM 10.9. *Any ideal $I \lhd R[X]$ has a reduced CSB. Any reduced CSB of I is minimal. A CSB χ of an ideal I is minimal if and only if $|\chi| = |\mathcal{O}(I)|$.*

10.2.2. *Canonical generating system of a polynomial ideal in k variables* If $k = 1$ then a CGS of an ideal $I \lhd \mathcal{R}$ (see Section 10.1) is exactly a reduced CSB.

The notion of a CGS generalizes to the case $k > 1$ in the following way [277,307]. A polynomial system φ is called *π-homogeneous polynomial system* if $\|F\| = \|\varphi\|$ for any $F \in \varphi$. Any set χ of polynomials can be represented as a disjoint union of π-homogeneous subsets:

$$\chi = \pi^{a_0}\chi_0 \cup \cdots \cup \pi^{a_t}\chi_t, \quad \|\chi_s\| = 0,\ s \in \{0, \ldots, t\}. \tag{10.26}$$

Let $Lm(\chi_s)[X\rangle$ be the ideal of the semigroup $[X\rangle$ generated by the set of monomials $Lm(\chi_s) = \{Lm(G) \mid G \in \chi_s\}$ and let $\mathcal{F}_s \subseteq \mathbb{N}_0^k$ be the subset consists of all rows $(i_1, \ldots, i_k) \in \mathbb{N}_0^k$ such that $x_1^{i_1} \cdots x_k^{i_k} \in [X\rangle \setminus Lm(\chi_s)[X\rangle$. Then $\mathcal{F}_s$ is a *Ferrer diagram*, i.e. for any $\mathbf{i} = (i_1, \ldots, i_k)$, $\mathbf{j} = (j_1, \ldots, j_k) \in \mathbb{N}_0^k$ if $\mathbf{j} \leqslant \mathbf{i}$ (coordinate-wise) then $\mathbf{i} \in \mathcal{F}_s \Rightarrow \mathbf{j} \in \mathcal{F}_s$. We call $\mathcal{F}_s$ the *support Ferrer diagram* of the system χ_s.

THEOREM 10.10. *(See [277].) Let $I \lhd R[X]$ be an ideal, $\preccurlyeq$ be a monomial ordering on $\mathbb{N}_0^k$. Then for some $t \in \{0, \ldots, n-1\}$ there exists a chain*

$$\mathcal{F}_t \subset \mathcal{F}_{t-1} \subset \cdots \subset \mathcal{F}_0 \subset \mathbb{N}_0^k \tag{10.27}$$

of Ferrer diagrams, strictly ordered by inclusion, and a series

$$0 \leqslant \|I\| = a_0 < a_1 < \cdots < a_t < a_{t+1} = n \tag{10.28}$$

of integers satisfying the following conditions:

(C1) *for all $F \in \mathcal{R}_k$ and $s \in \{0, \ldots, t\}$ $(F \in I, \operatorname{supp}(F) \subseteq \mathcal{F}_s) \Rightarrow (\|F\| \geqslant a_{s+1})$;*

(C2) *for each $s \in \{0, \ldots, t\}$ there exists a π-homogeneous system $\chi_s \subset R[X]$ of polynomials that is reduced relative to $\preccurlyeq$ such that $\|\chi_s\| = 0$, $\pi^{a_s}\chi_s \subset I$.*

Any system of polynomials

$$\chi_0, \chi_1, \ldots, \chi_t, \tag{10.29}$$

satisfying conditions (C1), (C2), *has also the following properties.*

(C3) *If $F \in I$ and $\|F\| = a$, where $a \geqslant a_s$, for some $s \in \{0, \ldots, t\}$, then*

$$F \in \mathcal{R}_k\big(\pi^a\chi_s \cup \pi^{a_{s+1}}\chi_{s+1} \cup \cdots \cup \pi^{a_t}\chi_t\big),$$

in particular,

$$I = \mathcal{R}_k\big(\pi^{a_0}\chi_0 \cup \pi^{a_1}\chi_1 \cup \cdots \cup \pi^{a_t}\chi_t\big).$$

(C4) *If $s \in \{0, \ldots, t\}$ and $a_s \leqslant a < a_{s+1}$, then*

$$I \cap \pi^a\mathcal{R}_k = \mathcal{R}_k\big(\pi^a\chi_s \cup \pi^{a_{s+1}}\chi_{s+1} \cup \cdots \cup \pi^{a_t}\chi_t\big), \tag{10.30}$$

$$\big(I : \pi^a\big) = \mathcal{R}_k\big(\chi_s \cup \pi^{a_{s+1}-a}\chi_{s+1} \cup \ldots, \pi^{a_t-a}\chi_t \cup \big\{\pi^{n-a}\big\}\big). \tag{10.31}$$

(C5) *If* $H(\mathbf{x}) = H_0(\mathbf{x}) \in \mathcal{R}_k$ *and* $H_1(\mathbf{x}), \dots, H_{t+1}(\mathbf{x})$ *are polynomials, constructed recursively by the rule*:

$$H_{s+1} \in \mathrm{NF}(H_s, \chi_s), \quad s \in \{0, \dots, t\},$$

then the inclusion $H \in I$ *is equivalent to the system of conditions*

$$\|H_s(x)\| \geqslant a_s, \quad s \in \{0, \dots, t+1\}. \tag{10.32}$$

(C6) *The set* $\chi = \pi^{a_0}\chi_0 \cup \pi^{a_1}\chi_1 \cup \cdots \cup \pi^{a_t}\chi_t$ *is a coordinate standard basis of the ideal* I *with respect to the ordering* $\preccurlyeq$, *i.e. for any polynomial* $F \in \mathcal{R}_k$ *the conditions* $F \in I$ *and* $F \xrightarrow{\chi} 0$ *are equivalent.*

(C7) *The ideal* I *is monic iff* $a_0 = 0$ *and the Ferrer diagram* $\mathcal{F}_0$ *is finite* (*i.e all of the diagrams* $\mathcal{F}_s$ *are finite*). *Suppose* I *is a monic ideal,* R *is a finite ring,* $\overline{R} = GF(q)$ *and* $|\mathcal{F}_s| = m_s$ *for each* $s \in \{0, \dots, t\}$. *Then* $S = \mathcal{R}_k/I$ *is a finite ring and*

$$|S| = q^{(m_0 - m_1)a_1 + \cdots + (m_{t-1} - m_t)a_t + m_t n}. \tag{10.33}$$

Now we can give the following definition. A subset χ of an ideal $I \lhd \mathcal{R}_k$ is called *canonical generating system* (*CGS*) *of* I (*corresponding to an admissible order* $\preccurlyeq$) if it is a union (10.26) of π-homogeneous subsets $\pi^{a_0}\chi_0, \dots, \pi^{a_t}\chi_t$ of the ideal I such that:

(i) $0 \leqslant a_0 < a_1 < \cdots < a_t < n = a_{t+1}$;

(ii) For every $s \in \{0, \dots, t\}$ the system χ_s is reduced (corresponding to $\preccurlyeq$) with support Ferrer diagram $\mathcal{F}_s$;

(iii) $\mathcal{F}_0 \supsetneqq \cdots \supsetneqq \mathcal{F}_t$;

(iv) $\forall F \in I, s \in \{0, \dots, t\}$: $(\mathrm{supp}(F) \subseteq \mathcal{F}_s) \Rightarrow (\|F\| \geqslant a_{s+1})$.

This definition is natural generalization of the definition of CGS of an ideal $I \in \mathcal{R}_1$ (Section 10.1.1). So Theorem 10.10 states that any non-zero ideal $I \lhd \mathcal{R}_k$ contains a CGS corresponding to any admissible order $\preccurlyeq$. Any CGS of I is its CSB of it [158]. In [158] an algorithm for the construction of a CGS of a non-zero ideal I, given a finite set of generators of I, was presented.

10.2.3. *Systems of polynomial equations* Consider a system of polynomial equations

$$\{F_i(x_1, \dots, x_k) = 0,\ i \in \{1, \dots, d\}\} \tag{10.34}$$

over a GE-ring R. Let us consider the vector-function $\varPhi(\mathbf{x}) = (F_1(\mathbf{x}), \dots, F_d(\mathbf{x}))^T \in \mathcal{R}_k^{(d)}$. Then the system (10.34) can be written in the following short form

$$\varPhi(\mathbf{x}) = 0. \tag{10.35}$$

Below it is assumed that $\overline{\varPhi}(x) \neq \bar{0}$. The simplest method of the description of all solutions of this system is:

METHOD OF COORDINATEWISE LINEARIZATION. For any vector $\mathbf{c} = (c_1, \dots, c_k) \in R^k$ and for every $j \in \{0, \dots, n-1\}$ vector $\gamma_j(\mathbf{c}) = (\gamma_j(c_1), \dots, \gamma_j(c_k)) \in \varGamma(R)^k$ is call the j-th *coordinate-vector* of $\mathbf{c}$. Then $\mathbf{c} = \gamma_0(\mathbf{c}) + \pi\gamma_1(\mathbf{c}) + \cdots + \pi^{n-1}\gamma_{n-1}(\mathbf{c})$. Let us put

$$\mathbf{c}^{[j]} = \gamma_0(\mathbf{c}) + \pi\gamma_1(\mathbf{c}) + \cdots + \pi^{j-1}\gamma_{j-1}(\mathbf{c}) \in \varGamma^k + \pi\varGamma^k + \cdots + \pi^{j-1}\varGamma^k,$$
$$j \in \{1, \dots, n\},$$

$D\Phi(\mathbf{x}) = (D_s F_i(\mathbf{x}))_{d\times k}$, where $D_s F_i(\mathbf{x})$ is the partial derivative of $F_i(\mathbf{x})$ with respect to x_s.

THEOREM 10.11. *A vector* $\mathbf{c} \in R^k$ *is a solution of the system of Eq.* (10.34) *iff the vector* $\gamma_0(\mathbf{c})$ *is a solution in* Γ^k *of the system of polynomial equations*

$$\Phi(\mathbf{z}) \equiv 0 \pmod{\pi R}, \tag{10.36}$$

and for $j \in \{1, \dots, n-1\}$ *the vector* $\gamma_j(\mathbf{c})$ *is a solution in* Γ^k *of the system of linear equations*

$$D\Phi\big(\gamma_0(\mathbf{c})\big) \cdot \mathbf{z} \equiv -\gamma_j\big(\Phi(\mathbf{c}^{[j]})\big) \quad \big(\mathrm{mod}\, \pi^j R\big). \tag{10.37}$$

It is important to note that (10.36) and (10.37) are in fact systems of equations over the field $\Gamma = GF(q)$. In particular, if the system (10.37) is solvable then it has q^{k-r} solutions, where $r = \operatorname{rank} \overline{D}\Phi(\gamma_0(\mathbf{c}))$.

COROLLARY 10.1. *A system* (10.34) *of* $d = k$ *polynomial equations has unique solution over the ring* R *iff the system* (10.36) *has unique solution* $\mathbf{c}_0$ *over the field* Γ, *and the equality* $\operatorname{rank} \overline{D}\Phi(\mathbf{c}_0) = k$ *holds.*

Another way to obtain solutions of the system (10.34) is given by the following (related) approach.

METHOD OF SOLVING EQUATIONS USING CANONICAL GENERATING SYSTEMS. Let $\preccurlyeq$ be some admissible ordering on $\mathbb{N}_0^k$ and χ be the CGS of the ideal $I = \mathcal{R}_k\{F_1, \dots, F_d\} \lhd \mathcal{R}_k$ generated by the polynomials from the left part of the system (10.34) corresponding to the ordering $\preccurlyeq$. Then the original system of equations is equivalent to the system

$$\chi(\mathbf{x}) = 0. \tag{10.38}$$

As before we suppose that $\overline{\Phi}(x) \neq \bar{0}$. That means that $\overline{\chi}(x) \neq \bar{0}$, i.e. the set (10.26) satisfies the condition $a_0 = 0$, and has the form

$$\chi = \chi_0 \cup \pi^{a_1}\chi_1 \cup \dots \cup \pi^{a_t}\chi_t.$$

In order to simplify the notations we consider a somewhat redundant system of generators of the same ideal I:

$$\Psi = \Psi_0 \cup \pi\Psi_1 \cup \dots \cup \pi^{n-1}\Psi_{n-1}, \quad \text{where } \Psi_j = \chi_i, \text{ if } a_i \leqslant j < a_{i+1}.$$

The system of equations (10.34) is equivalent to the system $\Psi(\mathbf{x}) = 0$, i.e. to the system

$$\Psi_0(\mathbf{x}) = 0, \qquad \pi\Psi_1(\mathbf{x}) = 0, \qquad \dots, \qquad \pi^{n-1}\Psi_{n-1}(\mathbf{x}) = 0. \tag{10.39}$$

THEOREM 10.12. *A vector* $\mathbf{c} \in R^k$ *is a solution of the system* (10.34) (*of the system* (10.39)) *iff*

(a) *the vector* $\gamma_0(\mathbf{c}) \in \Gamma^k$ *is a solution of the system of equations*

$$\Psi_{n-1}(\mathbf{z}) \equiv 0 \pmod{\pi R}, \tag{10.40}$$

(b) *for every* $j \in \{1, \ldots, n-1\}$ *the coordinate-vector* $\gamma_j(\mathbf{c}) \in \Gamma^k$ *is a solution of the system of linear equations*

$$D\Psi_{n-j-1}\big(\gamma_0(\mathbf{c})\big) \cdot \mathbf{z} \equiv -\gamma_j\big(\Psi_{n-j-1}\big(\mathbf{c}^{[j]}\big)\big). \tag{10.41}$$

10.2.4. *Some other applications of coordinated standard bases* Coordinated standard bases allow one to solve many computational problems in the polynomial ring $R[X]$.

IDEAL MEMBERSHIP PROBLEM. Given a finite set of polynomials $\psi \subset R[X]$ and a polynomial F, determine whether $F \in (\psi)$. This problem can be effectively solved using CSB's. Let χ be a CSB of the ideal (ψ) then

$$F \in (\psi) \quad \Leftrightarrow \quad F* \xrightarrow{\chi} 0.$$

A SET OF COSET REPRESENTATIVES. Given an ideal $I \lhd R[X]$, find a set of polynomials $C \subset R[X]$ such that for any $F \in R[X]$ there exists the only one polynomial $G \in C$: $F \equiv G \bmod (I)$. Let Γ be the coordinate field of R (Section 2.4). Take a CSB χ of the ideal I. Consider the set of terms

$$T = \big\{\pi^a u \;\big|\; a \in \{1, \ldots, n\},\ u \in [X\rangle,\ \text{there is no } G \in \chi \\ \text{such that } Lt(G) \text{ divides } \pi^a u\big\}.$$

THEOREM 10.13. *(See [159].) For any ideal* $I \lhd R[X]$ *the set of polynomials of the form*

$$\sum_{U \in T} a_U U$$

(*where only a finite number of the coefficients* $a_U \in \Gamma$ *is not equal* 0) *is a set of coset representatives for* $R[X]/I$.

THEOREM 10.14. *Let* I *be a monic ideal in* $R[X]$, *then the factor-ring* $R[X]/I$ *is finite and*

$$\big|R[X]/I\big| = q^{|T|},$$

where q *is the cardinality of the residue field* $\overline{R}$.

If in addition $\chi = \chi_0 \cup \cdots \cup \chi_t$ *is a CGS of the ideal* I *then*

$$|T| = (m_0 - m_1)a_1 + \cdots + (m_{t-1} - m_t)a_t + m_t n,$$

where $m_s = |\mathcal{F}_s|$ *and* $a_s = \|\chi_s\|$ *for* $s \in \{0, \ldots, t\}$ (*compare with* (10.15)).

The following theorem shows that the problem of the evaluation of the parameters $m_s = |\mathcal{F}_s| = |[X\rangle \setminus Lm(\chi_s)[X\rangle|$, $s \in \overline{0, t}$ is rather difficult.

THEOREM 10.15. *(See [37].) The following problem is NP-complete*:

Given a set of monomials L *and* $m \in \mathbb{N}$, *is* $|[X\rangle \setminus L[X\rangle| \leqslant m$*?*

Syzygy module generators. The *syzygy module* of a polynomial system $\psi = \{F_1, \dots, F_l\} \subset R[X]$ is defined to be

$$\operatorname{Syz}(F_1, \dots, F_l) = \bigl\{(H_1, \dots, H_l) \in R[X]^l \bigm| F_1H_1 + \dots + F_lH_l = 0\bigr\}. \tag{10.42}$$

The problem is to construct a set of generators of $\operatorname{Syz}(F_1, \dots, F_l)$.

Let $\chi = \{G_1, \dots, G_m\}$ be a CSB of the ideal (ψ). According to the Theorem 10.6, for any $i, j \in \{1, \dots, m\}$, $i \neq j$, the S-polynomial $S(G_i, G_j) = V_{ij}G_i - U_{ij}G_j$ possesses an H-presentation:

$$V_{ij}G_i - U_{ij}G_j = \sum_{\alpha=1}^{m}\left(\sum_{k=1}^{m_\alpha} a^k_{ij\alpha}u^k_{ij\alpha}\right)G_\alpha,$$

where $a^k_{ij\alpha} \in R$, $u^k_{ij\alpha} \in [X\rangle$ and $Lt(a^k_{ij\alpha}u^k_{ij\alpha}G_\alpha) \preccurlyeq Lt(S(G_i, G_j))$. Note that this representation can be effectively constructed using the reduction process. The vector

$$\mathbf{s}_{ij} = V_{ij}\mathbf{e}_i - U_{ij}\mathbf{e}_j - \sum_{\alpha=1}^{m}\left(\sum_{k=1}^{m_\alpha} a^k_{ij\alpha}u^k_{ij\alpha}\right)\mathbf{e}_\alpha \in R[X]^m \tag{10.43}$$

is called an *S-syzygy*. Here $\mathbf{e}_\alpha = (0, \dots, 1, \dots, 0) \in R[X]^m$ (the 1 occurs in the α-th location).

The vector

$$\mathbf{p}_k = \pi^{n-\|G_k\|}\mathbf{e}_k \in R[X]^m \tag{10.44}$$

is called a π-syzygy.

Theorem 10.16. *(See* [159].*) Let $G_1, \dots, G_m \in R[X]$ be a CSB of an ideal I. Then the syzygy module* $\operatorname{Syz}(G_1, \dots, G_m)$ *is generated by the system of vectors* $\{\mathbf{s}_{ij} \mid i, j \in \{1, \dots, m\},\ i \neq j\} \cup \{\mathbf{p}_k \mid k \in \{1, \dots, m\}\}$.

There exist matrices $T_{l\times m}$ and $S_{m\times l}$ with entries in $R[X]$ such that

$$(F_1, \dots, F_l) = (G_1, \dots, G_m)S \quad \text{and} \quad (G_1, \dots, G_m) = (F_1, \dots, F_l)T.$$

Theorem 10.17. *(See* [1].*) The syzygy module* $\operatorname{Syz}(F_1, \dots, F_l)$ *is generated by the columns of the matrix $E_l - TS$ and the vectors $T\mathbf{s}^t_{ij}$, $T\mathbf{p}^t_k$, $i, j, k \in \{1, \dots, m\}$, $i \neq j$, where E_l is the $l \times l$ identity matrix, and the vectors $\mathbf{s}_{ij}$ and $\mathbf{p}_k$ are S-syzygies and π-syzygies for $G_1, \dots, G_m$.*

Note that the matrix S can be obtained using the reduction of polynomials F_i modulo χ. The matrix T can be constructed during the evaluation of the standard basis χ. With that end in view, it is necessary to keep track of the reductions at every step of Algorithm 10.7.

11. Periods of polynomials and polynomial ideals over a commutative f.r. [226,227,296]

Here we suppose that R is a commutative f.r. with identity e.

11.1. *Periodic ideals of 1-variable polynomials, properties and parameters*

An ideal $I \lhd \mathcal{R}$ (respectively polynomial $F(x) \in \mathcal{R}$) is called *periodic*, if there exist numbers $d \in \mathbb{N}_0$, $t \in \mathbb{N}$ such that

$$x^d\bigl(x^t - e\bigr) \in I \quad \bigl(\text{respectively } F(x) \mid x^d\bigl(x^t - e\bigr)\bigr). \tag{11.1}$$

LEMMA 11.1. *For a periodic ideal I there exist parameters $D(I) \in \mathbb{N}_0$, $T(I) \in \mathbb{N}$ such that*

$$\forall d \in \mathbb{N}_0,\ t \in \mathbb{N}\colon\ (11.1) \quad \Longleftrightarrow \quad d \geqslant D(I)\ \&\ T(I)|t.$$

We call $D(I)$ the *defect* and $T(I)$ the *period* of the ideal I. For a periodic polynomial $F(x)$ we use correspondingly notations $D(F) = D(F(x)\mathcal{R})$, $T(F) = T(F(x)\mathcal{R})$.

An ideal $I \lhd \mathcal{R}$ is called *monic* if it contains a monic polynomial.

THEOREM 11.2. *For any ideal $I \lhd \mathcal{R}$ the following conditions are equivalent*:

(a) *$S = \mathcal{R}/I$ is a finite ring*;

(b) *I is a periodic ideal*;

(c) *I is a monic ideal.*

Under these conditions there are relations: $D(I) + T(I) \leqslant |S|$, *and if* $|S| > 2$, $D(I) + T(I) \leqslant |S| - 1$.

If $|S| > 2$, *then the equality* $D(I) + T(I) = |S| - 1$ *holds if and only if*

- *either* $S = GF(2)[x]/(x^2)$ (*and then* $D(I) + T(I) = 3$),
- *or I has a form $I = (F(x), \pi)$, where $F(x) \in \mathcal{R}$ is a monic polynomial of degree m and image $\overline{F}(x)$ under the canonical epimorphism $R \to \overline{R} = R/\pi R = GF(q)$ satisfying the condition $T(\overline{F}) = q^m - 1$. In the last case $S = GF(q^m)$, $D(I) = 0$, $T(I) = q^m - 1$.*

Let us call a periodic ideal I (respectively polynomial $F(x)$) *reversible*, if $D(I) = 0$ (respectively $D(F) = 0$), and *degenerate*, if $x^{D(I)} \in I$ (respectively $F(x) \mid x^{D(F)}$).

PROPOSITION 11.3. *Any monic polynomial $F(x) \in \mathcal{R}$ is a periodic one. If* $\deg F(x) = m$, $|R|^m > 2$, *then $D(F) + T(F) \leqslant |R|^m - 1$. The polynomial $F(x)$ is reversible if and only if $F(0) \in R^*$.*

PROPOSITION 11.4. *If I_1, I_2 are periodic ideals of the ring $\mathcal{R}$, then the ideal $I = I_1 \cap I_2$ is also periodic and*

$$D(I) = \max\bigl\{D(I_1), D(I_2)\bigr\}, \qquad T(I) = \bigl[T(I_1), T(I_2)\bigr].$$

If $F_1(x)$, $F_2(x) \in \mathcal{R}$ are coprime monic polynomials, then

$$D(F_1F_2) = \max\bigl\{D(F_1), D(F_2)\bigr\}, \qquad T(F_1F_2) = \bigl[T(F_1), T(F_2)\bigr].$$

PROPOSITION 11.5. *Any monic ideal $I \lhd \mathcal{R}$ has a unique representation in the form $I = I^{(r)} \cap I^{(d)}$, where $I^{(r)}$ is a reversible and $I^{(d)}$ is a degenerate ideal. These ideals are comaximal and also $I = I^{(r)}I^{(d)}$. There are the equalities*: $T(I) = T(I^{(r)})$, $D(I) = D(I^{(d)})$.

The calculation of parameters of a periodic ideal can be reduced to the case when R is a local ring. Indeed, a commutative f.r. R with identity e is a direct sum of local rings:

$$R = R_1 \dotplus \cdots \dotplus R_t \tag{11.2}$$

(see Theorem 1.7). Let $e = e_1 + \cdots + e_t$ be the corresponding decomposition of the identity of the ring R. Then e_s is the identity of R_s and $R_s = Re_s$. The decomposition (11.2) implies decompositions of any polynomial $F(x) \in \mathcal{R}$ and any ideal $I \lhd \mathcal{R}$:

$$I = I_1 \dotplus \cdots \dotplus I_t, \quad \text{where } I_s = e_s I. \tag{11.3}$$

$$F(x) = F_1(x) \dotplus \cdots \dotplus F_t(x), \quad \text{where } F_s(x) = e_s F(x) \in R_s[x] = e_s\mathcal{R}; \tag{11.4}$$

Proposition 11.6. *Under conditions (11.2)–(11.4) an ideal I is periodic if and only if everyone of its component I_s is periodic. If I is a periodic ideal, then*

$$T(I) = \big[T(I_1), \ldots, T(I_t)\big], \qquad D(I) = \max\big\{D(I_1), \ldots, D(I_t)\big\}.$$

For any monic polynomial $F(x)$

$$T(F) = \big[T(F_1), \ldots, T(F_t)\big], \qquad D(F) = \max\big\{D(F_1), \ldots, D(F_t)\big\}.$$

11.2. *Parameters of periodic ideals over a commutative local f.r.*

Below R is a titled ring, $\overline{R} = R/\mathfrak{N}(R) = GF(q)$, $q = p^r$, p is a prime, $p^d = \operatorname{char} R$ is the characteristic of the ring R, $n = \operatorname{ind}\mathfrak{N}(R)$ is the nilpotency index of the ring R.

11.2.1. *General estimates* Let $I \lhd \mathcal{R}$ be a monic (i.e. periodic) ideal. We call a monic polynomial $F(x) \in I$ of least degree a *main generator of the ideal I*. This definition is equivalent to the equality

$$I = F(x)\mathcal{R} + \mathfrak{N}(I), \quad \text{where } \mathfrak{N}(I) = I \cap \mathfrak{N}[x]. \tag{11.5}$$

Proposition 11.7. *Every monic polynomial $F(x)$ over a commutative local f.r. R has a unique decomposition into a product $F(x) = F^{(\mathrm{r})}(x)F^{(\mathrm{d})}(x)$ of a reversible polynomial $F^{(\mathrm{r})}(x)$ and a degenerating polynomial $F^{(\mathrm{d})}(x)$. Moreover, $T(F) = T(F^{(\mathrm{r})})$, $D(F) = D(F^{(\mathrm{d})})$.*

If F is a main generator of the ideal I, then

$$I^{(\mathrm{r})} = F^{(\mathrm{r})}(x)\mathcal{R} + \mathfrak{N}(I), \qquad I^{(\mathrm{d})} = F^{(\mathrm{d})}(x)\mathcal{R} + \mathfrak{N}(I).$$

Note, that in general, if R is not a local ring, the last proposition is not true. For example the polynomial $F(x) = x^2 - 4x - 3$ over the ring $\mathbb{Z}/(6)$ cannot be represented in the form $F = F^{(\mathrm{r})}F^{(\mathrm{d})}$.

As before we denote by $\overline{F}$ the image of a polynomial $F \in \mathcal{R}$ under the natural epimorphism $R \to \overline{R}$.

PROPOSITION 11.8. *The period and defect of a monic ideal I with a main generator $F(x)$ satisfy the relations*

$$D(\overline{F}) \leqslant D(I) \leqslant D(F) \leqslant nD(\overline{F}) \leqslant n \cdot \deg F(x), \tag{11.6}$$

$$T(\overline{F}) \mid T(I), \qquad T(I) \mid T(F). \tag{11.7}$$

If I is a reversible ideal, then

$$T(I) = T(\overline{F})p^{\alpha(I)}, \quad \text{where } \alpha(I) \leqslant d. \tag{11.8}$$

So the calculation of the period of a reversible ideal I reduced to the calculation of the period $T(\overline{F})$ of the polynomial $\overline{F}(x)$ over the field $\overline{R}$ and to the calculation of the parameter $\alpha(I)$, the minimal $\alpha \in \mathbb{N}_0$ with the property

$$\operatorname{Res}\bigl(x^{T(\overline{F})p^{\alpha}} - e \,/\, F(x)\bigr) \in I.$$

Below there is offered a more convenient approach to the calculation of the parameter $\alpha(I)$, unconnected with the calculation of the parameter $T(\overline{F})$. This approach is based on the following notion.

11.2.2. *Calculation of the period of a reversible ideal via distinguished polynomials* We call reversible polynomial $D(x) \in \mathcal{R}$ *distinguished*, if $T(D) = T(\overline{D})$. Say that $D(x)$ is a distinguished polynomial corresponding to polynomial $G(x) \in \mathcal{R}$ (or to the polynomial $g(x) \in \overline{R}[x]$), if $\overline{D}(x) = \overline{G}(x)$ (respectively $\overline{D}(x) = g(x)$). A polynomial $G(x) \in \mathcal{R}$ is called *separable*, if $(G(x), G(x)')_{\mathcal{R}} = (e)_{\mathcal{R}}$ (i. e. $(\overline{G}(x), \overline{G}(x)') = \bar{e}$), where $G(x)'$ is the derivative of $G(x)$.

Recall that the ring R contains a unique coefficient ring $K = GR(q^d, p^d)$, $p^d = \operatorname{char} R$, of the form $K = \Gamma(R) + \Gamma(R)p + \cdots + \Gamma(R)p^{d-1}$ (see (5.3)).

PROPOSITION 11.9. *To every reversible separable polynomial $G(x) \in \mathcal{R}$ corresponds a unique distinguished polynomial $G_*(x) \in \mathcal{R}$, this polynomial belongs to $K[x]$. A product of co-prime distinguished polynomials is a distinguished polynomial.*

Here we present a simple way for the construction of a polynomial $G_*(x)$ as in the last proposition. First of all note that in order to construct a distinguished polynomial corresponding to G we can choose any polynomial $G^{[0]}(x) \in K[x]$ with the property $\overline{G^{[0]}} = \overline{G}$ and then according to Proposition 11.9 $G_* = G_*^{[0]}$. So it is enough to construct a distinguished polynomial for the separable reversible polynomial $G(x) \in K[x]$.

There exists an extension $S = K[\xi]$ of the ring K such that

$$x^p - e = (x - e)(x - \xi) \cdots \bigl(x - \xi^{p-1}\bigr). \tag{11.9}$$

For example if $p = 2$, then $\xi = -e$ and $K[\xi] = K$. In general we can find an S in the form $S = K[x]/(E(x))$, where

$$E(x) = x^{p-1} + px^{p-2} + \binom{p}{2}x^{p-3} + \cdots + \binom{p}{2}x + p.$$

LEMMA 11.10. *For every separable reversible polynomial $G(x) \in K[x]$ of degree m the following equality over the ring S holds*

$$(-1)^{m(p-1)} \cdot \prod_{i=0}^{p-1} G\big(\xi^i x\big) = G^{[1]}\big(x^p\big), \tag{11.10}$$

where $G^{[1]}(x) \in K[x]$ is a monic polynomial with the property $\overline{G}^{[1]}(x^p) = \overline{G}(x)^p$.

PROPOSITION 11.11. *Let $G(x) \in K[x]$ be a separable reversible polynomial*, $\deg G = m$, *and let $G^{[0]}(x), G^{[1]}(x), \dots$ be a sequence of polynomials over K defined by the rule*

$$G^{[0]}(x) = G(x), \qquad G^{[k+1]}\big(x^p\big) = (-1)^{m(p-1)} \prod_{i=0}^{p-1} G^{[k]}\big(\xi^i x\big), \quad k \in \mathbb{N}_0. \tag{11.11}$$

Then the polynomial

$$G_*(x) = G^{[\varkappa r]}(x), \quad \text{where } \varkappa = \big](d-1)/r\big[, \tag{11.12}$$

is a distinguished polynomial corresponding to $G(x)$. The polynomial $G(x)$ is distinguished exactly if

$$G^{[r]}(x) = G(x). \tag{11.13}$$

In the important particular case $p = 2$ formulas (11.11) have essential simplifications. Any polynomial $G(x) \in K[x]$ has a unique representation in the form

$$G(x) = G_{(0)}\big(x^2\big) + xG_{(1)}\big(x^2\big).$$

PROPOSITION 11.12. *If under the conditions of Proposition 11.11 the equality $p = 2$ holds then the series of polynomials (11.11) can be constructed by the rule*

$$G^{[0]}(x) = G(x), \qquad G^{[k+1]}(x) = (-1)^m \big(G_{(0)}^{[k]}(x)^2 - xG_{(1)}^{[k]}(x)^2\big). \tag{11.14}$$

We define the *radical* of a reversible polynomial $F(x) \in \mathcal{P}$ as the distinguished polynomial $\operatorname{rad} F(x) \in \mathcal{P}$, which corresponds to the product of all different monic divisors of the polynomial $\overline{F}(x) \in \overline{\mathcal{P}}$ that are irreducible over $\overline{R}$. Note that $\operatorname{rad} F(x)$ is a separable polynomial and in order to construct $\operatorname{rad} F(x)$ it is necessary only to know $\operatorname{rad} \overline{F}(x)$. The last one can be found without the decomposition of $\overline{F}(x)$ into a product of irreducible polynomials. It is enough to use the operations of differentiation, calculation of gcd's and arithmetical operations on polynomials.

THEOREM 11.13. *Let I be a reversible ideal with a main generator $F(x)$, let $G(x) =$* $\operatorname{rad} F(x)$, *let k be the maximum of the multiplicities of irreducible divisors of $\overline{F}(x)$ over $\overline{R}$, and let $p^{a-1} < k \leqslant p^a$. Then*

$$T(I) = T(\overline{G})p^{a+\alpha(I)} = T(\overline{G})p^{\beta(I)}, \tag{11.15}$$

where $\alpha(I)$ is defined by (11.8) and $\beta(I)$ is the minimal $b \in \mathbb{N}_0$ with the property

$$G^{[b]}\big(x^{p^b}\big) \in I. \tag{11.16}$$

11.2.3. *Periods of reversible polynomials and ideals over a Galois ring* Let $R = K = GR(q^n, p^n) = GR(p^{rn}, p^n)$ be a Galois ring. Then, using the canonical generating system (10.14) of the ideal I:

$$I = \bigl(F(x) = F_0(x), p^{a_1}F_1(x), \dots, p^{a_t}F_t(x)\bigr),$$

we can deduce formulae for $\alpha(I)$ in (11.15).

Let $G(x) = \operatorname{rad} F_0(x)$ and let the parameters, k, a be as in Theorem 11.13. For any $b \in \mathbb{N}_0$, we suppose that

$$U_b(x) = \operatorname{Res}\bigl(G^{[b]}\bigl(x^{p^b}\bigr)/F(x)\bigr) \tag{11.17}$$

has the following decomposition in the system of radices $F_1(x), \dots, F_t(x)$:

$$U_b(x) = U_{b1}(x)F_1(x) + \cdots + U_{bt}(x)F_t(x) + U_{b,t+1}(x), \tag{11.18}$$

where $\deg U_{bi}(x)F_i(x) < \deg F_{i-1}(x)$, $i \in \{1, \dots, t+1\}$. Let

$$n_{bi} = \bigl\|U_{bi}(x)\bigr\|, \quad i \in \{1, \dots, t+1\}; \qquad n_b = \min\{n_{b1}, \dots, n_{b,t+1}\};$$
$$d_b(I) = \max\{a_1 - n_{b1}, \dots, a_t - n_{b,t}, n - n_{b,t+1}\}. \tag{11.19}$$

THEOREM 11.14. *(See [226].) Under the assumptions above, $T(I) = T(\overline{F})p^{\alpha(I)}$, where*

$$\alpha(I) = \begin{cases} d_a(I), & \text{if } p^{n_a} > 2, \text{ or } p^{n_a} = 2,\ d_a(I) \leqslant 1; \\ d_{a+1}(I) + 1, & \text{if } p^{n_a} = 2,\ d_a(I) > 1. \end{cases}$$

COROLLARY 11.1. *Let $F(x)$ be a reversible polynomial of degree m over the ring $R = GR(q^n, p^n)$, $n > 1$, let $G(x) = \operatorname{rad} F(x)$ and let k, a be as above. Then*

(a) *for some $n_a \in \{1, \dots, n\}$*

$$\operatorname{Res}\bigl(G^{[a]}\bigl(x^{p^a}\bigr)/F(x)\bigr) = p^{n_a}V_a(x), \quad \overline{V}_a(x) \neq \bar{0}, \tag{11.20}$$

and if $p^{n_a} > 2$, or if $p^{n_a} = 2, n = 2$, or if $p^{n_a} = 2, n > 2$ and

$$\overline{V}_a(x)\bigl(\overline{V}_a(x) + \bigl(x\overline{G}(x)'\bigr)^{2^a}\bigr) \not\equiv \bar{0} \pmod{\overline{F}(x)}, \tag{11.21}$$

then

$$T(F) = T(\overline{F})p^{n-n_a}; \tag{11.22}$$

(b) *if $p^{n_a} = 2, n > 2$, and (11.21) does not hold, then for some $n_{a+1} \in \{3, \dots, n\}$*

$$\operatorname{Res}\bigl(G^{[a+1]}\bigl(x^{2^{a+1}}\bigr)/F\bigr) = 2^{n_{a+1}}V_{a+1}(x), \quad \overline{V}_{a+1}(x) \neq \bar{0}, \tag{11.23}$$

and

$$T(F) = T(\overline{F})p^{n-n_{a+1}+1} < T(\overline{F})p^{n-n_a}. \tag{11.24}$$

Each of the polynomials considered satisfies the inequality

$$T(F) \leqslant \bigl(q^m - 1\bigr)p^{n-1}. \tag{11.25}$$

11.3. *Polynomials of maximal period over a Galois ring [226]*

In view of (11.25) we can call a reversible polynomial $F(x)$ of degree m over a Galois ring $R = GR(q^n, p^n)$ a *polynomial of maximal period* (MP-polynomial) if

$$T(F) = \big(q^m - 1\big)p^{n-1}. \tag{11.26}$$

THEOREM 11.15. *A reversible polynomial $F(x)$ of degree m over the ring $R = GR(q^n, p^n)$ is a polynomial of maximal period if and only if the following conditions hold*:

(a) *$\overline{F}(x)$ is a polynomial of maximal period over the field $\overline{R}$*;

(b) *if $F_*(x)$ is the distinguished polynomial corresponding to $F(x)$ (Section 11.2.2), then*

$$\begin{aligned} &F(x) = F_*(x) + pV(x), \quad \textit{where} &(11.27)\\ &\overline{V}(x) \neq \bar{0}, \quad \textit{and} \quad \textit{if } p = 2 < n, \textit{ then, in addition,} \\ &\overline{V}(x) \not\equiv x\overline{F}(x)' \quad \big(\mathrm{mod}\ \overline{F}(x)\big). &(11.28)\end{aligned}$$

COROLLARY 11.2. *An MP-polynomial of degree m over the ring $R = GR(q^n, p^n)$ exists iff $q^m > 2$ or $q^m = 2 = n$. Under these conditions for any MP-polynomial $f(x) \in \overline{\mathcal{R}}$ of degree m there exists an MP-polynomial $F(x) \in \mathcal{R}$ with $\overline{F}(x) = f(x)$, and the number of such polynomials is equal to*

$$\begin{aligned} &\big(q^m - 1\big)q^{(n-2)m}, \quad \textit{if } p > 2 \textit{ or } p = 2 = n;\\ &\big(q^m - 2\big)q^{(n-2)m}, \quad \textit{if } p = 2 < n. \end{aligned}$$

COROLLARY 11.3. *Let $F(x)$ be a reversible polynomial of degree m over a Galois ring $R = GR(q^n, p^n)$, $q = p^r$, and let $F^{[r]}(x)$ be the polynomial obtained from F by the rule (11.11). Then $F(x)$ is an MP-polynomial if and only if $T(\overline{F}) = q^m - 1$ and the polynomial $W(x)$ defined from the relation $F(x) - F^{[r]}(x) = pW(x)$ satisfies the conditions*

$$\overline{W}(x) \neq \bar{0}, \quad \textit{and} \quad \textit{if } p = 2 < n, \quad \overline{W}(x) \not\equiv x\overline{F}(x)' \quad \big(\mathrm{mod}\ \overline{F}(x)\big).$$

The above results make it possible to simplify the algorithm for the construction of MP-polynomials over $\mathbb{Z}_{p^n}$ with the help of the table of MP-polynomials over $\mathbb{Z}_p$ and some combinatorial conditions on the coefficients of polynomials.

THEOREM 11.16. *Let $F(x) = x^m + a_{m-1}x^{m-1} + \cdots + a_0$ be a reversible polynomial over $\mathbb{Z}_{p^n}$, $p > 2$, such that $T(\overline{F}) = p^m - 1$. Then $F(x)$ is an MP-polynomial if and only if*

$$\sum_{j_0,\ldots,j_{m-1}\in A} \frac{p!}{j_0!\cdots j_{m-1}!}\prod_{s=0}^{m}\big(a_s x^s\big)^{j_s} \not\equiv F\big(x^p\big) \quad \big(\mathrm{mod}\ p^2\big),$$

where $a_m = e$, A is the set of all rows $(j_0, \ldots, j_m)$ of numbers from $\{0, \ldots, p-1\}$ such that

$$j_0 + j_1 + \cdots + j_m = p, \qquad j_1 + 2j_2 + \cdots + mj_m \equiv 0 \quad (\mathrm{mod}\ p).$$

In particular this implies

THEOREM 11.17. *(See [226].) Let $F(x) = x^m + a_{m-1}x^{m-1} + \cdots + a_0$ be a polynomial over $\mathbb{Z}_{p^n}$, such that $T(\overline{F}) = p^m - 1$. Then $F(x)$ is an MP-polynomial in the following cases.*

1. $p > 2$ *and* $a_0^p \not\equiv a_0 \pmod{p^2}$, *or* $F(x) = x^m + a_k x^k + a_0, m \geqslant p - 2$.
2. $p = 2$ *and*:
   (a) *m is even, $a_0 \equiv e \pmod 4$; or*
   (b) *m is odd and*

$$a_1 \underset{4}{\equiv} \begin{cases} e + 2a_0a_2 & \text{if } \bar{a}_1 = \bar{e}, \\ 2(e + a_0a_2) & \text{if } \bar{a}_1 = \bar{0}; \end{cases}$$

or

   (c) $F(x) = x^m + a_k x^k + a_0, a_0, a_k \in \{-e, e\}, (m, a_0) \neq (2k, e)$.

11.4. *Periodic ideals of k-variable polynomials*

Let R be a commutative f.r. with identity e and $\mathcal{R}_k = R[x_1, \ldots, x_k] = R[\mathbf{x}]$ be the polynomial ring in k variables over R. We call an ideal $I \lhd \mathcal{R}_k$ *periodic* (*respectively reversible*) if there exist parameters $l_1, \ldots, l_k \in \mathbb{N}_0, t_1, \ldots, t_k \in \mathbb{N}$, such that

$$x_s^{l_s}\bigl(x_s^{t_s} - e\bigr) \in I \quad \text{for } s \in \{1, \ldots, k\}$$

(respectively

$$\bigl(x_s^{t_s} - e\bigr) \in I \quad \text{for } s \in \{1, \ldots, k\}).$$

A periodic ideal is said to be *degenerated* if $x^\ell \in I$ for some $\ell \in \mathbb{N}_0^k \backslash 0$.

The ring $S = \mathcal{R}_k/I = R[\vartheta_1, \ldots, \vartheta_k]$, where $\vartheta_s = x_s + I \in S$, will be called the *operator ring of the ideal I*. Denote by $\mathcal{O}(I)$ the subsemigroup $[e, \vartheta_1, \ldots, \vartheta_k\rangle$ of the semigroup $(S, \cdot)$ generated by $e, \vartheta_1, \ldots, \vartheta_k$, and call it the *orbital semigroup of the ideal I*.

PROPOSITION 11.18. *Let $I \lhd \mathcal{R}_k$. Then*

(a) *(I is a periodic ideal) $\Leftrightarrow$ (I is a monic ideal)$\Leftrightarrow$* $(|S| < \infty) \Leftrightarrow (|\mathcal{O}(I)| < \infty)$;
(b) *(I is a reversible ideal) $\Leftrightarrow$ (I contains some system of elementary reversible polynomials $F_1(x_1), \ldots, F_k(x_k)$)* $\Leftrightarrow (|S| < \infty,\ \mathcal{O}(I) < S^*)$;
(c) *(I is a degenerated ideal)* $\Leftrightarrow (|S| < \infty,\ 0 \in \mathcal{O}(I))$.

For a periodic ideal I some natural power of any element of the finite semigroup S is an idempotent. Let $\varepsilon_s = \varepsilon_s(I)$ be the unique idempotent of the semigroup $[\vartheta_s\rangle$, $s \in \{1, \ldots, k\}$, and let $\varepsilon = \varepsilon(I)$ be the product of all idempotents of the semigroup $\mathcal{O}(I)$.

PROPOSITION 11.19. *If I is a periodic ideal of $\mathcal{R}_k$, then $\varepsilon = \varepsilon_1 \cdots \varepsilon_k$ and $\varepsilon\mathcal{O}(I) = \mathcal{T}(I)$ is a subgroup of the semigroup $\mathcal{O}(I)$ with unit ε.*

We call the group $\mathcal{T}(I)$ the *cycle group* of the periodic ideal I, and we call its cardinality $T(I) = |\mathcal{T}(I)|$ the *period* of I. The parameter $D(I) = |\mathcal{O}(I)| - |\mathcal{T}(I)|$ is called the *defect* of the ideal I. If $D(I) > 0$, then the ideal I will be called *defected*.

PROPOSITION 11.20. *A periodic ideal I is reversible iff $\mathcal{T}(I) = \mathcal{O}(I)$ (i.e., $D(I) = 0$) and it is degenerated iff $\mathcal{T}(I) = 0$.*

We call a vector $\mathbf{t} \in \mathbb{N}_0^k \setminus 0$ a *vector-period* of the ideal I if

$$\exists\, \mathbf{l} \in \mathbb{N}_0^k\colon \quad \vartheta^{\mathbf{l}+\mathbf{t}} = \vartheta^{\mathbf{l}}$$

(i.e., $\mathbf{x}^{\mathbf{l}}(\mathbf{x}^{\mathbf{t}} - e) \in I$). The subgroup $\mathfrak{P}(I)$ of the group $(\mathbb{Z}^k, +)$ generated by all vector-periods of the ideal I is called its *group of periods*.

Note that if I is a periodic ideal, then each element $\varepsilon\vartheta_s$ of the group $\mathcal{T}(I)$ has finite order. For any vector $\mathbf{t} \in \mathbb{Z}^k$ define

$$\varepsilon\vartheta^{\mathbf{t}} = (\varepsilon\vartheta_1)^{t_1} \cdots (\varepsilon\vartheta_k)^{t_k}.$$

PROPOSITION 11.21. *If $I \lhd \mathcal{R}_k$ is a periodic ideal, then $\mathfrak{P}(I)$ is a subgroup of rank k of the group $(\mathbb{Z}^k, +)$, and*

$$\mathfrak{P}(I) = \{\mathbf{t} \in \mathbb{Z}^k \mid \varepsilon\vartheta^{\mathbf{t}} = \varepsilon\}; \qquad \mathcal{T}(I) \cong \mathbb{Z}^k/\mathfrak{P}(I); \qquad T(I) = [\mathbb{Z}^k\mathfrak{P}(I)].$$

12. Matrices and linear sequences over chain rings

12.1. *Similarity of matrices*

Let R be a commutative f.r. with the identity, and let $M_m(R)$ be the ring of all $m \times m$-matrices over R. We call two matrices $A, B \in M_m(R)$ *similar* and write $A \approx B$ if $B = T^{-1}AT$ for some (*transforming*) matrix $T \in M_m(R)^*$.

12.1.1. *Problems of similarity*

PROBLEM 1. For the given matrices $A, B \in M_m(R)$ to determine when the relation $A \approx B$ is true?

Of course this problem can be solved by a brute force algorithm. We call this approach *trivial*. The question of Problem 1 alludes to the matter of the existence of a nontrivial algorithm.

Let us call a system $I_1(X), \ldots, I_s(X)$ of functions on $M_m(R)$ with values in any sets *a system of invariants* if

$$\forall A, B \in M_m(R)\colon \quad A \approx B \quad \Rightarrow \quad \bigl(I_1(A) = I_1(B), \ldots, I_s(A) = I_s(B)\bigr). \tag{12.1}$$

This system is called a *full system of invariants* in a class matrices $\mathcal{M}$ if for any two matrices $A, B \in \mathcal{M}$ of the same size the inverse to the implication (12.1) is true.

We have one *trivial* full invariant: $I(A) = [A]_{\approx}$, the class of all matrices similar to A. However calculation of $I(A)$ is a hard problem.

PROBLEM 2. To find a nontrivial full system of invariants on $M_m(R)$.

In general Problems 1, 2 have no simple solutions, because they are not more simple than well-known hard "problem of a pair of matrices over a field" (also called a "wild" problem). The last one has the following formulation. For a given field P and for given matrices $A_1, B_1, A_2, B_2 \in M_k(P)$ to find a solution $X \in M_k(P)^*$ of the system of equations

$$X^{-1}A_1X = B_1, \qquad X^{-1}A_2X = B_2.$$

The hardness ("wildness") of this problem is confirmed by the following results. For any polynomial $F(x) \in \mathcal{R}$ let

$$\mathcal{V}\big(F, M_m(R)\big) = \big\{A \in M_m(R)\colon\ F(A) = 0\big\}.$$

The problem of similarity of matrices implies the problem of a pair of $m \times m$-matrices over a field $\mathbb{Z}_p$ in the classes $\mathcal{V}(x^3, M_{3m}(\mathbb{Z}_{p^n}))$, $\mathcal{V}(x^{p^2} - e, M_{4m}(R))$ [51], and in the class $\mathcal{V}(x^2 - e, M_m(R))$, when $\operatorname{char} R = 2^n (p = 2)$ and $\mathfrak{N}(R)$ is not a principal ideal [58].

Classes of similar matrices in $\mathcal{V}(x^2 - e, M_m(\mathbb{Z}_{2^n}))$ were described in [58,143].

Approaches to Problems 1, 2 in particular cases are connected with the following notions.

To every matrix A over R corresponds its *characteristic* matrix $xE - A$ over $\mathcal{R}$. We call the matrices $xE - A$ and $xE - B$ *equivalent* and write $xE - A \sim xE - B$ if $xE - B$ can be obtained from $xE - A$ by a finite series of elementary transformations. The polynomial $\chi_A(x) = |xE - A|$ is called the *characteristic polynomial* of the matrix A.

Let $R^{(m)}$ be the R-module of all columns of the length m over the ring R. Any matrix $A \in M_m(R)$ defines a structure of an $\mathcal{R}$-module on $R^{(m)}$ with multiplication of $a^{\downarrow} \in R^{(m)}$ by $F(x)$ of the form

$$F(x)a^{\downarrow} = F(A)a^{\downarrow}. \tag{12.2}$$

We denote this module ${}_{\mathcal{R}}R^{(m)}$ by $M(A)$.

THEOREM 12.1. *(See* [53].*) For any two matrices $A, B \in M_m(R)$ the following conditions are equivalent*:

(a) $A \approx B$;
(b) $(xE - A) \sim (xE - B)$;
(c) ${}_{\mathcal{R}}M(A) \cong {}_{\mathcal{R}}M(B)$.

PROBLEM 3. For a given matrix $A \in M_m(R)$ to check whether it is *reducible*, i.e. similar to some decomposed matrix:

$$\operatorname{Diag}(A_1, A_2) = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}?$$

In the theory under consideration matrix of the special form

$$S = \begin{pmatrix} 0 & 0 & \dots & 0 & f_0 \\ e & 0 & \dots & 0 & f_1 \\ 0 & e & \dots & 0 & f_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & e & f_{m-1} \end{pmatrix} \tag{12.3}$$

is important. It is called the *companion matrix of the polynomial* $F(x) = x^m - f_{m-1}x^{m-1} - \dots - f_1 x - f_0$ and denoted by $S = S(F)$. We call a matrix A *normal* if

$$\begin{aligned} &A \approx \operatorname{Diag}\bigl(S(F_1), \dots, S(F_t)\bigr) \\ &\quad \text{for some monic polynomials } F_1(x), \dots, F_t(x) \in \mathcal{R}. \end{aligned} \tag{12.4}$$

PROBLEM 4. To check when a given matrix $A \in M_m(R)$ normal?

THEOREM 12.2. *(See* [291]*.)*
(a) *A matrix $A \in M_m(R)$ is reducible if and only if the module $M(A)$ is reducible*;
(b) *the module $M(A)$ is cyclic if and only if $A \approx S(\chi_A(x))$*;
(c) *a matrix $A \in M_m(R)$ is normal if and only if the module $M(A)$ is a direct sum of cyclic submodules*.

Let $\operatorname{Ann}(A) = \{G(x) \in \mathcal{R}\colon\ G(A) = 0\}$ be the *annihilator* of A in $\mathcal{R}$. The main result of the article [263] states that if R is a principal ideal ring and $\operatorname{Ann}(A)$ is a principal ideal of $\mathcal{R}$, then A is normal. This is incorrect. Here is a simple counterexample: the matrix is

$$A = \operatorname{Diag}\left(\begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\right) \in M_4(\mathbb{Z}_4).$$

The annihilator of A is $\operatorname{Ann}(A) = (x^2)$, however A is not normal. Indeed,

$$\bar{A} = \operatorname{Diag}\left((\bar{0}), (\bar{0}), \begin{pmatrix} 0 & 0 \\ \bar{1} & 0 \end{pmatrix}\right) = \operatorname{Diag}\bigl(\bar{S}(x), \bar{S}(x), \bar{S}\bigl(x^2\bigr)\bigr).$$

Since $\operatorname{Ann}(A) = (x^2)$, if A were normal it would be similar to some matrix of the form

$$\operatorname{Diag}\bigl(S(x-2a), S(x-2b), S\bigl(x^2\bigr)\bigr) = \operatorname{Diag}\left(2a, 2b, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\right).$$

Testing things by exhausting all a, b shows that last statement is false.

12.1.2. *Decompositions of a matrix* Problem 1 can be further reduced to the same problem for matrices over a local f.r. with primary characteristic polynomial.

Indeed, using the decomposition (11.2) of the ring R into a direct sum of local rings, we obtain a decomposition

$$M_m(R) = M_m(R_1) \dot{+} \dots \dot{+} M_m(R_t), \qquad M_m(R_s) = e_s M_m(R), \tag{12.5}$$

where e_s is the identity of the ring R_s, and corresponding decompositions of matrices:

$$A = A_1 \dot{+} \dots \dot{+} A_t, \qquad B = B_1 \dot{+} \dots \dot{+} B_t, \qquad T = T_1 \dot{+} \dots \dot{+} T_t. \tag{12.6}$$

Now observe that

$$T \in M_m(R)^* \quad \Longleftrightarrow \quad T_s \in M_m(R_s)^*, \quad s \in \{1, \dots, t\},$$

and

$$A \approx B\big(T^{-1}AT = B\big) \quad \Longleftrightarrow \quad A_s \approx B_s \quad \big(B_s = T_s^{-1}A_sT_s\big),\ s \in \{1, \dots, t\}.$$

So all main the problems connected with the solution of the equality $X^{-1}AX = B$ are reduced to the case when R is a local ring.

The following result gives some sort of universal approach to the decomposition of a matrix.

THEOREM 12.3. *Let $A \in M_m(R)$ and $F(x) \in \mathcal{R}$ be a monic polynomial such that $F(A) = 0$ and*

$$F(x) = F_1(x) \cdot \dots \cdot F_t(x), \tag{12.7}$$

where $F_1, \dots, F_t$ are monic mutually coprime polynomials and $|F_i(A)| \notin R^*$, $i \in \{1, \dots, t\}$. Then

$$A \approx \operatorname{Diag}(A_1, \dots, A_t), \qquad F_i(A_i) = 0, \quad i \in \{1, \dots, t\}. \tag{12.8}$$

Moreover, if $F(x) = \chi_A(x)$, then $F_i(x) = \chi_{A_i}(x)$, $i \in \{1, \dots, t\}$.

Theorems 12.3, 12.2, 2.12 imply

COROLLARY 12.1. *A square matrix A over a commutative local f.r. R is normal exactly if*

$$A \approx \operatorname{Diag}\big(S\big(G_1(x)\big), \dots, S\big(G_r(x)\big)\big), \tag{12.9}$$

where $G_1, \dots, G_r \in \mathcal{R}$ are primary monic polynomials. The matrix on the right hand in (12.9) is defined uniquely up to a permutation of blocks.

12.1.3. *Quasi-canonical polynomial matrix and Fitting invariants* The solution of Problems 1, 2 for matrices over a field is based on Theorem 12.1 and the fact that any characteristic matrix $Ex - A$ is equivalent to a unique canonical matrix (see e.g. [236]). Here instead of this result we have only

THEOREM 12.4. *Let A be an $m \times m$-matrix over a local commutative f.r. R. Then the matrix $xE - A$ is equivalent to a matrix*

$$\mathcal{K}(x) = \big(K_{ij}(x)\big), \tag{12.10}$$

where $K_{11}, \dots, K_{mm}$ are monic polynomials such that

$$\sum \deg K_{ii}(x) = m, \quad \overline{K}_{ii}(x) \mid \overline{K}_{i+1i+1}(x),\ i \in \{1, \dots, m-1\};$$

and if $i \neq j$, then $\overline{K}_{ij}(x) = \bar{0}$, $\deg K_{ij}(x) < \min\{\deg K_{ii}(x),\ \deg K_{jj}(x)\}$.

For any matrix (12.10) with the properties above there exists a matrix $A \in M_m(R)$ such that $xE - A \sim \mathcal{K}(x)$.

We call the matrix (12.10) the *quasi-canonical* matrix equivalent to $xE - A$ or *quasi-canonical form* of $xE - A$. In [291] a constructive algorithm for finding such a form for $xE - A$ has been given.

Unfortunately, in general the quasi-canonical form of the matrix $xE - A$ is not uniquely defined. For example, if $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 3 \end{smallmatrix}\right) \in M_2(\mathbb{Z}_4)$ then some quasi-canonical forms of $xE - A$ are

$$\begin{pmatrix} x-1 & 0 \\ 0 & x-3 \end{pmatrix}, \quad \begin{pmatrix} x-1 & 2 \\ 0 & x-3 \end{pmatrix}, \quad \begin{pmatrix} x-1 & 0 \\ 2 & x-3 \end{pmatrix},$$
$$\begin{pmatrix} x-1 & 2 \\ 2 & x-3 \end{pmatrix}, \quad \begin{pmatrix} x-3 & 0 \\ 0 & x-1 \end{pmatrix}, \quad \dots .$$

For any polynomial matrix $\mathfrak{A}(x) \in M_m(\mathcal{R})$ and for every $s \in \{1, \dots, m\}$ we define the s-th *Fitting invariant* as the ideal $\mathcal{D}_s(\mathfrak{A}(x)) \lhd \mathcal{R}$ generated by all minors of the matrix $\mathfrak{A}(x)$ of order s This definition corresponds to original definition in [123] (see [291]).

PROPOSITION 12.5. *Let $A, B \in M_m(R)$, then*

$$(B \approx A) \quad \Rightarrow \quad \big(\mathcal{D}_s(xE - B) = \mathcal{D}_s(xE - A),\ s \in \{1, \dots, m\}\big). \tag{12.11}$$

So the system of ideals

$$\mathcal{D}_1(xE - A), \dots, \mathcal{D}_m(xE - A) \tag{12.12}$$

is a system of invariants on $M_m(R)$. It is useful to note that this system contains the characteristic polynomial:

$$\mathcal{D}_m(xE - A) = \mathcal{R}\chi_A(x),$$

and allows one to describe one more classical invariant, since according to [259]

$$\operatorname{Ann}(A) = \mathcal{D}_m(xE - A) : \mathcal{D}_{m-1}(xE - A). \tag{12.13}$$

THEOREM 12.6. *(See [291].) For a matrix A over a commutative local f.r. R the following conditions are equivalent*:

(a) $xE - A \sim \operatorname{diag}(K_1(x), \dots, K_m(x))$, $K_i(x) \mid K_{i+1}(x)$, $i \in \{1, \dots, m-1\}$;

(b) *there exists a unique quasi-canonical matrix equivalent to $xE - A$*;

(c) *all Fittings invariants of $xE - A$ are principal ideals.*

If minimal polynomial of the matrix $\overline{A}$ over the field $\overline{R}$ equals to $\overline{\chi}_{\overline{A}}(x)$ then $A \approx S(\chi_A(x))$.

It is interesting to note that if R is a field then any matrix $A \in M_m(R)$ satisfies all of conditions listed in previous theorem.

12.1.4. *Canonically defined matrices* In general the system (12.12) is not a full system of invariants on $M_m(R)$, i.e. the proposition inverse to Proposition 12.5 is not true. We call a matrix $A \in M_m(R)$ *canonically defined* if for any matrix $B \in M_m(R)$ the implication inverse to (12.11) is true. Any matrix over a field is canonically defined.

Theorem 12.6 implies

PROPOSITION 12.7. *If all Fitting invariants of a matrix $A \in M_m(R)$ are principal ideals then A is canonically defined.*

THEOREM 12.8. *(See [291].) A normal matrix $A \in M_m(R)$ is canonically defined precisely if all invariants (12.12) are principal ideals.*

However, there exists matrices over rings R which are not canonically defined.

EXAMPLE 12.9. Matrix $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 3 \end{smallmatrix}\right) \in M_2(\mathbb{Z}_4)$ is a normal matrix with $\mathcal{D}_1(xE - A) = (x - 1, 2)_{\mathcal{R}}$ not a principal ideal. So A is not a canonically defined matrix.

These results and some other examples give reason for following

CONJECTURE CANDEF. *A matrix $A \in M_m(R)$ is canonically defined precisely if all invariants (12.12) are principal ideals.*

Additional support for this conjecture is provided by following results.

THEOREM 12.10. *(See [291].) Let R be a GE-ring with radical $\mathfrak{N}(R) = \pi R$, and* $\operatorname{ind}(\mathfrak{N}(R)) = n > 1$. *Then for $A \in M_2(R)$ only one of the following situations is possible*:

(a) $\mathcal{D}_1(xE - A) = (e)$, $\operatorname{Ann}(A) = (\chi_A(x))$, *and* $A \approx S(\chi_A(x))$;

(b) $\mathcal{D}_1(xE - A) = (x - r)$, $\operatorname{Ann}(A) = (x - r)$, *for some* $r \in R$, *and* $A = rE$;

(c) $\mathcal{D}_1(xE - A) = (x - r, \pi^k)$, $\operatorname{Ann}(A) = (\chi_A(x), \pi^{n-k})$, *for some* $r \in R$, $k \in \{1, \dots, n-1\}$, *and* $A \approx rE + \pi^k S(G(x))$, $G(x) = x^2 - ax - b$, $\chi_A(x) = (x - r)^2 - \pi^k(x - r) - \pi^{2k}b$.

The matrix A is canonically defined precisely if $\mathcal{D}_1(xE - A)$ is a principal ideal.

THEOREM 12.11. *(See [291].) Let R be a GE-ring with radical $\mathfrak{N}(R) = \pi R$, and* $\operatorname{ind}(\mathfrak{N}(R)) = 2$. *Then a matrix $A \in M_3(R)$ is canonically defined exactly if $\mathcal{D}_1(xE - A)$, $\mathcal{D}_2(xE - A)$ are principal ideals.*

THEOREM 12.12. *(See [142].) Let R be a GE-ring. Then a matrix $A \in M_m(R)$ with the property $\overline{A} = a\overline{E}$, $a \in \overline{R}$, is canonically defined precisely if all invariants (12.12) are principal ideals.*

12.1.5. *Polynomial defined matrices and radical identities* Here R is a commutative local f.r. with radical $\mathfrak{N} = \mathfrak{N}(R)$ and residue field $\overline{R} = R/\mathfrak{N} = GF(q)$. The images of a matrix $A \in M_m(R)$, a polynomial $F(x) \in \mathcal{R}$ and a subset $I \subseteq \mathcal{R}$ under the natural homomorphism $R \to \overline{R}$ will be denoted correspondingly by $\overline{A}$, $\overline{F}(x)$, $\overline{I}$.

Note that for any matrices $A, B \in M_m(R)$ the equalities $\overline{\mathcal{D}_s(xE - A)} = \mathcal{D}_s(x\overline{E} - \overline{A})$, $s \in \{1, \dots, m\}$ hold and the implication

$$(A \approx B) \quad \Rightarrow \quad \big(\operatorname{Ann}(A) = \operatorname{Ann}(B),\ \overline{A} \approx \overline{B}\big) \tag{12.14}$$

is fulfilled. We call a matrix $A \in M_m(R)$ *polynomially defined* if for every $B \in M_m(R)$ the implication inverse to (12.14) is true. Every polynomially defined matrix is canonically defined.

THEOREM 12.13. *(See [291].) A matrix $A \in M_m(R)$ is polynomially defined if and only if the following conditions hold*:

(a) $\operatorname{Ann}(A)$ *is a principal ideal*;

(b) *any two not coprime elementary divisors of the matrix $x\overline{E} - \overline{A}$ are equal.*

Under these conditions all Fitting invariants of the matrix $xE - A$ are principal ideals and if $\operatorname{Ann}(F(x))$ *where $F(x)$ has the canonical decomposition $F(x) = F_1(x) \cdots F_t(x)$, then*

$$A \approx \operatorname{Diag}\bigl(S\bigl(F_1(x)\bigr), S\bigl(F_2(x)\bigr), \ldots, S\bigl(F_t(x)\bigr)\bigr).$$

We call a monic polynomial $F(x) \in \mathcal{R}$ a *radical identity* if

$$\forall A, B \in \mathcal{V}\bigl(F, M_m(R)\bigr): \quad A \approx B \quad \Longleftrightarrow \quad \overline{A} \approx \overline{B}.$$

THEOREM 12.14. *(See [291].) Let a monic polynomial $F(x) \in \mathcal{R}$ be a product $F(x) = H_1(x) \cdots H_t(x)$ of mutually coprime Galois–Eisenstein polynomials* (*see Section* 5.1.1). *Then F is a radical identity. Moreover, $\mathcal{V}(F, M_m(R))$ is the set of all matrices satisfying some relation of the form*

$$A \approx \operatorname{Diag}\bigl(S(H_{i_1}), \ldots, S(H_{i_s})\bigr).$$

COROLLARY 12.2. *Any monic separable polynomial $G(x) \in \mathcal{R}$ is a radical identity and any matrix $A \in \mathcal{V}(G, M_m(R))$ is normal.*

Examples such are the polynomials
$x^t - e \in \mathcal{R}$ if $(t, \operatorname{char} R) = 1$ [318,366];
$x^2 - x \in \mathcal{R}$ [176].

12.1.6. *Kurakin invariants for the similarity of matrices [225]* Let R be a GE-ring. Then according to Proposition 5.2 there exists a local principal ideal domain Z with $\mathfrak{N}(Z) = \pi Z$ such that $R \cong Z/\pi^n Z$ for some natural number n. Let $\varphi : M_m(Z) \to M_m(R)$ be the natural morphism with kernel $\pi^n Z$. For a matrix $A_1 \in M_m(Z)$ let $\mathcal{D}_k^Z(Ex - A_1)$ be the Fitting invariant of A_1 in $Z[x]$. Denote by $\widehat{\mathcal{D}}_k(Ex - A)$ the ideal in $Z[x]$ generated by the sets $\mathcal{D}_k^S(Ex - A_1)$ for all $A_1 \in M_m(S)$ such that $\varphi(A_1) = A$:

$$\widehat{\mathcal{D}}_k(Ex - A) = \sum_{A_1 \in \varphi^{-1}(A)} \mathcal{D}_k^Z(Ex - A_1).$$

PROPOSITION 12.15. *Let $A \in M_m(R)$. Then for any fixed matrix $A_1 \in \varphi^{-1}(A)$ and for every $k \in \{1, \ldots, m\}$ the equality*

$$\widehat{\mathcal{D}}_k(Ex - A) = \mathcal{D}_k^Z(Ex - A_1) + \pi^n \mathcal{D}_{k-1}^Z(Ex - A_1) + \cdots + \pi^{(k-1)n} \mathcal{D}_1^Z(Ex - A_1) + \pi^{kn} Z[x].$$

holds.

THEOREM 12.16.

$$\forall A, B \in M_m(R): \quad (A \approx B) \\ \Rightarrow \quad \big(\widehat{\mathcal{D}}_k(Ex - A) = \widehat{\mathcal{D}}_k(Ex - B),\ k \in \{1, \ldots, m\}\big).$$

If $m = 2$ the converse implication is true.

So the *Kurakin invariants* $\widehat{\mathcal{D}}_k(Ex - A)$ of a matrix A are stronger than the Fitting invariants (see Theorem 12.10). There exist examples of non-similar 3×3-matrices A, B, such that $\widehat{\mathcal{D}}_k(Ex - A) = \widehat{\mathcal{D}}_k(Ex - B)$ for $k \in \{1, 2, 3\}$ [364].

Related sources: [50,83,84,122,256].

12.2. *Multiplicative group of a matrix ring over a local f.r. [76,199,200,276,296]*

12.2.1. *Exponents and elements of maximal order* Here S is a local f.r. with radical $\mathfrak{N}$ of nilpotency index n and residue field $\bar{S} = GF(q)$ of characteristic p. We consider the multiplicative group $M_m(S)^*$ of the ring $M_m(S)$ of all $m \times m$-matrices over S for $m > 1$ (for the case $m = 1$ see Sections 2.2, 4.5, 5.1.3).

It is evident that

$$\exp M_m(S) \leqslant \exp M_m(\bar{S}) \exp\big(E + M_m(\mathfrak{N})\big), \tag{12.15}$$

$$\exp M_m(\bar{S}) = \big[q^2 - 1, \ldots, q^m - 1\big] p^{\sigma}, \quad \sigma =]\log_p m[. \tag{12.16}$$

(Here $[a, b, \ldots, c]$ is the lcm of the integers $a, b, \ldots, c$).

THEOREM 12.17. *Let $\nu =]\log_p n[$, $\operatorname{char} \mathfrak{N}^{p^s} = p^{d_s}$ for $s \in \{0, \ldots, \nu\}$, $\omega = \max\{s + d_s: s \in \{0, \ldots, \nu\}\}$. Then*

$$\exp\big(E + M_m(\mathfrak{N})\big) = p^{\mu(S)}, \quad \mu(S) \leqslant \omega \leqslant d_0 + \nu - 1. \tag{12.17}$$

If S is a balanced local f.r. with ramification index ε (see Section 2.2), then

$$\omega = \left]\frac{n - p^b}{\varepsilon}\right[+ b, \quad \text{where } b = \max\left\{0, \left]\log_p \frac{\varepsilon}{p - 1}\right[\right\}. \tag{12.18}$$

If S is a GE-ring then

$$\exp\big(E + M_m(\mathfrak{N})\big) = p^{\omega}. \tag{12.19}$$

THEOREM 12.18. *Let $R = GR(q^n, p^n)$ be a Galois ring. Then*

$$\exp M_m(R)^* = \exp M_m(\bar{R}) \exp\big(E + M_m(pR)\big) \\ = \big[q^2 - 1, \ldots, q^m - 1\big] p^{\rho}, \quad \rho =]\log_p m[+ n - 1.$$

Let S be a GE-ring with radical $\mathfrak{N} = \pi S$ and $\pi^{\varepsilon} = p v_{\pi}$. Let $T_{\max}(S, m)$ be maximum of the orders $T(A)$ of invertible matrices $A \in M_m(S)^*$. Then

$$T_{\max}(S, m) \leqslant T_{\max}(\bar{S}, m) \exp\big(E + M_m(\mathfrak{N})\big) = \big(q^m - 1\big) \exp\big(E + M_m(\mathfrak{N})\big).$$

For any $\lambda \in \mathbb{N}$ define

$$b(\lambda) = \max\left\{0, \left]\log_p\left(\frac{\varepsilon}{\lambda(p-1)}\right)\right[\right\},$$

$$\omega(\lambda) = \left]\frac{n - \lambda p^{b(\lambda)}}{\varepsilon}\right[+ b(\lambda). \tag{12.20}$$

THEOREM 12.19. *Let S be a GE-ring. Then*

(a) *under condition* (2.10) $T_{\max}(S, m) = T_{\max}(\bar{S}, m)\exp(E + \mathfrak{N}) = (q^m - 1)p^{\omega}$.

(b) *If $A \in M_m(S)^*$, $T(A) = T_{\max}(S, m)$, then*

(b1) $\chi_{\bar{A}}(x) = g(x) \in \bar{S}[x]$ *is a polynomial of maximal period* $q^m - 1$;

(b2) *the matrix A has the form $A = A_*(E + \pi^{\lambda}H(A_*))$, where $\chi_{A_*}(x) = G(x)$ is a distinguished polynomial corresponding to $g(x)$* (*see Section* 11.2.2),

$$H(x) \in S[x], \quad \deg H(x) < m,$$
$$\overline{H}(x) \neq \bar{0}, \quad 1 \leqslant \lambda < n, \quad \omega(\lambda) = \omega.$$

(c) *If a matrix $A \in M_m(S)^*$ satisfies conditions* (b1), (b2) *and if, given the condition*

$$\lambda(p-1)p^{b(\lambda)} = \varepsilon, \quad \lambda p^{b(\lambda)+1} < n,$$

an element $v \in S^*$ with the property $\bar{v}^{p^{b(\lambda)}} = \bar{v}_{\pi}$ satisfies the condition

$$\overline{H}(x)^p + \bar{v}\overline{H}(x) \neq 0 \quad \bigl(\operatorname{mod} \overline{G}(x)\bigr),$$

then $T(A) = T_{\max}(S, m)$.

12.2.2. *Cyclic types of linear substitutions over a commutative f.r.* Let R be any commutative f.r. with identity, let $A \in M_m(R)^*$ be an invertible $(m \times m)$-matrix over R and let φ_A be the *linear substitution* on the module $R^{(m)}$ of all columns of the length m, acting on a column $\alpha \in R^{(m)}$ by the rule: $\varphi(\alpha) = A\alpha$. Let $\mathbb{Z}_1[y]$ be the set of all polynomials over the ring $\mathbb{Z}$ with zero free coefficient. The *cyclic type* of the substitution φ_A is the polynomial $C_A(y) \in \mathbb{Z}_1[y]$ given by

$$C_A(y) = \sum_{t=1}^{N} c_A(t)y^t, \quad N = \bigl|R^{(m)}\bigr|,$$

where $c_A(t)$ is the number of cycles of length t in φ_A.

Let us define a composition $*$ of elements $B(y) = \sum_{t\in\mathbb{N}} b(t)y^t$, $C(y) = \sum_{t\in\mathbb{N}} c(t)y^t \in \mathbb{Z}_1[y]$ in the following way:

$$B(y) * C(y) = D(y) = \sum_{t\in\mathbb{N}} d(t)y^t, \quad \text{where } d(t) = \sum_{r,s\in\mathbb{N},\ [r,s]=t} (r,s)b(r)c(s)$$

(here (r, s) is the gcd and $[r, s]$ is the lcm of two numbers $r, s \in \mathbb{N}$).

PROPOSITION 12.20. *The algebra $(\mathbb{Z}_1[y], +, *)$ is a commutative ring with the identity y.*

It is useful to note that if

$$C_1(y) = \sum_{t\in\mathbb{N}} c_1(t)y^t, \quad \ldots, \quad C_k(y) = \sum_{t\in\mathbb{N}} c_k(t)y^t \in \mathbb{Z}_1[y]$$

then

$$C_1(y) * \cdots * C_k(y) = \sum_{t_1\in\mathbb{N}} \cdots \sum_{t_k\in\mathbb{N}} \frac{t_1 \cdots t_k}{[t_1, \ldots, t_k]} c_1(t_1) \cdots c_k(t_k) y^{[t_1,\ldots,t_k]}.$$

Let

$$R = R_1 \dot{+} \cdots \dot{+} R_k \tag{12.21}$$

be a decomposition of the ring R into a direct sum of commutative local f.r. $R_1, \ldots, R_k$ with identities $e_1, \ldots, e_k$. Then any matrix $A \in M_m(R)$ has decomposition

$$A = A_1 + \cdots + A_k, \quad A_i = e_i A \in M_m(R_i), \quad i \in \{0, \ldots, k\}, \tag{12.22}$$

and

$$A \in M_m(R)^* \iff A_i \in M_m(R_i)^*, \quad i \in \{0, \ldots, k\},$$

PROPOSITION 12.21. *Let $A \in M_m(R)^*$ and either the conditions (12.21), (12.22) hold or the matrix A is decomposable*: $A \approx \mathrm{Diag}(A_1, \ldots, A_k)$. *Then there is the equality*

$$C_A(y) = C_{A_1}(y) * \cdots * C_{A_k}(y). \tag{12.23}$$

12.2.3. *Calculation of cyclic type in the primary case* In view of Corollary 1.1 and Theorem 12.3 Proposition 12.21 shows that the problem of the calculation of the cyclic type of an invertible matrix over a finite commutative ring R is reduced to the case when R is a local ring and the characteristic polynomial $\chi_A(x)$ of the matrix A is primary.

Let now R be a commutative local f.r. and $A \in M_m(R)^*$ be a matrix with primary characteristic polynomial such that $\overline{\chi}_A(x) = \chi_{\overline{A}}(x) = g(x)^k$, where $g(x) \in \overline{R}[x]$ is an irreducible polynomial.

Note that for any $s \in \mathbb{N}_0$ there is the equality $g(x)^{p^s} = g^{(s)}(x^{p^s})$, where $g^{(s)}(x) \in \overline{R}[x]$ is some irreducible polynomial of period $\tau = T(g^{(s)}) = T(g)$. Let $G^{(s)}(x) \in \mathcal{R}$ be the distinguished polynomial corresponding to $g^{(s)}(x)$. For any $s \in \mathbb{N}_0$

$$T\big(G^{(s)}\big) = T\big(g^{(s)}\big) = T(g) = \tau, \quad (\tau, p) = 1.$$

Below, for any $m \times m$-matrix B over the ring R the set of all solutions in $R^{(m)}$ of the equation $Bx^{\downarrow} = 0^{\downarrow}$ will be denoted by $\mathfrak{N}(B)$.

THEOREM 12.22. *Let the minimal polynomial of the matrix $\overline{A}$ be $g(x)^l$. Then the cyclic type of the substitution φ_A has the form*

$$C_A(y) = y + \sum_{s=0}^{\sigma} c_A\big(\tau p^s\big) y^{\tau p^s}, \quad \textit{where}$$

$$c_A(\tau) = \frac{1}{\tau}\bigl(\bigl|\mathfrak{R}\bigl(G^{[0]}(A)\bigr)\bigr| - 1\bigr),$$
$$c_A\bigl(\tau p^s\bigr) = \frac{1}{\tau p^s}\bigl(\bigl|\mathfrak{R}\bigl(G^{[s]}\bigl(A^{p^s}\bigr)\bigr)\bigr| - \bigl|\mathfrak{R}\bigl(G^{[s-1]}\bigl(A^{p^{s-1}}\bigr)\bigr)\bigr|\bigr), \tag{12.24}$$

and under the condition of Theorem 12.17 $]\log_p l[\ \leqslant \sigma \leqslant\ ]\log_p l[+ \omega$.

The additional assumption that $R = S$ is a GE-ring allows one to obtain a more interesting formulas for the coefficients in (12.24). Below we use notions and notations of Section 5.1.2. Let for $s \geqslant 0$

$$\operatorname{sign}\bigl(G^{[s]}\bigl(A^{p^s}\bigr)\bigr) = \bigl[d_1(s), \ldots, d_m(s)\bigr], \quad d(s) = d_1(s) + \cdots + d_m(s). \tag{12.25}$$

Let s_1 be the minimum of the values $s \in \mathbb{N}_0$ with the property

$$d_1(s) > \frac{\varepsilon}{2}, \quad \text{if } p > 2; \qquad d_1(s) > \varepsilon, \quad \text{if } p = 2,\ n > \varepsilon;$$
$$d_1(s) = \varepsilon, \quad \text{if } p = 2,\ n = \varepsilon.$$

THEOREM 12.23. *Under the assumptions above the order of the substitution φ_A is*

$$T(A) = \tau p^\sigma, \quad \textit{where } \sigma = s_1 + \left]\frac{n - d_1(s_1)}{\varepsilon}\right[,$$

and the coefficients in (12.24) satisfy the equalities

$$c_A(\tau) = \frac{1}{\tau}\bigl(q^{d(0)} - 1\bigr), \qquad c_A\bigl(\tau p^s\bigr) = \frac{1}{\tau p^s}\bigl(q^{d(s)} - q^{d(s-1)}\bigr).$$

Moreover, if $T(\overline{A}) = \tau p^\xi$ and $b(\lambda)$ is defined by (12.20) then

$$s_1 \leqslant b\bigl(d_1(\xi)\bigr) + \xi + 1$$

and for every $s > s_1$ the parameter $d(s)$ can be calculated from (12.25) and there are the equalities

$$d_j(s) = \min\bigl\{n,\ d_j(s_1) + (s - s_1)\varepsilon\bigr\}, \quad j \in \{1, \ldots, m\}.$$

13. (Poly)-linear recurrences over a finite module

13.1. *General theory [226,233–235,278,297–299,301,303]*

Let ${}_AM$ be a faithful finite module over a f.r. A with the identity e. For a fixed $k \in \mathbb{N}$ any function $u : \mathbb{N}_0^k \to M$ will be called *a k-sequence over M*. Let $M^{\langle k\rangle}$ be the A-module of all k-sequences over M (with usual pointwise addition of functions and multiplication by constants from A). We define on $M^{\langle k\rangle}$ a structure of a left module over the polynomial ring $\mathcal{A}_k = A[\mathbf{x}] = A[x_1, \ldots, x_k]$. Any $H(\mathbf{x}) \in \mathcal{A}_k$ has the form $H(\mathbf{x}) = \sum_{\mathbf{j}\in\mathbb{N}_0^k} c_{\mathbf{j}}\mathbf{x}^{\mathbf{j}}$, where $\mathbf{x}^{\mathbf{j}} = x_1^{j_1} \cdots x_k^{j_k}$ for $\mathbf{j} = (j_1, \ldots, j_k) \in \mathbb{N}_0^k$. The product of $u \in M^{\langle k\rangle}$ by $H(\mathbf{x})$ is defined as

$$H(\mathbf{x})u = v \in M^{\langle k\rangle}, \qquad v(\mathbf{i}) = \sum_{\mathbf{j}\in\mathbb{N}_0^k} c_{\mathbf{j}} v(\mathbf{i}+\mathbf{j}), \quad \mathbf{i} \in \mathbb{N}_0^k.$$

13.1.1. *Relations between LRS-families* A k-sequence $u \in M^{\langle k\rangle}$ is called *a k-linear recurring sequence (k*-LRS*)* over the module ${}_AM$, if it admits a *system of elementary characteristic polynomials*, i.e. a system of monic polynomials

$$F_s(x_s) = x_s^{m_s} - c_{s,m_s-1}x_s^{m_s-1} - \cdots - c_{s1}x_s - c_{s0} \in \mathcal{A}_k, \quad s \in \{1,\ldots,k\}, \tag{13.1}$$

such that $F_s(x_s)u = 0$, $s \in \{1,\ldots,k\}$. For $k = 1$ this definition turns into the usual definition of an LRS over ${}_AM$. The set $\mathcal{L}_AM^{\langle k\rangle}$ of all k-LRS over ${}_AM$ is an $\mathcal{A}_k$-submodule of $M^{\langle k\rangle}$. But this is not true in general (if ${}_AM$ is infinite module).

An ideal of the ring $\mathcal{A}_k$ is called *monic*, if it contains some system (13.1) of monic polynomials. A subset $\chi \subset \mathcal{A}_k$ is called *left monic*, if the left ideal $\mathcal{A}_k\chi$ generated by χ is a monic.

Below we use notations of Section 6. Evidently a sequence $u \in M^{\langle k\rangle}$ is a k-LRS over the ring A precisely if its left annihilator in $\mathcal{A}_k$: $\lambda(u) = \lambda_{\mathcal{A}_k}(u)$ is a monic ideal. If I is a monic ideal of $\mathcal{A}_k$ then its right annihilator in $M^{\langle k\rangle}$: $\rho(I) = \rho_{M^{\langle k\rangle}}(I)$ is a finite set of k-LRS called *k-LRS family with characteristic ideal I*. Moreover if the ideal I contains a system (13.1) of monic polynomials, then $|\rho(I)| \leqslant |M|^{m_1m_2\cdots m_k}$.

If ${}_AM$ is a bimodule ${}_AM_B$ (e.g. $B = \mathrm{End}({}_AM)$) in a manner very similar to that of the left module case we define the product $uH(\mathbf{x})$ of $u \in M^{\langle k\rangle}$ and $H(\mathbf{x}) \in \mathcal{B}_k = B[\mathbf{x}] = B[x_1,\ldots,x_k]$ and define the notion of k-LRS over a module M_B. The set of all k-LRS over M_B will be denoted by $\mathcal{L}M_B^{\langle k\rangle}$.

For any left ideals I, I_1, and I_2 of $\mathcal{A}_k$ and right submodules R, R_1, and R_2 of $M_{\mathcal{B}_k}^{\langle k\rangle}$

$$\lambda\bigl(\rho(I)\bigr) \supseteq I, \qquad \rho\bigl(\lambda(R)\bigr) \supseteq R, \tag{13.2}$$

$$\rho(I_1 \cap I_2) \supseteq \rho(I_1) + \rho(I_2), \qquad \lambda(R_1 \cap R_2) \supseteq \lambda(R_1) + \lambda(R_2), \tag{13.3}$$

$$\rho(I_1 + I_2) = \rho(I_1) \cap \rho(I_2), \qquad \lambda(R_1 + R_2) = \lambda(R_1) \cap \lambda(R_2). \tag{13.4}$$

For right ideals of $\mathcal{B}_k$ and left submodules of ${}_{\mathcal{A}_k}M^{\langle k\rangle}$ the obvious left-right versions of these relations are also true.

It is well known that if $M = A = B$ is a field and $k = 1$ then relations (13.2), (13.3) are equalities. In order to develop a profound enough theory of LRS's over ${}_AM$ by analogy with the theory of LRS's over a field we need to preserve these equalities. A generalization of these relations is connected with the notion of QF-bimodule (Section 6).

THEOREM 13.1. *For a finite faithful bimodule ${}_AM_B$ the following statements are equivalent.*

(a) *${}_AM_B$ is a QF-bimodule.*

(b) *For any ideals $I \leqslant {}_{\mathcal{A}_k}\mathcal{A}_k$ and $J \leqslant \mathcal{B}_{k\mathcal{B}_k}$ and for any finite submodules $R < M_{\mathcal{B}_k}^{\langle k\rangle}$ and $L < {}_{\mathcal{A}_k}M^{\langle k\rangle}$ the equalities $I = \lambda(\rho(I))$, $R = \rho(\lambda(R))$, $J = \rho(\lambda(J))$, $L = \lambda(\rho(L))$ hold.*

(c) *The mappings $I \mapsto \rho(I)$ and $R \mapsto \lambda(R)$ are mutually inverse Galois correspondences between the set of monic left ideals I of the ring $\mathcal{A}_k$ and the set of right finite submodules $R < M_{\mathcal{B}_k}^{\langle k\rangle}$. The mappings $J \mapsto \lambda(J)$ and $L \mapsto \rho(L)$ are mutually inverse Galois correspondences between the set of monic right ideals J of the ring $\mathcal{B}_k$ and the set of left finite submodules $L < {}_{\mathcal{A}_k}M^{\langle k\rangle}$.*

(d) *If S is a finite $(\mathcal{A}_k, \mathcal{B}_k)$-subbimodule of $M^{\langle k\rangle}$ and $I = \lambda(S)$, $J = \rho(S)$, then the $(\mathcal{A}_k/I, \mathcal{B}_k/J)$-bimodule S is a QF-bimodule.*

COROLLARY 13.1. *Let ${}_AM_B$ be a finite QF-bimodule. Then*

(a) *If I is a two-sided monic ideal of $\mathcal{A}_k$, then $S = \rho(I)$ is a finite $(\mathcal{A}_k, \mathcal{B}_k)$-subbimodule of $M^{\langle k\rangle}$ and $J = \rho(S)$ is a two-sided monic ideal of $\mathcal{B}_k$.*

(b) *The mappings $I \mapsto \rho(I)$, $S \mapsto \lambda(S)$ and $J \mapsto \lambda(J)$, $S \mapsto \rho(S)$ are one-to one correspondences between three sets: the set of two-sided monic ideals I of $\mathcal{A}_k$, the set of finite $(\mathcal{A}_k, \mathcal{B}_k)$-subbimodules of $M^{\langle k\rangle}$ and the set of two-sided monic ideals J of the ring $\mathcal{B}_k$.*

(c) *For a monic left ideal I of $\mathcal{A}_k$ the family $L_M(I)$ is a left $\mathcal{A}_k$-module if and only if I is a two-sided ideal of $\mathcal{A}_k$.*

(d) *For any monic left ideals I_1, I_2 of $\mathcal{A}_k$ and finite right submodules R_1, R_2 of $M^{\langle k\rangle}_{\mathcal{B}_k}$ the inclusions (13.3) and their left-right versions are equalities.*

(e) *$\mathcal{L}_AM^{\langle k\rangle} = \mathcal{L}M^{\langle k\rangle}_B$ is a $(\mathcal{A}_k, \mathcal{B}_k)$-subbimodule of $M^{\langle k\rangle}$.*

In the commutative case property (d) of Corollary 13.1 implies that ${}_AM_B$ is a QF-bimodule, in the noncommutative case the correctness of this assertion is an *open question*.

13.1.2. *Criterion of cyclicity of an LRS-family over a QF-module with commutative coefficient ring* It is well known that if $M = A = B$ is a field and $k = 1$ then for any monic ideal $I \lhd \mathcal{A}_1$ the LRS-family $\rho(I)$ is a cyclic $\mathcal{A}_1$-module.

A generalization of this result for LRS's over a module ${}_RM$ with commutative coefficient ring R is the following. Here we use the notation $L_M(I) = \rho_{M^{\langle k\rangle}}(I)$ for a k-LRS family with characteristic ideal I.

If R is a commutative f.r. then the nilradical $\mathfrak{N} = \mathfrak{N}(R)$ of R is the set of all nilpotent elements of R: '$\mathfrak{N}(R) = \sqrt{0}$. It is not difficult to see that the radical $\sqrt{I}$ of any ideal $I \lhd \mathcal{R}_k$ satisfies the following relations: $\mathfrak{N}\mathcal{R}_k \subseteq \sqrt{I}$, $\mathfrak{N}(I : \sqrt{I}) \subseteq I$. This means that we can consider the R-modules $\mathcal{R}_k/\sqrt{I}$ and $(I : \sqrt{I})/I$ as modules over the top-factor $\overline{R} = R/\mathfrak{N}$ of the ring R. The following result here is the endproduct of a series of result's [226,253,295,297,298,301]

THEOREM 13.2. *(See [160].) Let ${}_RQ$ be a quasi-Frobenius module over a commutative f.r. R. Then for any monic ideal $I \lhd \mathcal{R}_k$ the LRS-family $\mathcal{M} = L_Q(I)$ is a QF-module over the commutative f.r. $S = \mathcal{R}_k/I$ and the following conditions are equivalent:*

(a) *$I = \lambda(u)$ for some recurrence $u \in \mathcal{L}Q^{\langle k\rangle}$;*

(b) *$\mathcal{M}$ is a cyclic $\mathcal{R}_k$-module: $\mathcal{M} = \mathcal{R}_ku$;*

(c) *S is a quasi-Frobenius ring;*

(d) *there exists an isomorphism of modules ${}_{\overline{R}}(\mathcal{R}_k/\sqrt{I}) \cong {}_{\overline{R}}((I : \sqrt{I})/I)$.*

In [295] for the case when $k = 1$ and R is a GE-ring an algorithm was given for checking the conditions of Theorem 13.2 using a canonical generating system of the ideal I.

13.1.3. *Periodic k-sequences* Let ${}_RM$ be a finite faithful module over a commutative f.r. R. For $\mu \in M^{\langle k\rangle}$ the set $\mathcal{O}(\mu)$ of all k-sequences $\nu \in M^{\langle k\rangle}$ of the form $\nu = \mathbf{x}^{\mathbf{i}}\mu$, $\mathbf{i} \in \mathbb{N}_0^k$, is called the *trajectory* of μ.

A sequence $\mu \in M^{\langle k\rangle}$ is called *periodic* if its trajectory $\mathcal{O}(\mu)$ is finite, and it is called *reversible* if $\mathcal{O}(\mu) = \mathcal{O}(\mathbf{x}^{\mathbf{i}}\mu)$ for every $\mathbf{i} \in \mathbb{N}_0^k$.

For a periodic sequence μ the set $\mathcal{T}(\mu)$ of all reversible elements of its trajectory $\mathcal{O}(\mu)$ is called the *cycle* of the sequence μ, and its cardinality $T(\mu) = |\mathcal{T}(\mu)|$ is called the *period* of the sequence μ. The elements of the set $\mathcal{D}(\mu) = \mathcal{O}(\mu) \setminus \mathcal{T}(\mu)$ are called *defect elements* of the trajectory $\mathcal{O}(\mu)$ and its cardinality $D(\mu) = |\mathcal{D}(\mu)|$ is called the *defect* of the sequence μ. The sequence μ is said to be *degenerated* if it is periodic and its cycle contains only the zero sequence, i.e., $\mathcal{T}(\mu) = \{0\}$.

Thus, $D(\mu) + T(\mu) = |\mathcal{O}(\mu)|$, and a periodic sequence μ is reversible iff $D(\mu) = 0$, i.e., $\mathcal{T}(\mu) = \mathcal{O}(\mu)$. A periodic sequence is degenerated iff $\mathbf{x}^{\mathbf{i}} \in \mathrm{An}(\mu)$ for some $\mathbf{i} \in \mathbb{N}_0^k$.

Proposition 13.3. *A k-sequence $\mu \in M^{\langle k\rangle}$ is a periodic* (*respectively reversible*, *degenerated*) *sequence exactly if its annihilator $I = \mathrm{An}(\mu) \lhd \mathcal{R}_k$ is a periodic* (*respectively reversible*, *degenerated*) *ideal* (*see Section* 11.4).

If μ is a periodic sequence, *then $T(\mu) = T(I)$*, *$D(\mu) = D(I)$.*

The reversibility of the sequence $\mu \in M^{\langle k\rangle}$ is equivalent to the condition

$$\forall \mathbf{i} \in \mathbb{N}_0^k \exists \mathbf{j} \in \mathbb{N}_0^k\colon \quad \mathbf{x}^{\mathbf{j}}\big(\mathbf{x}^{\mathbf{i}}\mu\big) = \mu.$$

Proposition 13.4. *A k-sequence μ over a finite module ${}_RM$ is periodic iff it is a k-LRS.*

A nonzero vector $\mathbf{t} \in \mathbb{N}_0^k$ is called a *vector-period* of the sequence $\mu \in M^{\langle k\rangle}$ if $\mathbf{x}^{\mathbf{l}}(\mathbf{x}^{\mathbf{t}} - e)\mu = 0$ for some $\mathbf{l} \in \mathbb{N}_0^k$. A subgroup $\mathfrak{P}(\mu)$ of the group $(\mathbb{Z}^k, +)$, generated by all vector-periods of μ, will be called its *group of periods*.

Proposition 13.5. *For every $\mu \in M^{\langle k\rangle}$ the set $\mathfrak{P}^+(\mu)$ of all nonzero nonnegative vectors from $\mathfrak{P}(\mu)$ coincides with the set of all vector-periods of the sequence μ. For any subgroup $\mathcal{G} < \mathbb{Z}^k$ which is generated by the set $\mathcal{G}^+$ of all of its nonnegative vectors*, *there exists a k-sequence $\mu \in R^{\langle k\rangle}$ such that $\mathfrak{P}(\mu) = \mathcal{G}$. If μ is a periodic k-sequence with annihilator $I = \mathrm{An}(\mu)$ then $\mathfrak{P}(\mu) = \mathfrak{P}(I)$*, $\operatorname{rank}\mathfrak{P}(\mu) = k$, *and $T(\mu) = [\mathbb{Z}^k : \mathfrak{P}(\mu)]$.*

Note that if $\operatorname{rank}\mathfrak{P}(\mu) = k$, then μ is not necessarily a periodic sequence. The sequence $\mu \in M^{\langle 2\rangle}$ of the form

$$\begin{array}{|ccccccccccc}
\hline
\alpha & 0 & \alpha & 0 & 0 & \alpha & 0 & 0 & 0 & \alpha & \dots \\
0 & \alpha & 0 & \alpha & 0 & \alpha & \dots & & & & \\
\alpha & 0 & \alpha & 0 & \alpha & 0 & \dots & & & & \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & & & &
\end{array}$$

where $\alpha \neq 0$, is nonperiodic, but $\operatorname{rank}\mathfrak{P}(\mu) = 2$.

Let now ${}_RM$ be a finite faithful module over a commutative f.r. R with identity. Then the annihilator $\mathrm{An}(\mu) = \lambda(\mu)$ of any $\mu \in M^{\langle k\rangle}$ is an ideal of the polynomial ring $\mathcal{R}_k = R[x_1, \dots, x_k]$. The quotient ring $S = \mathcal{R}_k/An(\mu)$ is called the *operator ring* of the sequence μ. It has a form $S = R[\vartheta_1, \dots, \vartheta_k]$, where $\vartheta_s = x_s + \mathrm{An}(\mu)$, THEOREM 13.6. *The sequence μ is reversible if and only if its operator ring S is finite and $\vartheta_1, \dots, \vartheta_k \in S^*$. If μ is a reversible sequence, then*

$$T(\mu) = \bigl|\langle\vartheta_1, \dots, \vartheta_k\rangle\bigr| \leqslant |S^*| \leqslant |S| - 1,$$

where $\langle\vartheta_1, \dots, \vartheta_k\rangle$ is the subgroup of the group S^* generated by $\vartheta_1, \dots, \vartheta_k$. The equality

$$T(\mu) = |S^*|$$

holds iff

$$S^* = \langle\vartheta_1, \dots, \vartheta_k\rangle. \tag{13.5}$$

If μ is a faithful reversible sequence, then

$$T(\mu) = |S| - 1$$

if and only if the following three conditions hold:

(a) *$R = GF(q)$ is a Galois field*;

(b) $\mathrm{An}(\mu)$ *is a maximal ideal of the ring $\mathcal{R}_k = GF(q)[x_1, \dots, x_k]$* (*i.e.*, *$S = GF(q^n)$ for some $n \in \mathbb{N}$*);

(c) *the equality* (13.5) *holds.*

13.2. *Cycles of reversible LRS-families [226]*

Let ${}_RM$ be a finite faithful module over a commutative f.r. R with identity and let $I \lhd \mathcal{R}_k$ be a monic ideal. A k-LRS-family $L_M(I)$ is called *reversible* if any sequence $\mu \in L_M(I)$ is reversible.

PROPOSITION 13.7. *If $I \lhd \mathcal{R}_k$ is a reversible ideal, then $L_M(I)$ is a reversible family. If $L_M(I)$ is a reversible family, then the ideal $I' = \mathrm{An}(L_M(I))$ is reversible and $L_M(I) = L_M(I')$.*

Let $F_1(x_1), \dots, F_k(x_k) \in \mathcal{R}_k$ be a system of monic elementary polynomials. Then the family $L_M(F_1, \dots, F_k)$ is reversible iff the polynomials $F_1(x_1), \dots, F_k(x_k)$ are reversible.

Define the relation $\sim$ on the k-LRS-family $L_M(I)$ by

$$\mu \sim \nu \quad \Longleftrightarrow \quad \exists \mathbf{i} \in \mathbb{N}_0^k\colon \ \ \mathbf{x}^{\mathbf{i}}\mu = \nu \quad \bigl(\mu, \nu \in L_M(I)\bigr).$$

Note that for every ideal $I \lhd \mathcal{R}_k$ we can consider the left $\mathcal{R}_k$-module $L_M(I)$ as a natural left module over the operator ring $S = \mathcal{R}_k/I$ of the ideal I (see Section 11.4). In particular for any $\mu \in L_M(I)$ the product $\mathcal{T}(I)\mu$ is well defined.

PROPOSITION 13.8. *The relation $\sim$ on $L_M(I)$ is an equivalence relation iff $L_M(I)$ is a reversible family. In the last case the relation $\sim$ decomposes $L_M(I)$ into classes of equivalent sequences, and the class $[\mu]_\sim$ of sequences equivalent to $\mu \in L_M(I)$ is $[\mu]_\sim = \mathcal{T}(\mu) = \mathcal{T}(I)\mu$, i.e., $[\mu]_\sim$ is the cycle of μ.*

13.2.1. *The cyclic type of reversible k-LRS-family* Let $\mathcal{H}_k$ be the set of all subgroups $\mathcal{G} < \mathbb{Z}^k$ of rank k such that $\mathcal{G}$ is generated by the set $\mathcal{G}^+$ of its nonnegative vectors. Then $\mathcal{H}_k$ is closed relative to the operation $\cap$ of intersection of subgroups. Thus there is a semigroup $(\mathcal{H}_k, \cap)$.

Denote by $\mathbb{Z}[\mathcal{H}_k]$ the semigroup algebra of a semigroup $\mathcal{H}_k$ over $\mathbb{Z}$.

The *cyclic type* of a finite reversible k-LRS-family $L_M(I)$ is the element Z_I^M of the algebra $\mathbb{Z}[\mathcal{H}_k]$ of the form

$$Z_I^M = \sum_{\mathcal{G}\in\mathcal{H}_k} z_I^M(\mathcal{G})\mathcal{G},$$

where $z_I^M(\mathcal{G})$ is the number of cycles $\mathcal{T}(\mu) \subseteq L_M(I)$ with $\mathfrak{P}(\mu) = \mathcal{G}$.

In the case $k = 1$ the last definition can be simplified. In fact, each subgroup $\mathcal{G} \in \mathcal{H}_1$ is generated by the number $t = [\mathbb{Z} : \mathcal{G}]$, and instead of the cyclic type $Z_I^M = \sum_{t\geqslant 1} z_I^M(\langle t\rangle)\langle t\rangle$ we can consider a cyclic type $C_I^M(y) = \sum_{t\geqslant 1} c_I^M(t)y^t \in \mathbb{Z}_1[y]$, where $c_I^M(t)$ is the number of cycles $\mathcal{T}(\mu) \subseteq L_M(I)$ with $T(\mu) = t$. In this form the notion of cyclic type coincide with the same notion in Section 12.2.2.

The composition of elements $A = \sum_{\mathcal{G}\in\mathcal{H}_k} a_{\mathcal{G}}\mathcal{G}$ and $B = \sum_{\mathcal{G}\in\mathcal{H}_k} b_{\mathcal{G}}\mathcal{G}$ in the ring $\mathbb{Z}[\mathcal{H}_k]$ is the element $A * B = C = \sum_{\mathcal{G}\in\mathcal{H}_k} c_{\mathcal{G}}\mathcal{G}$, where

$$c_{\mathcal{G}} = \sum_{\mathcal{G}_1\cap\mathcal{G}_2=\mathcal{G}} \big[\mathbb{Z}^k : (\mathcal{G}_1 + \mathcal{G}_2)\big] a_{\mathcal{G}_1} b_{\mathcal{G}_2} = \sum_{\mathcal{G}_1\cap\mathcal{G}_2=\mathcal{G}} \frac{[\mathbb{Z}^k : \mathcal{G}_1][\mathbb{Z}^k : \mathcal{G}_2]}{[\mathbb{Z}^k : \mathcal{G}]} \cdot a_{\mathcal{G}_1} b_{\mathcal{G}_2}.$$

THEOREM 13.9. *The algebra* $(\mathbb{Z}[\mathcal{H}_k], +, *)$ *is a commutative ring with unit* $\mathbb{Z}^k$. *If* $B_s = \sum_{\mathcal{G}\in\mathcal{H}_k} b_{\mathcal{G}}^{(s)}\mathcal{G} \in \mathbb{Z}[\mathcal{H}_k]$ *for* $s \in \{1, \dots, r\}$, *then*

$$B_1 * \cdots * B_r = \sum_{\mathcal{G}\in\mathcal{H}_k} \left(\sum_{\mathcal{G}_1\cap\cdots\cap\mathcal{G}_r=\mathcal{G}} \frac{[\mathbb{Z}^k : \mathcal{G}_1]\cdots[\mathbb{Z}^k : \mathcal{G}_r]}{[\mathbb{Z}^k : \mathcal{G}]} \cdot b_{\mathcal{G}_1}^{(1)} \cdots b_{\mathcal{G}_r}^{(\mathrm{r})} \right) \mathcal{G}.$$

The ring $(\mathbb{Z}[\mathcal{H}_k], +, *)$ is called the *ring of cyclic types* of k-LRS-families.

THEOREM 13.10. *Let* $L_M(I)$ *be a finite reversible k-LRS-family. If* $M = M_1 \dot{+} M_2$ *is the direct sum of submodules, then* $Z_I^M = Z_I^{M_1} * Z_I^{M_2}$. *If* $I = I_1 I_2$ *is the product of comaximal ideals, then* $Z_I^M = Z_{I_1}^M * Z_{I_2}^M$.

13.2.2. *Full-cycle and k-maximal ideals and recurrences* Let R be a commutative f.r. and I be a reversible ideal of $\mathcal{R}_k$ with operator ring $S = \mathcal{R}_k/I$. Then S is a f.r.

We call a reversible ideal $I \lhd \mathcal{R}_k$ *full-cycle* if $I \cap R = 0$, $L_R(I)$ is a cyclic S-module, and $\mathcal{T}(I) = S^*$. A k-sequence $u \in R^{\langle k\rangle}$ is called *full-cycle* if $I = \mathrm{An}(u)$ is a full-cycle ideal and $L_R(I) = Su$.

PROPOSITION 13.11. *Let* $I \lhd \mathcal{R}_k$ *be a reversible ideal such that* $I \cap R = 0$, $L_R(I) = Su$. *Then* $\mathfrak{P}(u) = \mathfrak{P}(I)$, *and* I *is a full-cycle ideal iff*

$$\forall v \in L_R(I) \quad \big(\mathfrak{P}(v) = \mathfrak{P}(I)\big) \quad \Rightarrow \quad \big(v \in \mathcal{T}(u)\big).$$

PROPOSITION 13.12. *Let R, Q be finite quasi-Frobenius rings, $R < Q$. Then there exists a full-cycle recurrence u over R such that the ring S of operators of u is isomorphic to Q.*

A k-sequence $u \in R^{\langle k \rangle}$ is called *exact* if $R \cap \operatorname{Ann}(u) = \emptyset$.

An exact reversible k-LRS u over a Galois ring $R = GR(q^n, p^n)$ is called a *k-maximal recurrence* if its operator ring $S = \mathcal{R}_k / \operatorname{An}(u)$ satisfies the conditions

$$S = GR\big(q^{mn}, p^n\big), \quad \text{for some } m \in \mathbb{N}, \quad \text{and} \quad T(u) = |S^*|. \tag{13.6}$$

i.e., if u is a full-cycle recurrence over R, and its operator ring is a Galois ring. If here $S = GR(q^{mn}, p^n)$, then we say that u is a *k-max-LRS of rank m*.

By Theorem 13.6 condition (13.6) is equivalent to the condition

$$S = GR\big(q^{mn}, p^n\big), \qquad S^* = \langle \vartheta_1, \dots, \vartheta_k \rangle, \quad \text{where } \vartheta_s = x_s + I,\ s \in \overline{1, k}. \tag{13.7}$$

The description of such recurrences is based on the notion of trace in a Galois ring (Section 4.4).

THEOREM 13.13. *Let $Q = GR(q^{mn}, p^n)$ be a Galois extension of the Galois ring $R = GR(q^n, p^n)$. Then a k-sequence $u \in R^{\langle k \rangle}$ is a k-max-LRS of rank m over R iff there exist elements $\xi, \alpha_1, \dots, \alpha_k \in Q^*$ such that*

$$u(\mathbf{z}) = \operatorname{Tr}_R^Q\big(\xi \alpha^{\mathbf{z}}\big) = \operatorname{Tr}_R^Q\big(\xi {\alpha_1}^{z_1} \cdot \dots \cdot {\alpha_k}^{z_k}\big), \quad Q^* = \langle \alpha_1, \dots, \alpha_k \rangle. \tag{13.8}$$

Any such LRS has period $T(u) = (q^m - 1)q^{m(n-1)}$.

We can construct a k-max-LRS of rank m over a finite field, i.e. over a Galois ring $R = GR(q^n, p^n)$ with $n = 1$, for any $k, m \in \mathbb{N}$. In the case $n > 1$ this is not true. If $R = GR(q^n, p^n)$, $q = p^r$, $n > 1$, then there exists a k-max-LRS of rank m over R if and only if

$$k \geqslant mr, \quad \text{if } p > 2 \text{ or } p = n = 2;$$
$$k \geqslant mr + 1, \quad \text{if } p = 2 < n.$$

These restrictions on k follow from (13.7) and Theorem 4.12.

We call a reversible ideal $I \lhd \mathcal{R}_k$ a *k-maximal ideal or ideal of maximal period* over a Galois ring R if its operator ring S satisfies (13.7) for some $m \in \mathbb{N}$.

THEOREM 13.14. *Let I be an ideal of maximal period over $R = GR(q^n, p^n)$ and suppose that (13.7) holds. Then the group of periods of I is $\mathfrak{P}(I) = \{\mathbf{t} \in \mathbb{Z}^k \mid \vartheta^{\mathbf{t}} = e\}$, and*

$$T(I) = \big(q^m - 1\big)q^{m(n-1)}, \qquad \mathbb{Z}^k / \mathfrak{P}(I) \cong GR\big(q^{mn}, p^n\big)^*.$$

The cyclic type of the family $L_R(I)$ is given by

$$Z_I^R = 1 \cdot \mathbb{Z}^k + 1 \cdot \mathfrak{P}(I_1) + \dots + 1 \cdot \mathfrak{P}(I_{n-1}) + 1 \cdot \mathfrak{P}(I),$$

where $I_s = I + p^s \mathcal{R}_k$, $s \in \{1, \dots, n-1\}$. Moreover, $\mathcal{T}(I_s) \cong GR(q^{ms}, p^s)^*$.

13.2.3. *Calculation of the cyclic type of a reversible 1-LRS-family [46,68,146]* Let now I be a reversible ideal of the ring $\mathcal{R} = R[x]$. According to Theorem 13.10 the description of the cyclic type of the family $L_M(I)$ is reduced to the case where R is a local ring and I is a primary ideal; we assume this in what follows.

Let I be a primary ideal with main generator $F(x)$ (see Section 11.2.1) and with a generating system $F(x), G_1(x), \ldots, G_t(x)$. Let $\deg F(x) = m$ and

$$\overline{F}(x) = g(x)^k, \tag{13.9}$$

where $g(x)$ is an irreducible polynomial over the field $\overline{R}$,

$$T\bigl(g(x)\bigr) = \tau, \quad p^{a-1} < k \leqslant p^a. \tag{13.10}$$

Then $\operatorname{rad} F(x) = G(x)$ is a distinguished polynomial corresponding to $g(x)$ (see Section 11.2.2), and by (11.15) $T(I) = \tau p^{\beta(I)}$.

For an $(m \times l)$-matrix B over R, let $\mathfrak{L}_M(B)$ be the R-module of all solutions $(\mu_1, \ldots, \mu_m) \in M^m$ of the system of linear equations $(x_1, \ldots, x_m)B = (0, \ldots, 0)$. Write

$$B_s = \bigl(G^{[s]}(S), G_1(S), \ldots, G_t(S)\bigr)_{m\times m(t+1)}, \quad s \geqslant 0, \tag{13.11}$$

where $G^{[s]}(x)$ is the polynomial defined in (11.11), and $S = S(F)$ is the companion matrix of the polynomial $F(x)$.

THEOREM 13.15. *Under the above assumptions, the cyclic type of the family $L_M(I)$ is*

$$C_I^M(y) = y + \sum_{s=0}^{\beta(I)} c_I^M\bigl(\tau p^s\bigr) y^{\tau p^s}, \quad \textit{where}$$

$$c_I^M(\tau) = \frac{1}{\tau}\bigl(\bigl|\mathfrak{L}_M(B_0)\bigr| - 1\bigr),$$

$$c_I^M\bigl(\tau p^s\bigr) = \frac{1}{\tau p^s}\bigl(\bigl|\mathfrak{L}_M(B_s)\bigr| - \bigl|\mathfrak{L}_M(B_{s-1})\bigr|\bigr), \quad s \in \bigl\{1, \ldots, \beta(I)\bigr\}.$$

If ${}_RM$ is a QF-module or $I = \mathcal{R}F(x)$ is a principal ideal, then $c_I^M(\tau p^{\beta(I)}) \neq 0$, i.e., there exists a recurrence $\mu \in L_M(I)$ such that $T(\mu) = T(I)$.

Note that if R is a field then there exists $\mu \in L_M(I)$ such that $\operatorname{An}(\mu) = I$. In general case this statement is not true. However, there is

COROLLARY 13.2. *Let M be any faithful finite module over a commutative f.r. R and let I be a reversible ideal of $\mathcal{R}$. Then the length of any cycle of the family $L_M(I)$ divides the maximum of the lengths of cycles of this family. If M is a QF-module or I is a principal ideal, then there exists an LRS $\mu \in L_M(I)$ with $T(\mu) = T(I)$.*

The last result implies

CONJECTURE. Let M be a faithful finite module over a commutative f.r. R and let I be a reversible ideal of $\mathcal{R}$. Then there exists an LRS $\mu \in L_M(I)$ with $T(\mu) = T(I)$.

Theorem 13.15 is not convenient to use because there is no a good theory which makes it possible to evaluate the number of solutions of a system of linear equations over an arbitrary local ring R [11, 12]. But for a GE-ring R there is such theory.

In what follows, we assume that R is a GE-ring with parameters (5.4).

In this case, if $m \leqslant l$, then each $(m \times l)$-matrix B over R is equivalent to a unique diagonal matrix D in canonical form (see Section 5.1.2):

$$B \sim D = \operatorname{diag}\bigl(\pi^{d_1}, \dots, \pi^{d_m}\bigr), \quad 0 \leqslant d_1 \leqslant \cdots \leqslant d_m \leqslant n.$$

Recall that the parameter $\operatorname{def} B = d_1 + \cdots + d_m$ is called the defect of B.

PROPOSITION 13.16. *Let R be a GE-ring, and let $I \lhd \mathcal{R}$ be a primary reversible ideal with a main generator F satisfying conditions* (13.9), (13.10), *and let the matrices B_s in* (13.11) *satisfy the condition*: $\operatorname{def} B_s = d(s)$ *for* $s \geqslant 0$. *Then the cyclic type of the family $L_R(I)$ is given by*

$$\begin{aligned} C_I^R(y) &= y + \sum_{s=0}^{\beta(I)} c_I^R\bigl(\tau p^s\bigr) y^{\tau p^s}, \quad \textit{where} \\ c_I^R(\tau) &= \frac{1}{\tau}\bigl(q^{d(0)} - 1\bigr), \\ c_I^R\bigl(\tau p^s\bigr) &= \frac{1}{\tau p^s}\bigl(q^{d(s)} - q^{d(s-1)}\bigr), \\ s &\in \bigl\{1, \dots, \beta(I)\bigr\}. \end{aligned} \tag{13.12}$$

Moreover, $c_I^R(\tau p^{\beta(I)}) \neq 0$.

If $I = \mathcal{R}F(x)$ is a principal ideal, there are some simplifications in the calculation of the cyclic type of $L_R(I) = L_R(F)$, using the result of Theorem 12.23 for the case $A = S(F)$.

THEOREM 13.17. *Under the conditions of Proposition 13.16 let $F(x)$ be a primary reversible polynomial over R satisfying conditions* (13.9), (13.10), $G(x) = \operatorname{rad} F(x)$, *and for* $s \geqslant 0$

$$\operatorname{sign} G^{[s]}\bigl(S^{p^s}\bigr) = \bigl[d_1(s), \dots, d_m(s)\bigr], \qquad d(s) = d_1(s) + \cdots + d_m(s). \tag{13.13}$$

Let s_1 be the least $s \in \mathbb{N}_0$ such that

$$\begin{aligned} &d_1(s) > \frac{\varepsilon}{2}, \quad \textit{if } p > 2; \qquad d_1(s) > \varepsilon, \quad \textit{if } p = 2,\ n > \varepsilon; \\ &d_1(s) = n, \quad \textit{if } p = 2,\ n = \varepsilon. \end{aligned} \tag{13.14}$$

Then

$$T(F) = \tau p^{\sigma}, \quad \textit{where } \sigma = s_1 + \left]\frac{n - d_1(s_1)}{\varepsilon}\right[,$$

and the family $L_R(F)$ has the cyclic type

$$C_F^R(y) = y + \sum_{s=0}^{\sigma} c_F^R\bigl(\tau p^s\bigr) y^{\tau p^s},$$

where the coefficients $c_F^R(\tau p^s)$ *are defined by* (13.12). *Moreover, for each* $s > s_1$ *the parameters* $d_j(s)$ *in* (13.13) *can be expressed with the help of* $[d_1(s_1), \dots, d_m(s_1)]$ *in the form*

$$d_j(s) = \min\{n, d_j(s_1) + (s - s_1)\varepsilon\}, \quad j \in \{1, \dots, m\}.$$

The parameter s_1 *satisfies the inequality* $s_1 \leqslant b(d_1(a)) + a + 1$, *where* a *is defined in* (13.10), $b(x) = \max\{0, \log_p \frac{\varepsilon}{x(p-1)}\}$.

13.3. *Linear recurrences of maximal period over a Galois ring*

Let $R = GR(q^n, p^n)$ be a Galois ring of order q^n and of characteristic p^n (where p is a prime, $q = p^r$), and let $F(x) \in \mathcal{R}$ be a monic polynomial of degree m. Then the periods of $F(x)$ and any LRS $u \in L_R(F(x))$ (see Section 11.1) satisfy the inequalities $T(F) \leqslant (q^m - 1)p^{n-1}$.

Denote by $\bar{u}$ and $\overline{F}$ the images respectively of a sequence u and a polynomial F under the natural homomorphism $R \to \overline{R} = R/pR$. If $\deg F(x) = m$, then

$$T(u) \mid T(F), \; T(F) \mid T(\overline{F}(x))p^{n-1} \leqslant (q^m - 1)p^{n-1}.$$

If $T(F) = (q^m - 1)p^{n-1}$, then F is called a *polynomial of maximal period* (*MP-polynomial of degree m*) over R.

If $T(u) = (q^m - 1)p^{n-1}$ then the sequence u is called an *LRS of maximal period* (MP-*recurrence*) *of rank m*.

PROPOSITION 13.18. *An LRS* $u \in L_R(F)$ *is an MP-recurrence of rank m if and only if* $F(x)$ *is an MP-polynomial over* R *and* $\bar{u} \neq \bar{0}$.

The following representation of MP-recurrences over a Galois ring via the trace function (Section 4.4) is a fundamental result in the theory of such recurrences.

THEOREM 13.19. *Let* $R = GR(q^n, p^n)$, *let* $F(x) \in \mathcal{R}$ *be an MP-polynomial of degree m over* R, *and let* $Q = GR(q^{nm}, p^n)$ *be a Galois extension of degree m of the ring* R. *Then* $Q = R[\vartheta]$, *where* $F(\vartheta) = 0$, *and the family* $L_R(F)$ *is exactly the family of all sequences* $u \in R^{\langle 1 \rangle}$ *of the form*

$$u(i) = \mathrm{Tr}_R^Q(\xi\vartheta^i), \quad i \in \mathbb{N}_0, \; \xi \in Q. \tag{13.15}$$

Here $\|u\| = \|u[\{0, \dots, m-1\}]\| = \|\xi\| = \nu \in \{1, \dots, n\}$ (*see* (10.3)) *and if* $\nu < n$, *then* $T(u) = \tau p^{n-\nu-1}$, *where* $\tau = q^m - 1$. *The cyclic type of the family* $L_R(F)$ *is*

$$C_F^R(y) = y + \sum_{s=0}^{n-1} \left(\frac{q^m}{p}\right) y^{\tau p^s}.$$

Below we restrict ourselves to some illustrations only of the contention that linear sequences over a Galois ring are good source of pseudorandom sequences. More detailed information on this subject is contained in [226–229,232,233,238,239].

13.3.1. *Linear complexity of coordinate sequences of MP-recurrence* Let u be an MP-recurrence of period $(q^m - 1)p^{n-1} = \tau p^{n-1}$ with minimal polynomial $G(x)$ of degree m over the ring $R = GR(q^n, p^n)$. Any term $u(i)$ of the sequence u has a standard *p-adic decomposition*:

$$\begin{aligned} u(i) &= u_0(i) + u_1(i)p + \cdots + u_{n-1}(i)p^{n-1}, \\ u_s(i) &= \gamma_s\bigl(u(i)\bigr) \in \Gamma(R), \end{aligned} \tag{13.16}$$

(see Section 4.2). The latter gives us n sequences $u_0, \dots, u_{n-1}$ over the field $\Gamma(R) = GF(q)$. For sufficiently large but acceptable values of m and s the sequence u_s is a good source of pseudorandom numbers. Of course u_s is an LRS over $\Gamma(R)$. Let $\operatorname{rank} u_s$ be the *rank* or *linear complexity* of u_s: the degree of its minimal polynomial. Apparently we can consider u_s as an "approximation" of a random sequence only if $\operatorname{rank} u_s$ is large enough. The simplest estimate of this parameter is the following.

$T(u_s) = \tau p^s$, $s \in \{0, \dots, n-1\}$, and u_0 is an MP-recurrence of rank m over the field Γ,

$$\forall s \in \{1, \dots, n-1\}: \quad \operatorname{rank} u_s > m\bigl(p^{s-1} + 1\bigr).$$

In the case $R = \mathbb{Z}_{p^n}$ instead of the p-adic decomposition (13.16) one can consider the *p-ary decomposition* of $u(i)$:

$$\begin{aligned} u(i) &= v_0(i) + v_1(i)p + \cdots + v_{n-1}(i)p^{n-1}; \\ v_s(i) &= \delta_s\bigl(u(i)\bigr) \in \{0, \dots, p-1\}. \end{aligned} \tag{13.17}$$

Any of the sequences $v_0, \dots, v_{n-1}$ is an LRS over the field $\mathbb{Z}_p$ and $\operatorname{rank} v_0 = m$.

Let p be an odd prime number. If $s_k(p) = 1^k + 2^k + \cdots + (p-1)^k$, $k \geqslant 1$, then

$$\begin{aligned} &p \mid s_k(p), \quad 1 \leqslant k \leqslant p-2, \qquad p^2 \nmid s_1(p), \\ &p^2 \mid s_{2a+1}(p), \quad 3 \leqslant 2a+1 \leqslant p-2. \end{aligned}$$

A pair of numbers $(p, 2a)$, $2 \leqslant 2a \leqslant p-3$, is called *regular* if $p^2 \nmid s_{2a}(p)$ [52].

THEOREM 13.20. *(See [223,226].) The rank of the first p-ary coordinate sequence v_1 of MP-LRS u satisfies the relations*

$$\operatorname{rank} v_1 = m + \binom{m}{2}, \quad \textit{if } p = 2;$$

$$\operatorname{rank} v_1 = m + \sum_{l \in V} \binom{m+l-1}{l} + \binom{m+p-2}{p-1} + \binom{m+p-1}{p},$$

$$\textit{if } p \geqslant 3,$$

where V is the set of odd numbers $l \in \{3, \dots, p-2\}$ such that $(p, p-l)$ is a regular pair.

UPPER BOUNDS ON $\operatorname{rank} v_s$. Let $\left\{ {m \atop r} \right\}$ be number of combinations of r identical balls distributed over m boxes under the extra condition that each box contains $p-1$ balls or less. By [57, p. 215]

$$\left\{ {m \atop r} \right\} = \sum_{j \geqslant 0} (-1)^j \binom{m}{j} \binom{r+m-pj-1}{m-1}, \quad r \geqslant 1.$$

THEOREM 13.21. *(See [226–228].) Let u be an MP-linear recurrence of rank m over the ring $R = \mathbb{Z}_{p^n}$, then* $\operatorname{rank} v_0 = m$, $\operatorname{rank} u_1 = m + \binom{m+p-1}{p}$ *and for $s \geqslant 2$ there is an upper bound*:

$$\operatorname{rank} v_s \leqslant \left\{ {m \atop p^s} \right\} + \left(p^{s-1} + 1\right)m + \sum_{l=1}^{p^{s-1}-1} (l+1) \cdot \sum_{t=b_s(l+1)+1}^{b_s(l)} \left\{ {m \atop t} \right\},$$

where $b_s(l) = p^s - pl + \pi_p(l)$, and

$$\pi_p(l) = \begin{cases} l, & \text{if } l \leqslant p-1; \\ 1 + [(l-1) \bmod (p-1)], & \text{if } l \geqslant p \geqslant 3; \\ 1 + [l \bmod 2], & \text{if } l \geqslant p = 2. \end{cases} \tag{13.18}$$

LOWER BOUNDS ON $\operatorname{rank} v_s$. Below we fix $R = \mathbb{Z}_{p^n}$ and an MP-polynomial $F(x)$, satisfying all conditions of Theorem 13.19. Then for every $s \in \{0, \ldots, n-2\}$ there are the relations

$$x^{\tau p^s} \equiv e + p^{s+1}\Phi_{s+1}(x) \quad \big(\operatorname{mod} F(x)\big), \quad \deg \Phi_{s+1}(x) < m, \quad \overline{\Phi_{s+1}}(x) \neq \bar{0}. \tag{13.19}$$

Let us write $w_s = \overline{\Phi_s(\vartheta)}$, $s \in \{1, \ldots, n-1\}$, and say that a polynomial $F(x)$ *satisfies the* (s,t)*-condition* if $[\overline{R}(w_s) : \overline{R}] = t$ (where, of course, $t \mid m$). We say also that $F(x)$ *satisfies the* $(s,*)$*-condition* if $w_s, w_s^p, \ldots, w_s^{p^{m-1}}$ is a basis of $\overline{Q}_{\overline{R}}$. An algorithm for constructing an MP-polynomial F with a fixed polynomial $\overline{\Phi_s}$ is presented in [226].

THEOREM 13.22. *(See [226–228].) Let u be an MP-linear recurrence over the ring $R = \mathbb{Z}_{p^n}$ with characteristic polynomial $F(x)$. Then*

(a) *if F satisfies* $(2,1)$*-condition, then*

$$\operatorname{rank} v_s \geqslant \sum_{l \in V(2,1,p)} (l+1) \left\{ {m \atop b_s(l)} \right\}, \quad \textit{where}$$

$$V(2,1,p) = \left\{0, \ldots, p^{s-1}\right\}, \quad \text{if } p \geqslant 3,$$

$$V(2,1,p) = \left\{0, \ldots, 2^{s-1}\right\} \cap 2\mathbb{N}, \quad \text{if } p = 2.$$

(b) *if F satisfies* $(2,h)$*-condition, and $h > p$, then*

$$\operatorname{rank} v_s \geqslant m\left(p^{s-1} + 1\right) + \left\{ {m \atop p^s} \right\} + \frac{h-p}{h} \sum_{l \in V(2,h,p)} (1+l)\left(\left\{ {m \atop b_s(l)} \right\} - \delta_k\right), \quad \textit{where}$$

$$V(2,h,p) = \left\{l \in \left\{1, \ldots, p^{s-1} - 1\right\} \colon \pi_p(l) = 1,\ b_s(l) \leqslant m(p-1)\right\},$$

$$\delta_l = \begin{cases} e, & \text{if } m \mid b_s(l),\ \operatorname{tr}_p^{p^m}(w_2) = 0; \\ 0 & \text{otherwise.} \end{cases}$$

(c) *if F satisfies $(1,*)$-condition, then*

$$\operatorname{rank} v_s \geqslant \sum_{l\in V_1(1,*,p)} (l+1)\left\{ \begin{matrix} m \\ b_s(l) \end{matrix} \right\} + \frac{m}{m+p} \sum_{l\in V_2(1,*,p)} (l+1)\left\{ \begin{matrix} m \\ b_s(l) \end{matrix} \right\},$$

where

$$V_1(1,*,p) = \left\{l \in \left\{0,\dots,p^{s-1}\right\}\colon \pi_p(l) = 1,\ \textit{and}\ b_s(l) < m(p-1)\right\},$$
$$V_2(1,*,p) = \left\{l \in \left\{2,\dots,p^{s-1}\right\}\colon \pi_p(l) = 2,\ \textit{and}\ b_s(l) < m(p-1)\right\}.$$

There are also simpler but less precise lower bounds.

THEOREM 13.23. *(See [226–228,239].) For every $m > 1$ there exists an MP-polynomial $F(x) \in \mathbb{Z}_{p^n}[x]$ of degree m such that any MP-linear recurrence $u \in L_{\mathbb{Z}_{p^n}}(F)$ satisfies the following condition: for $s \geqslant 2$*

$$\operatorname{rank} v_s \geqslant \left\{ \begin{matrix} m \\ p^s \end{matrix} \right\} + \left(p^{s-1}+1\right)m + \sum_{k=2}^{s-2}\left(p^{s-1}-p^k+1\right)\left\{ \begin{matrix} m \\ p^{k+1}+p-1 \end{matrix} \right\}.$$

CONJECTURE. If $p = 2$, then the MP-polynomial $F(x)$ can be chosen such that the last sum contains an additional term $\binom{m}{4}2^{s-1}$.

If p, s are fixed and $m \to \infty$, then $\operatorname{rank} u_s = \binom{m}{p^s}(1 + o(1/m))$.

13.3.2. *Frequency characteristics of MP-recurrences* Another important requirement for a pseudo-random sequence is a "uniformity" condition on the distribution of elements and of k-tuples on long enough segments from it. Results on this topic are summarized in [226, 227]. We present here only one of the newer results. Let $F(x)$ be an MP-polynomial of degree m over a Galois ring $R = GR(q^n, p^n)$, and let $u \in L_R(F)$ be an MP-recurrence of period $T = (q^m - 1)p^{n-1}$. Let $0 \leqslant i_1 \leqslant \cdots \leqslant i_k < T$ be fixed integer numbers, and let $a_1, \dots, a_k$ be fixed elements of R. Denote by $\nu_u(\mathbf{a}/\mathbf{i})$ the number of solutions $i \in \{0, \dots, T-1\}$ of the system of equations

$$u(i+i_1) = a_1, \quad \dots, \quad u(i+i_k) = a_k.$$

THEOREM 13.24. *(See [192].) If the system of residues of polynomials $\bar{x}^{i_1}, \dots, \bar{x}^{i_k} \in \overline{\mathcal{R}}$ modulo $\overline{F}(x)$ is linearly independent over the field $\overline{R}$, then*

$$\left|\nu_u(\mathbf{a}/\mathbf{i}) - \frac{T}{|R|^k}\right| \leqslant p^{2(n-1)}q^{m/2} \approx p^{3(n-1)/2}\sqrt{T}.$$

There are more precise results on the distribution of elements on cycles of MP-linear recurring sequences over $R = \mathbb{Z}_{p^n}$. Note that if $F(x) \in \mathcal{R}$ is an MP-polynomial with the parameters pointed out above, then for every $s \in \{1, \dots, n-1\}$ the following relations hold:

$$x^{\tau p^{s-1}} \equiv e + p^s \Phi^{(s)}(x) \quad \left(\operatorname{mod} F(x)\right), \quad \deg \Phi^{(s)}(x) < m,\ \overline{\Phi^{(s)}}(x) \neq \bar{0}.$$

Let $N_u(a) = \nu_u(a/0)$ be the number of solutions $i \in \{0, \dots, T - 1\}$ of the equation $u(i) = a$ for a given $a \in R$.

THEOREM 13.25. *(See [227].) Under the above assumptions,*

(a) *if $p \geqslant 3$, then*

$$\nu_u(a) \geqslant \frac{p-1}{p} \cdot \frac{T}{|R|} \quad \textit{for any } a \in R \textit{ in case } \deg \overline{\Phi^{(1)}}(x) > 0,$$
$$\nu_u(a) = p^{m-1} \quad \textit{for any } a \in R^* \textit{ in case } \deg \overline{\Phi^{(1)}}(x) = 0;$$

(b) *if $p = 2$, then*

$$\nu_u(a) \geqslant \frac{1}{4} \cdot \frac{T}{|R|} \quad \textit{for any } a \in R \textit{ in case } \deg \overline{\Phi^{(2)}}(x) > 0,$$
$$\nu_u(a) \geqslant \frac{1}{2} \cdot \frac{T}{|R|} \quad \textit{for any } a \in R^* \textit{ in case } \deg \overline{\Phi^{(2)}}(x) = 0.$$

Let now $R = GR(q^2, 4)$ be a Galois ring of characteristic 4. Then a full description of the possible variants of distributions of elements on cycles of MP-recurrences can be given.

Let $F(x) \in \mathcal{R}$ be a monic polynomial of degree m such that its period $T = T(F)$ is equal to $\tau = q^m - 1$ (distinguished polynomial) or to 2τ (MP-polynomial). In such a case there is a description of the possible types $[N_u(c)\colon\ c \in R]$ for $u \in L_R(F)$ based on the presentation of an LRS u via the trace-function in Galois rings [226,239,294] and on the theory of quadratic forms over Galois fields of characteristic 2 (see [240]).

Let $\lambda = [m/2]$ be the integer part of $m/2$ and let $\delta_{c,0}$ be the Kronecker delta.

THEOREM 13.26. *(See [230,231,241].)*

(a) *If $F(x)$ is a distinguished polynomial then for any $c \in R$*

$$N_u(c) = q^{m-2} \pm wq^{\lambda-1} - \delta_{c,0},$$

where $w \in \{1,\ q-1\}$ if $m = 2\lambda + 1$, and $w \in \{0,\ 1,\ q-1\}$ if $m = 2\lambda$. There exist not more than $2q + 1$ different types $[N_u(c)\colon\ c \in R]$ in $L_R(F)$.

(b) *If $F(x)$ is an MP-polynomial then for any $c \in R$*

$$N_u(c) = 2q^{m-2} \pm wq^{\lambda-1} - 2\delta_{c,0},$$

where $w \in \{0, 2, q-2, q, 2(q-1)\}$ if $m = 2\lambda+1$, and $w \in \{0, 1, 2, q-1, 2(q-1)\}$ if $m = 2\lambda$. There exist not more than $2q+1$ different types $[N_u(c)\colon\ c \in R]$ in $L_R(F)$.

14. Linear codes over finite rings and modules

14.1. *Main definitions. Parameters of codes [168,267]*

Let M be a finite left module over a finite ring R with identity. The *Hamming weight* of a *word* $\vec{\alpha} = (\alpha_1, \dots, \alpha_n) \in M^n$ is defined as $w_{\mathrm{H}}(\vec{\alpha}) = \|\vec{\alpha}\| = |\{i \in \{1, \dots, n\}\colon\ \alpha_i \neq 0\}|$.

The *(Hamming) distance $d(\vec{\alpha}, \vec{\beta})$ between words $\vec{\alpha}, \vec{\beta} \in M^n$* is defined as

$$d(\vec{\alpha}, \vec{\beta}) = \left|\{i \in \{1, \dots, n\}\colon \alpha_i \neq \beta_i\}\right| = \|\vec{\alpha} - \vec{\beta}\|.$$

Any submodule $\mathcal{K} \leqslant {}_R M^n$, $n \in \mathbb{N}$, is called a *linear n-code* over ${}_R M$.

The *(Hamming) distance of a code $\mathcal{K} \leqslant M^n$* is

$$d(\mathcal{K}) = \min\{d(\vec{\alpha}, \vec{\beta})\colon \vec{\alpha}, \vec{\beta} \in \mathcal{K},\ \vec{\alpha} \neq \vec{\beta}\} = \min\{\|\vec{\alpha}\|\colon \vec{\alpha} \in \mathcal{K} \setminus 0\}.$$

If $d(\mathcal{K}) = d$ we call $\mathcal{K}$ an $(n, C, d)_{|M|}$*-code*, where $C = |\mathcal{K}|$, or $[n, k, d]_{|M|}$*-code*, where $k = \log_{|M|} |\mathcal{K}|$. If $d(\mathcal{K}) = d \geqslant 2t + 1$ the code *notes $d - 1$ and corrects t errors*.

The main problems of the coding theory are as follows: for a given $|M| = q$ and with two of the parameters n, k, d fixed to find the extreme (maximal or minimal) possible value of the third of them. In this connection there are some well-known bounds.

SINGLETON BOUND. $d \leqslant n - k + 1$. A code with $d = n - k + 1$ is called an *MDS-code*. Example: the linear $[q, k, q - k + 1]$ Reed–Solomon codes over $GF(q)$.

PLOTKIN BOUND. $d \leqslant q^{-1}(C - 1)^{-1}(q - 1)Cn$.

Equality holds for the *simplex code* over $GF(q)$: the linear $[q^m - 1, m, q^{m-1}(q - 1)]_q$-code.

SPHERE PACKING (HAMMING) BOUND. If $d \geqslant 2t + 1$, then

$$C \leqslant q^n \Bigg(\sum_{i \in \{0, \dots, t\}} \binom{n}{i} (q - 1)^i \Bigg)^{-1}.$$

If the last inequality is equality a code is called *perfect*. Examples: the linear $[n, n - m, 3]_q$*-Hamming code* over $GF(q)$, $n = (q - 1)^{-1}(q^m - 1)$; the *Golay* $[23, 12, 7]_2$-code and the $[11, 6, 5]_3$-code.

14.2. *General results*

14.2.1. *The socle and the distance of a linear code [163]* If $\mathcal{K} \leqslant {}_R M^n$ then it is possible to simplify the calculation of $d(\mathcal{K})$ using the notion of socle (Section 6.1.2). Recall that the socle $\mathfrak{S}({}_R M)$ is a left module over the top-factor $\overline{R}$ of the ring R.

PROPOSITION 14.1. *The socle $\mathfrak{S}(\mathcal{K})$ of a linear code $\mathcal{K} \leqslant_R M^n$ is a linear code over ${}_{\overline{R}}\mathfrak{S}(M)$:*

$$\mathfrak{S}(\mathcal{K}) = \mathcal{K} \cap \big(\mathfrak{S}(M)\big)^n \leqslant {}_{\overline{R}}\mathfrak{S}(M)^n; \tag{14.1}$$

moreover,

$$d(\mathcal{K}) = d\big(\mathfrak{S}(\mathcal{K})\big). \tag{14.2}$$

According to Theorem 1.1 the ring $\overline{R}$ has a decomposition

$$\overline{R} = \overline{R}_1 \oplus \cdots \oplus \overline{R}_t, \tag{14.3}$$

where $\overline{R}_i = M_{m_i}(P_i)$ is the ring of $m_i \times m_i$-matrices over $P_i = GF(q_i)$ with identity e_i, $i \in \{1, \dots, t\}$. Then $e_1, \dots, e_t$ is a system of pairwise orthogonal central idempotents of the ring $\overline{R}$ and $e_1 + \cdots + e_t = \bar{e}$ is the identity of $\overline{R}$. According to (14.3)

$$\mathfrak{S}(M) = \mathfrak{S}_1 \oplus \cdots \oplus \mathfrak{S}_t, \quad \text{where } \mathfrak{S}_i = e_i\mathfrak{S}(M) = M_{m_i,n_i}(P_i) \tag{14.4}$$

is a space of $m_i \times n_i$-matrices over P_i, $\mathfrak{S}_i$ is a left $\overline{R}_i$-module and $n_i \geqslant 0$ for $i \in \{1, \dots, t\}$.

Under the conditions (14.3), (14.4) the code $\mathfrak{S}(\mathcal{K})$ in (14.2) has a decomposition

$$\mathfrak{S}(\mathcal{K}) = \mathcal{L}_1 \oplus \cdots \oplus \mathcal{L}_t, \quad \text{where } \mathcal{L}_i = e_i\mathfrak{S}(\mathcal{K}) \leqslant {}_{\overline{R}_i}\mathfrak{S}_i^n,\ i \in \{1, \dots, t\}. \tag{14.5}$$

PROPOSITION 14.1. *Under the conditions (14.3), (14.4)*

$$d(\mathcal{K}) = \min\{d(\mathcal{L}_1), \dots, d(\mathcal{L}_t)\}. \tag{14.6}$$

So the calculation of the distance of any linear code over an arbitrary finite module is reduced to the same problem for codes over the modules $\mathfrak{S}_i = M_{m_i,n_i}(P_i)$ with simple coefficient rings $\overline{R}_i = M_{m_i}(P_i)$.

14.2.2. *Dual code over the character module [91,163,233]* Let $M_R^\flat$ be the character module of a finite module ${}_RM$ (Section 6.2.1). We can consider any row $\vec{\varphi} = (\varphi_1, \dots, \varphi_n) \in (M^\flat)^n$, as an element of $(M^n)^\flat$, acting on elements $\vec{\alpha} = (\alpha_1, \dots, \alpha_n) \in M^n$ by the rule

$$\vec{\varphi}(\vec{\alpha}) = \varphi_1(\alpha_1) + \cdots + \varphi_n(\alpha_n) \in \mathbb{Q}/\mathbb{Z}.$$

Then $(M^\flat)^n$ is the character group of M^n: $(M^\flat)^n = (M^n)^\flat$.

For every $\mathcal{K} \leqslant (M^n, +)$ we define its *dual code in Delsartes form*:

$$\mathcal{K}^\perp := \{\vec{\varphi} \in (M^\flat)^n\colon\ \vec{\varphi}(\vec{\alpha}) = 0 \text{ for all } \vec{\alpha} \in \mathcal{K}\}.$$

Then $\mathcal{K}^\perp \leqslant ((M^\flat)^n, +)$, and if $\mathcal{K} \leqslant {}_RM^n$ then $\mathcal{K}^\perp \leqslant (M^\flat)^n_R$, and also $\mathcal{K} \subseteq \mathcal{K}^{\perp\perp}$. Moreover, we have the following generalization of the well-known relations between a linear code over a finite field and its dual code.

PROPOSITION 14.2. *The equality $\mathcal{K}^{\perp\perp} = \mathcal{K}$ holds and there is a group isomorphism $\mathcal{K}^\perp \cong M^n/\mathcal{K}$. In particular $|\mathcal{K}^\perp| \cdot |\mathcal{K}| = |M|^n$.*

Let $\mathcal{K} \leqslant_R M^n$ and let $\vec{g}_1, \dots, \vec{g}_k \in \mathcal{K}$ form a generating set of the module ${}_R\mathcal{K}$. Then the $(k \times n)$-matrix

$$G_{k\times n} = \begin{bmatrix} \vec{g}_1 \\ \vdots \\ \vec{g}_k \end{bmatrix}$$

with entries in M is called the *generating matrix* of $\mathcal{K}$. A generating matrix of the code $\mathcal{K}^{\perp}$ over $M^{\flat}$ is called the *check matrix* of $\mathcal{K}$. Let

$$\varphi_i = (\varphi_{i1}, \ldots, \varphi_{in}) \in \left(M^{\flat}\right)^n, \quad i \in \{1, \ldots, l\},$$

be a generating system of the module ${}_R\mathcal{K}^{\perp}$. Then $\Phi = (\varphi_{ij})_{l\times n}$ is the check matrix of the code $\mathcal{K}$. We can consider Φ as a group homomorphism $\Phi : M^n \to (\mathbb{Q}/\mathbb{Z})^{(l)}$ into the group of all l-columns over $\mathbb{Q}/\mathbb{Z}$, acting on $\vec{\alpha} \in M^n$ by the rule

$$\forall \vec{\alpha} \in M^n\colon \quad \Phi(\vec{\alpha}) = \left(\varphi_1(\vec{\alpha}), \ldots, \varphi_l(\vec{\alpha})\right)^T \in (\mathbb{Q}/\mathbb{Z})^{(l)}.$$

Then, just as for a linear code over a field we have

PROPOSITION 14.3. $\mathcal{K} = \operatorname{Ker} \Phi$.

The Hamming distance of a code $\mathcal{K} \leqslant {}_RM^n$ can be characterized by inspecting a check matrix Φ for $\mathcal{K}$. Any column $\Phi_j^{\downarrow}$, $j \in \{1, \ldots, n\}$, of the matrix Φ is a homomorphism $\Phi_j^{\downarrow} : M \to (\mathbb{Q}/\mathbb{Z})^{(l)}$ according to the rule

$$\forall \alpha \in M^n\colon \quad \Phi_j^{\downarrow}(\alpha) = \left(\varphi_{1j}(\alpha), \ldots, \varphi_{lj}(\alpha)\right)^T \in (\mathbb{Q}/\mathbb{Z})^{(l)}.$$

We say that a system $\Phi_{j_1}^{\downarrow}, \ldots, \Phi_{j_s}^{\downarrow}$ of s columns of Φ is *linearly independent* over M, if

$$\Phi_{j_1}^{\downarrow}(\alpha_1) + \cdots + \Phi_{j_s}^{\downarrow}(\alpha_s) \neq 0^{\downarrow} \quad \text{for any } (\alpha_1, \ldots, \alpha_s) \in M^s \setminus \vec{0}.$$

Let us define the *assured (guaranteed) rank* $\varkappa_M(\Phi)$ of the matrix Φ relatively to M as the maximal $s \in \mathbb{N}$ such that any system $\Phi_{j_1}, \ldots, \Phi_{j_s}$ of s columns of Φ is linearly independent over M. Then there is the following generalization of a well-known classical result.

PROPOSITION 14.4. *Let $\mathcal{K} \leqslant {}_RM^n$ be a linear code with check matrix Φ. Then $d(\mathcal{K}) = \varkappa_M(\Phi) + 1$.*

Note that any linear code over a QF-bimodule ${}_RQ_R$ has a check matrix over R, and certainly the foregoing results hold for these codes and these check matrices.

14.2.3. *MacWilliams identity [163,233]* Let $|M| = m$. For any $s \in M$ and $\vec{\alpha} \in M^n$ write $\sigma_s(\vec{\alpha}) = |\{i \in \{1, \ldots, n\}\colon \alpha_i = s\}|$, and define the *specification* of $\vec{\alpha}$ as the vector

$$\vec{\sigma}(\vec{\alpha}) = \left(\sigma_s(\vec{\alpha})\colon s \in M\right) \in \mathbb{N}_0^M = \mathbb{N}_0^m.$$

Let $\mathbb{Z}[\mathbf{x}]$ be the polynomial ring in m indeterminates $\mathbf{x} = (x_s\colon s \in M)$. For any $\vec{\sigma} \in \mathbb{N}_0^M$ write $\mathbf{x}^{\vec{\sigma}} = \prod_{s\in M} x_s^{\sigma_s}$ and define the *complete weight enumerator* (*c.w.e.*) of $\mathcal{K} \leqslant {}_RM^n$ as:

$$W_{\mathcal{K}}(\mathbf{x}) := \sum_{\vec{\alpha}\in\mathcal{K}} \mathbf{x}^{\vec{\sigma}(\vec{\alpha})} = \sum_{\vec{\alpha}\in\mathcal{K}} \prod_{s\in M} x_s^{\sigma_s(\vec{\alpha})} \in \mathbb{Z}[\mathbf{x}].$$

This can be rewritten as

$$W_{\mathcal{K}}(\mathbf{x}) := \sum_{\vec{\sigma}\in\mathbb{N}_0^M} A_{\mathcal{K}}(\vec{\sigma})\mathbf{x}^{\vec{\sigma}}, \quad \text{where } A_{\mathcal{K}}(\vec{\sigma}) = \left|\left\{\vec{\alpha} \in \mathcal{K}\colon \vec{\sigma}(\vec{\alpha}) = \vec{\sigma}\right\}\right|.$$

For a linear code $\mathcal{K}$ the system of coefficients $A_{\mathcal{K}}(\vec{\sigma})$, $\vec{\sigma} \in \mathbb{N}_0^M$, gives full description of the possibility of the code to correct errors.

Similarly for a linear code $\mathcal{L} \leqslant (M^\flat)^n_R$ over $M^\flat_R$ (of the same cardinality m) and for a vector $\mathbf{y} = (y_\mu\colon \mu \in M^\flat)$ of m indeterminates we have the complete weight enumerator

$$W_{\mathcal{L}}(\mathbf{y}) := \sum_{\vec{\omega}\in\mathcal{L}} \prod_{\mu\in M^\flat} y_\mu^{\sigma_\mu(\vec{\omega})} \in \mathbb{Z}[\mathbf{y}], \quad \sigma_\mu(\vec{\omega}) = \left|\left\{i \in \{1,\dots,n\}\colon \omega_i = \mu\right\}\right|,$$

$$W_{\mathcal{L}}(\mathbf{y}) := \sum_{\vec{\sigma}\in\mathbb{N}_0^M} A_{\mathcal{L}}(\vec{\sigma})\mathbf{y}^{\vec{\sigma}}, \quad \text{where } A_{\mathcal{L}}(\vec{\sigma}) = \left|\left\{\boldsymbol{\omega} \in \mathcal{L}\colon \vec{\sigma}(\boldsymbol{\omega}) = \vec{\sigma}\right\}\right|.$$

The system of coefficients $\{A_{\mathcal{K}^\perp}(\vec{\sigma}),\ \vec{\sigma} \in \mathbb{N}_0^M\}$ is well defined by the system of coefficients $\{A_{\mathcal{K}}(\vec{\sigma}),\ \vec{\sigma} \in \mathbb{N}_0^M\}$.

THEOREM 14.5. *For any $\mathcal{K} \leqslant {}_RM^n$ the weight enumerator $W_{\mathcal{K}^\perp}(\mathbf{y})$ is determined by the relation*

$$W_{\mathcal{K}^\perp}(\mathbf{y}) = \frac{1}{|\mathcal{K}|} W_{\mathcal{K}}(\mathbf{y}A), \quad \textit{where}$$
$$A = (a_{\mu s})_{m\times m}, \quad a_{\mu s} := \exp\bigl(2\pi i\mu(s)\bigr) \textit{ for all } s \in M,\ \mu \in M^\flat.$$

This result is a generalization of the McWilliams theorem for a linear codes over fields [267].

The *Hamming weight enumerator* of a code $\mathcal{K} \leqslant {}_RM^n$ is defined as

$$W^{\mathrm{H}}_{\mathcal{K}}(x,y) := \sum_{\vec{\alpha}\in\mathcal{K}} x^{n-w_{\mathrm{H}}(\vec{\alpha})} y^{w_{\mathrm{H}}(\vec{\alpha})} = W_{\mathcal{K}}(x,y,\dots,y),$$

$$W^{\mathrm{H}}_{\mathcal{K}}(x,y) = \sum_{i\in\overline{0,n}} A_{\mathcal{K}}(i) x^{n-i} y^i, \quad A_{\mathcal{K}}(i) = \left|\left\{\vec{\alpha} \in \mathcal{K}\colon w_{\mathrm{H}}(\vec{\alpha}) = i\right\}\right|.$$

COROLLARY 14.6. *(See [91].) The Hamming weight enumerator of $\mathcal{K}^\perp$ satisfies the equality*

$$W^{\mathrm{H}}_{\mathcal{K}^\perp}(x,y) = \frac{1}{|\mathcal{K}|} W^{\mathrm{H}}_{\mathcal{K}}\bigl(x + (m-1)y, x - y\bigr).$$

14.2.4. *Equivalence of linear codes* We start from the following classical and well-known result of A.A. Markov Jr. Let M be a finite alphabet. We call a bijection $\varphi : M^n \to M^n$ an *isometry relative to the Hamming metric*, or a d_{H}*-isometry*, if

$$\forall \mathbf{x}, \mathbf{y} \in M^n\colon \quad d_{\mathrm{H}}\bigl(\varphi(\mathbf{x}), \varphi(\mathbf{y})\bigr) = d_{\mathrm{H}}(\mathbf{x}, \mathbf{y}).$$

We say that $\varphi : M^n \to M^n$ is *monomial*, if there exist a permutation $\sigma \in S_n$ and $\beta_1, \ldots, \beta_n \in S(M)$ such that

$$\forall \vec{\alpha} = (\alpha_1, \ldots, \alpha_n) \in M^n: \quad \varphi(\vec{\alpha}) = \bigl(\beta_1(\alpha_{\sigma(1)}), \ldots, \beta_n(\alpha_{\sigma(n)})\bigr). \tag{14.7}$$

It is evident that any monomial bijection is a Hamming isometry. But the following is not so evident.

THEOREM 14.2. *(A.A. Markov Jr.,* 1956*). Any Hamming isometry $\varphi : M^n \to M^n$ is monomial.*

Below the case when M is a left A-module and φ is a linear transformation is important.

COROLLARY 14.1. *Let ${}_AM$ be a finite module and $B = \mathrm{End}({}_AM)$. A linear transformation $\psi : {}_AM^n \to {}_AM^n$ is a Hamming isometry exactly if there exist a permutation $\sigma \in S_n$ and elements $b_1, \ldots, b_n \in B^* = \mathrm{Aut}({}_AM)$ such that*

$$\forall \vec{\alpha} = (\alpha_1, \ldots, \alpha_n) \in M^n: \quad \psi(\vec{\alpha}) = \bigl(\alpha_{\sigma(1)} b_1, \ldots, \alpha_{\sigma(n)} b_n\bigr). \tag{14.8}$$

Under condition (14.8) we shall say that ψ is a *monomial* (*over the ring B*) *linear transformation* of the module ${}_AM^n$. Evidently, for any such ψ and any $\mathcal{K} \leqslant {}_AM^n$ the code $\mathcal{L} = \psi(\mathcal{K})$ has the same Hamming weight enumerator as $\mathcal{K}$. The codes $\mathcal{L}$ and $\mathcal{K}$ are called (*linearly*) *equivalent*.

Let us say that two linear codes $\mathcal{K}, \mathcal{L} \leqslant {}_AM^n$ are (*linearly*) *isometric* if there exists a module isomorphism:

$$\tau : {}_A\mathcal{K} \to {}_A\mathcal{L} \tag{14.9}$$

such that

$$\forall \vec{\alpha}, \vec{\beta} \in \mathcal{K}: \quad d_{\mathrm{H}}\bigl(\tau(\vec{\alpha}), \tau(\vec{\beta})\bigr) = d_{\mathrm{H}}(\vec{\alpha}, \vec{\beta}). \tag{14.10}$$

In the light of these definitions the following questions are interesting. Are any two isometric linear codes equivalent? Is any linear isometry (14.9) monomial, i.e. can it be extended to a linear isometry (monomial transformation) (14.8) on M^n? The first classical result in this direction is the MacWilliams extension theorem:

THEOREM 14.7. *(See* [266]*.) Let $P = GF(q)$, $\mathcal{K} \leqslant P_P^n$ and let $\tau : \mathcal{K}_P \to P_P^n$ be a linear* Hamming isometric (isometric imbedding), *i.e.* $w_{\mathrm{H}}(\tau(\vec{\alpha})) = w_{\mathrm{H}}(\vec{\alpha})$ *for all $\vec{\alpha} \in \mathcal{K}$. Then τ can be extended to a* linear monomial transformation $\psi : P_P^n \to P_P^n$:

$$\forall \vec{\alpha} = (a_1, \ldots, a_n) \in P^n: \quad \psi(\vec{\alpha}) = (a_{\sigma(1)} u_1, \ldots, a_{\sigma(n)} u_n)$$

for some fixed permutation $\sigma \in S_n$ and $u_1, \ldots, u_n \in P^*$.

This theorem is the basis of the notion of equivalence for classical algebraic coding theory and has been extended to the ring-linear context in different ways [78,162,196,398–400]). The following theorem is a generalization of all these results to linear codes over QF-bimodules.

THEOREM 14.8. *(See [163].) Let ${}_AM_B$ be a finite QF-bimodule, such that $\mathfrak{S}({}_AM)$ is a cyclic A-module. Then for any $\mathcal{K} \leqslant {}_AM^n$ a linear mapping ${}_A\mathcal{K} \xrightarrow{\tau} {}_AM^n$ is a Hamming isometry (isometric imbedding) exactly if $\tau = \psi \mid \mathcal{K}$ is the restriction to $\mathcal{K}$ of a (right) monomial transformation $\psi : {}_AM^n \to {}_AM^n$ over B:*

$$\forall \vec{\alpha} = (\alpha_1, \ldots, \alpha_n) \in M^n : \quad \psi(\vec{\alpha}) = (\alpha_{\sigma(1)} b_1, \ldots, \alpha_{\sigma(n)} b_n)$$

for some fixed permutation $\sigma \in S_n$ and $b_1, \ldots, b_n \in B^*$.

This theorem is true for example if A is any f.r. with the identity and ${}_AM_B = {}_AA_A^\flat$ (see Theorem 6.16).

14.2.5. *Comparison of codes over fields, spaces and modules* As before ${}_RM$ is a finite module.

PROPOSITION 14.3. *(See [300].) Let R be a commutative local f.r. If there exists a linear $[n, k, d]$-code $\mathcal{L}$ over the field $\overline{R}$, then there exists a linear $[n, k, d]$-code $\mathcal{K}$ over ${}_RM$ with parity-check matrix over R. In particular if $\mathcal{L}$ is an MDS-code, then so is the code $\mathcal{K}$. If $(n, |\overline{R}|) = 1$ and $\mathcal{L}$ is a cyclic code, then $\mathcal{K}$ can be chosen to be cyclic also.*

This result generalizes some partial results from the papers [46,47,345,359,360].

Let L be an elementary Abelian p-group of order $q = p^t$, i.e. a finite linear space over $GF(p)$. If $t > 1$ then there exist linear codes over L which are better than linear codes over $GF(q)$. Let $B_L(n, 3)$ (respectively $B_q(n, 3)$) be the maximum of the cardinalities of linear n-codes over $(L, +)$ (respectively over $GF(q)$) with the distance 3.

PROPOSITION 14.4. *(See [304].) If $p^{\delta-1}(q-1)^{-1}(q^r-1) < n \leqslant p^\delta(q-1)^{-1}(q^r-1) - (p^\delta - 1)$ for some $k \geqslant 2$ and $\delta \in \{1, \ldots, t-1\}$, then $B_L(n, 3) = p^{t-\delta} B_q(n, 3)$.*

This is a generalization of an earlier result from [177] for some cases when $p = 2$. Attempts to construct linear codes over modules which are better than linear codes over linear spaces unexpectedly failed.

We say that a n-code $\mathcal{K}$ over M is *majored* by a n-code $\mathcal{L}$ over some alphabet L if $|L| = |M|$, $|\mathcal{L}| \geqslant |\mathcal{K}|$ and $d(\mathcal{L}) \geqslant d(\mathcal{K})$.

THEOREM 14.5. *(See [300].) Let ${}_RM$ be a module over a commutative local f.r. R, and let ${}_{\overline{R}}L$ be a linear space of cardinality $|L| = |M|$ over the residue field $\overline{R} = R/\mathfrak{N}$. Then any linear code $\mathcal{K} < {}_RM^n$ is majored by some linear code $\mathcal{L} < {}_{\overline{R}}L^n$. If M is a finite Abelian group and L is a direct sum of elementary Abelian groups of cardinality $|L| = |M|$, then any linear n-code over M is majored by some linear n-code over L.*

14.3. *Linear presentation of codes*

The investigation of linear codes over modules is not so important for the construction of codes which are better than codes over fields as for the description of new linear representations of these codes. The representation of codes over fields by linear codes over modules is closely related with the following notions.

14.3.1. *Homogeneous and egalitarian weights [163,169,180,232]* Let ${}_RM$ be a faithful finite module over a f.r. R. A function $w: M \to \mathbb{R}$ is called a *weight* if:

(W1) $\forall x \in M$: $w(x) \geqslant 0$, $w(x) = 0 \Leftrightarrow x = 0$;

(W2) $\forall x \in M$: $w(x) = w(-x)$*;*

(W3) $\forall x, y \in M$: $w(x + y) \leqslant w(x) + w(y)$.

For any weight $w: M \to \mathbb{R}$ the function $\rho_w(x, y) = w(x - y)$ defines a translation-invariant metric on M, i.e. a metric with the property $\rho_w(x + z, y + z) = \rho_w(x, y)$. Every translation-invariant metric ρ on M arises in this way from the weight $w_\rho(x) = \rho(x, 0)$.

We call a function $w: M \to \mathbb{R}$ *egalitarian*, if:

(H1) there exists $\zeta \in \mathbb{R}$ such that $\sum_{x \in U} w(x) = \zeta \cdot |U|$ for any nonzero submodule $U \leqslant M$. This function is called *homogeneous* if in addition

(H2) $\forall x \in M, \forall u \in R^*$: $w(x) = w(ux)$.

A module ${}_RM$ is called *weighted* if it admits an egalitarian weight w. In this case it admits also a homogeneous weight: $w^*(x) = |R^*|^{-1} \cdot \sum_{u \in R^*} w(ux)$. Note that the Hamming weight $w = w_{\text{Ham}}$ on ${}_RM$ is homogeneous if and only if the module ${}_RM$ is simple.

In [77] the following motivation for introducing the egalitarity axiom (H1) was given. For an arbitrary weight w on ${}_RM$ and $n \in \mathbb{N}$ the weight $w^n : {}_RM^n \to \mathbb{R}$ defined by $w^n(\mathbf{x}) = w(x_1) + \cdots + w(x_n)$ for $\mathbf{x} = (x_1, \ldots, x_n) \in M^n$ turns M^n into a translation-invariant metric space. Let now $\mathcal{K}$ be a linear code over ${}_RM$, i.e. a submodule of ${}_RM^n$. Then the projection $\mathcal{K}_i$ of $\mathcal{K}$ onto the i-th coordinate is a submodule of ${}_RM$. For information-theoretic purposes it is natural to require that $\mathcal{K}_i \neq 0$ for every i (i.e., $\mathcal{K}$ is a *full-length code*) and that the numbers $W_i = \sum_{x \in \mathcal{K}} w(x_i)$ satisfy the condition $W_1 = W_2 = \cdots = W_n$ The second condition holds for full-length linear codes over fields. It is satisfied by every full-length linear code over ${}_RM$ if and only if w satisfies (H1).

THEOREM 14.6. *Under conditions (14.3), (14.4) a module ${}_RM$ is weighted if and only if $\mathfrak{S}(M)$ is a cyclic $\overline{R}$-module (i.e. $n_j \leqslant m_j$ for $j \in \{1, \ldots, t\}$) and the decomposition (14.4) does not contain $GF(2) \oplus GF(2)$ or $GF(2) \oplus GF(3)$ as a direct summand.*

COROLLARY 14.2.

(a) *A finite Abelian group of order m is weighted if and only if it is cyclic and $m \not\equiv 0 \pmod 6$.*

(b) *A faithful module ${}_RM$ over a finite commutative local ring is weighted if and only if it is a QF-module.*

Corollary 14.2(a) implies that the Constantinescu—Heise criterion [77] is true not only for cyclic groups but for all finite Abelian groups.

THEOREM 14.7. *For a finite ring R both modules ${}_RR$ and R_R are weighted if and only if R is a Frobenius ring (Section 6.3) and the decomposition (14.3) does not contain $GF(2) \oplus GF(2)$ or $GF(2) \oplus GF(3)$. In this case the left and right homogeneous weights on R coincide.*

We denote by $\mathcal{F}_R$ the class of all finite left R-modules and define the *Euler function* $\mathcal{E}_R : \mathcal{F}_R \to \mathbb{N}$ as $\mathcal{E}_R(M) = |\{x \in M\colon\ M = Rx\}|$ and the *Möbius function* $\mu : \mathcal{F}_R \to \mathbb{Z}$ by the recursion formulae: $\mu(0) = 1$, and if $M \in \mathcal{F}_R \setminus 0$ then $\sum_{U \leqslant_R M} \mu(U) = 0$. In particularly if ${}_RM$ is an irreducible module then $\mu(M) = -1$, $\mathcal{E}_R(M) = |M| - 1$.

THEOREM 14.8. *For a weighted module ${}_RM$ there exists the unique homogeneous weight $w_h(x)$ such that $\sum_{x \in U} w_h(x) = |U|$ for any nonzero submodule $U \leqslant {}_RM$. This weight is given by*

$$w_h(x) = 1 - \frac{\mu_R(Rx)}{\mathcal{E}_R(Rx)} \quad \textit{for all } x \in M. \tag{14.11}$$

14.3.2. *Scaled isometries and presentation of codes* We describe here a rather general technique, based on the concept of a scaled isometry, which yields constructions of presentations of linear codes over weighted modules. We also give some examples of efficient applications of this technique.

For a weighted module ${}_RM \in \mathcal{F}_R$ fix some egalitarian weight w_R. It is extended to $w_R^n : M^n \to \mathbb{R}$ by setting $w_R^n(\mathbf{x}) = \sum_{i=1}^n w_R(x_i)$, and generates a metric $\rho_R^n(\mathbf{x}, \mathbf{y}) = w_h^n(\mathbf{x} - \mathbf{y})$ on M^n.

Let now ${}_SN$ be another weighted module over some ring S with an egalitarian weight w_S. Suppose that for some $d \in \mathbb{N}$ and $\zeta \in \mathbb{R}$ there exists a mapping $\sigma : M \to N^d$ such that

$$\forall a, b \in M\colon \quad \rho_S^d\bigl(\sigma(a), \sigma(b)\bigr) = \zeta \rho_R(a, b).$$

Then σ is called *scaled isometry* (with *scale factor* ζ). It induces for every $n \in \mathbb{N}$ a scaled isometry $\sigma^n : (M^n, \rho_R^n) \to (N^{dn}, \rho_S^{nd})$ with the same scale factor. With every code $C \subseteq M^n$ we associate the code $C' = \sigma^n(C) \subseteq N^{nd}$ and call C' a *σ-representation of the code C*. Note that if C is distance invariant (relative to the metric ρ_R^n) then so is C' (relative to the metric ρ_S^{nd}). If C is a linear code over ${}_RM$, we call C' a *σ-linear code* (and sometimes briefly an *${}_RM$-linear code*). An ${}_RM$-linear code C' is distance invariant but may be nonlinear. This approach allows to construct some new good codes and to find new compact representations of some well-known codes.

In [166] an isometry between $(\mathbb{Z}_4, \rho_{\mathbb{Z}_4})$ and $(\mathbb{F}_2^2, \rho_{\mathrm{Ham}})$ was rediscovered (the so-called Gray mapping $\Phi : \mathbb{Z}_4 \to \mathbb{F}_2^2$, $0 \mapsto 00$, $1 \mapsto 01$, $2 \mapsto 11$, $3 \mapsto 10$), and the term $\mathbb{Z}_4$-*linear code* was introduced for what we call a Φ-linear code. This approach allows one to repeat the proof of the $\mathbb{Z}_4$-linearity of the binary Kerdock code [293,294] and to prove the $\mathbb{Z}_4$-linearity of the Preparata, Delsarte–Goethals and some other codes. A more general form of this mapping for Galois rings is given in [231,302].

See also [237].

14.3.3. *A generalized Kerdock code over $GF(2^l)$* Let $R = GR(q^2, 4)$ be the Galois ring of characteristic 4 and cardinality q^2, $q = 2^l$, $l \geqslant 1$, with coordinate field $\Gamma(R) = GF(q)$ (see Sections 4.1, 4.2). A generalized Kerdock code $K_q(m+1)$ over $\Gamma(R)$ (m is odd) is a Reed–Solomon presentation of the so called base linear code $\mathcal{K}_R(m)$ over R.

Let $S = GR(q^{2m}, 4)$ be the extension of degree m of the ring R and ϑ be a primitive element of the coordinate field $\Gamma(S)$. The *base code* $\mathcal{K}_R(m)$ is defined as the linear code of length $h = q^m$ over R consisting of all words $\vec{v} = (v(0) \cdots v(h-1))$ such that for some $\xi \in S$, $c \in R$

$$v(i) = \mathrm{Tr}_R^S\bigl(\xi \vartheta^i\bigr) + c, \quad i = \{0, \dots, h-2\}, \qquad v(h-1) = c, \tag{14.12}$$

where $\mathrm{Tr}_R^S(x)$ is the *trace-function* from S onto R (Section 4.4).

Let now $\Gamma(R) = \{\omega_0 = 0,\ \omega_1 = e, \dots, \omega_{q-1}\}$. Define $\gamma_* : R \to \Gamma(R)^q$ for an element $r \in R$ of the form $r = r_0 + 2r_1$, $r_s \in \Gamma(R)$, as $\gamma_*(r) = (r_1, r_1 \oplus \omega_1 r_0, \dots, r_1 \oplus \omega_{q-1} r_0)$. Then $\gamma_*(R)$ is a Reed–Solomon $[q, 2, q-1]_q$-code over $\Gamma(R)$ and therefore the mapping γ_* is called the *RS-mapping* [240,302]. It is easy to see that $\gamma_*(R)$ is a scaled isometry of the space R with homogeneous weight into the Hamming space $\Gamma(R)^q$. The code $K_q(m+1)$ is a concatenation of the code $\mathcal{K}_R(m)$ (linear over R) and the code $\gamma_*(R)$ (linear over $\Gamma(R)$). It is the code of length $n = q^{m+1}$ consisting of all words $\gamma_*^h(\vec{u}) = (\gamma_*(u(0)), \dots, \gamma_*(u(h-1)))$, $\vec{u} \in \mathcal{K}_R(m)$. If $q = 2$, i.e. $R = \mathbb{Z}_4$, this code is the original binary Kerdock code [267].

THEOREM 14.9. *(See [231,240,241].) Let $m = 2\lambda + 1 \geqslant 3$. Then the code $K_q(m+1)$ is an R-linear $(n, n^2, \frac{q-1}{q}(n - \sqrt{n}\,))_q$-code with complete weight enumerator*

$$\begin{aligned}
&W_{K_q(m+1)}(x_0, \dots, x_{q-1}) \\
&\quad = \sum_{j=0}^{q-1} x_j^n + \bigl(q^{m+2} - q\bigr) \prod_{j=0}^{q-1} x_j^{n/q} \\
&\qquad + \frac{1}{2} q\bigl(q^m - 1\bigr)\bigl(q^m + q^{\lambda+1}\bigr) \prod_{j=0}^{q-1} x_j^{\frac{n}{q} - q^\lambda} \sum_{j=0}^{q-1} x_j^{q^{\lambda+1}} \\
&\qquad + \frac{1}{2} q\bigl(q^m - 1\bigr)\bigl(q^m - q^{\lambda+1}\bigr) \prod_{j=0}^{q-1} x_j^{\frac{n}{q} + q^\lambda} \sum_{j=0}^{q-1} x_j^{-q^{\lambda+1}}.
\end{aligned}$$

The paper [241] contains also a description of the c.w.e. $W_{\mathcal{K}_R(m)}$ of the base linear code. In particular for $R = \mathbb{Z}_4$ and $m = 2\lambda + 1 \geqslant 3$ we have

$$\begin{aligned}
&W_{\mathcal{K}_{\mathbb{Z}_4}(m)}(x_0, \dots, x_3) \\
&\quad = \sum_{r=0}^{3} x_r^{2^m} + 2 \cdot \bigl(2^m - 1\bigr)\bigl((x_0 x_2)^{2^{m-1}} + (x_1 x_3)^{2^{m-1}}\bigr) \\
&\qquad + 2^m \bigl(2^m - 1\bigr)(x_0 x_1 x_2 x_3)^{2^{m-2}} \sum_{r=0}^{3} \bigl(x_r^{-1} x_{r+1}^{-1} x_{r+2} x_{r+3}\bigr)^{2^{\lambda-1}}.
\end{aligned} \tag{14.13}$$

14.3.4. *Presentations of the extended binary Golay code [169,180]* The binary Golay code can be presented as a linear code over the ring $R = \mathbb{F}_2 \oplus \mathbb{F}_4$. Note that smaller rings of such form ($\mathbb{F}_2 \oplus \mathbb{F}_2$, $\mathbb{F}_2 \oplus \mathbb{F}_3$) are not weighted.

PROPOSITION 14.10. *Let $e = e_1 + e_2$ be the corresponding decomposition of the identity of R, and $\mathbb{F}_4 = \mathbb{F}_2(\alpha)$. Let $\sigma : R \to \mathbb{F}_2^3$ be the $\mathbb{F}_2$-linear map defined by $e_1 \mapsto 111$, $e_2 \mapsto 110$, $\alpha \mapsto 011$ and let $\mathcal{K} \leqslant {}_R M^8$ be the R-linear code with parity-check matrix*

$$\begin{bmatrix} e_1 & e_2 & e & e & e & 0 & 0 & 0 \\ e & e_1 & e_2 & e & 0 & e & 0 & 0 \\ e & e & e_1 & e_2 & 0 & 0 & e & 0 \\ e_2 & e & e & e_1 & 0 & 0 & 0 & e \end{bmatrix}. \tag{14.14}$$

Then the mapping σ is a scaled isometry from (R, ρ_{hom}) onto $(\mathbb{F}_2^3, \rho_{\text{Ham}}^3)$ with scale factor $\frac{3}{2}$, and the code $\sigma^8(\mathcal{K})$ is the linear binary (*Golay*) $[24, 12, 8]$-*code.*

Another isometric representation of the same code is based on some egalitarian, but not homogeneous, weight. Let $R = \mathbb{F}_2[x]/(x^4) = \mathbb{F}_2[z]$, where $z = x + (x^4)$ is the image of x in $\mathbb{F}_2[x]/(x^4)$. Every $a \in R$ has the unique representation $a = a_0 + a_1 z + a_2 z^2 + a_3 z^3$ with $a_j \in \mathbb{F}_2$. Define $\tau : R \to \mathbb{F}_2^4$ and $w_R : R \to \mathbb{R}$ by setting $\tau(a) = (a_0 + a_1 + a_2 + a_3, a_1 + a_3, a_2 + a_3, a_3)$ and $w_R(a) = w_{\text{H}}(\tau(a))$. The function w_R is an egalitarian weight on ${}_R R$ and σ is a scaled isometry $(R, w_R) \to (\mathbb{F}_2^4, w_{\text{H}})$.

PROPOSITION 14.11. *Let $\mathcal{K} \leqslant {}_R R^6$ be the linear code with parity-check matrix*

$$\begin{bmatrix} 1 & 0 & 0 & v & z & z \\ 0 & 1 & 0 & z & v & z \\ 0 & 0 & 1 & z & z & v \end{bmatrix}, \tag{14.15}$$

where $v = 1 + z^3$. The code $\tau^6(\mathcal{K})$ is the linear (*Golay*) $[24, 12, 8]$-*code over* $\mathbb{F}_2$.

14.3.5. *Presentation of the ternary Golay code [169]* Let $R = \mathbb{F}_3[x]/(x^3) = \mathbb{F}_3[z]$ with $z = x + (x^3)$. Then any $a \in R$ has a unique representation $a = a_0 + a_1 z + a_2 z^2$ with $a_j \in \mathbb{F}_3$. Let now $\sigma : R \to \mathbb{F}_3^3$ be defined by $\sigma(a) = (a_0 - a_1 + a_2, a_1 + a_2, a_2)$. Then $w_R : R \to \mathbb{R}$, defined by $w_R(a) = w_{\text{H}}^3(\sigma(a))$, is an egalitarian weight on ${}_R R$. So σ gives a scaled isometry $(R, w_R) \to (\mathbb{F}_3^3, w_{\text{H}})$.

PROPOSITION 14.12. *Let $\mathcal{K} \leqslant {}_R R^4$ be the linear code with parity-check matrix*

$$\begin{bmatrix} 1 & 0 & v & v^2 \\ 0 & 1 & v^2 & -v \end{bmatrix}, \tag{14.16}$$

where $v = 1 + z^2$. The code $\sigma^4(\mathcal{K})$ is a linear (*Golay*) $[12, 6, 6]$-*code over* $\mathbb{F}_3$.

14.3.6. *Scaled isometry over a commutative QF-ring [169,180]* Let now R be a finite commutative local QF-ring with socle $\mathfrak{S}(R) = S$ (Section 6.4). We construct a scaled isometry from the weighted R-module ${}_R R$ into a suitable Hamming space $\mathbb{F}_q^n$. There exists a system of elements $\pi_0, \ldots, \pi_l \in R$ such that π_l is a generator of S and every $x \in R$ has a unique representation

$$x = a_0 \pi_0 + \cdots + a_l \pi_l \quad \text{with } a_i \in \Gamma(R) \text{ for } i \in \{0, \ldots, l\}. \tag{14.17}$$

Consider the $((l+1)\times q^l)$-matrix G over $\Gamma(R)$ whose columns are all the vectors $(g_0,\dots,g_{l-1},1)$, $(g_0,\dots,g_{l-1})\in\Gamma(R)^l$, in some fixed order. The q-ary linear over $\Gamma(R)$ code C with generator matrix G is a generalized Reed–Muller $[q^l,l+1,q^l-q^{l-1}]$-code, cf. [168, Ch. 9.5]. It is a two-weight code with nonzero weights q^l-q^{l-1} and q^l.

PROPOSITION 14.13. *The mapping* $\sigma : R\to\Gamma(R)^{q^l}$, *defined by*

$$\sigma(x)=(a_0,a_1,\dots,a_l)\cdot G,$$

is a scaled isometry with scale factor q^l-q^{l-1} *from* (R,ρ_h) *onto* (C,d_{H}).

Some particular cases of this result can be found in [77,226,302]. For a generalization to arbitrary finite commutative local rings (using the notion of a quasi-Frobenius module) we refer to [180]. For linearly representable codes over chain rings see also [178].

14.4. *Loop codes [87]*

Let R be a f.r., let $G=\{g_1,\dots,g_n\}$ be a finite loop and let $I\leqslant{}_AA$ be a left ideal of the loop ring $A=RG$. Then the set $\mathcal{K}(I)$ of all words $(r_1,\dots,r_n)\in R^n$ such that $\sum r_ig_i\in I$ is a linear n-code over the ring R (a submodule of the module ${}_RR^n$). We shall call any code $\mathcal{K}$ of the above form a *loop-code* or a *G-code*.

There exist a large number of results about codes of such type for the case when $R=\mathbb{F}_q$ is a finite field and G is an Abelian, par excellence cyclic, group, see e.g. [168,266]. For non-Abelian groups there are results of [334–336] where mainly the ideals of the semisimple $\mathbb{F}_q$-algebras $\mathbb{F}_qG$, $((q,|G|)=1)$ were considered.

The paper [87] contains the computational results for parameters of codes $\mathcal{K}(I)$ where I spans the left ideals of loop-algebras $\mathbb{F}G$ for the fields $\mathbb{F}\in\{\mathbb{F}_2,\mathbb{F}_3,\mathbb{F}_4,\mathbb{F}_5\}$ and for all loops G of orders $\{4,\dots,7\}$. Separately, parameters of the group-codes for the dihedral groups $\mathcal{D}_m$, $m=4,5,6$; the group of quaternions Q_8; the group A_4 of even permutations and all Abelian groups of the same orders are presented.

Any loop-algebra $A=\mathbb{F}_qG$ contains two *trivial* MDS-codes: the $[n,1,n]$-code $\mathcal{K}(I_0)$ corresponding to the ideal $I_0=A(\sum_{g\in G}g)=\mathbb{F}_q(\sum_{g\in G}g)$, and the $[n,n-1,2]$-code $\mathcal{K}(\Delta)$ corresponding to the fundamental ideal $\Delta\lhd A$ which is the left and right annihilator of the ideal I_0. However some of the investigated algebras contain nontrivial MDS-codes. For example the algebra $\mathbb{F}_5S_3$ contains a $[6,3,4]$-code (see below).

Following [168] we shall call a (generally nonlinear) $[n,k,d]$-code $C\subseteq\mathbb{F}_q^n$ *optimal* if $|C|=q^k$ is the maximum of all possible cardinalities of n-codes with the distance d. In accord with to this definition we shall call a linear $[n,k,d]_q$-code a *linearly optimal* if k is the maximum of the dimensions of the linear over the field $\mathbb{F}_q$ n-codes with distance d. It is evident that any MDS-code is optimal.

Let us denote by $n(k,q)$ (respectively $m(k,q)$) the maximum of the lengths of MDS-codes C of combinatorial dimension $k=\log_q|C|$ over an alphabet of q elements (respectively for a primary q, of linear MDS codes over the field $\mathbb{F}_q$). Of course $m(k,q)\leqslant n(k,q)$.

PROPOSITION 14.14. *Let n, k be natural numbers such that $n > m(k+1, q)$ for a given primary q. Then any $[n, k, n-k]_q$-code that is linear over $\mathbb{F}_q$ is linearly optimal.*

Group codes satisfying the conditions of the last proposition really exist. For example, each of algebras $\mathbb{F}_q\mathcal{D}_4, \mathbb{F}_q Q_8, q \in \{2, \dots, 5\}$ contains an $[8, 4, 4]_q$-code. All these codes are linearly optimal. We announce the following result about first series of such codes (E. Couselo, S. Gonzáles, V. Markov, C. Martínez, A. Nechaev, 2007, unpublished).

THEOREM 14.15. *Let G be a group of order $2q$, containing an elementary Abelian p-group of order $q = p^l > 2$. Then there exists a $[2q, 2q-3, 3]_q$ linearly optimal G-code.*

THEOREM 14.16. *Let $q = p^l > 2$ be a primary number such that $3 \mid q-1$. Then for some groups G of order $3q$, containing an elementary Abelian p-group of order q there exists a $[3q, 3q-3, 3]_q$ linearly optimal G-code. In particular this holds if either G is commutative or $3 \nmid p-1$.*

Examples for the preceding theorems are $[6, 3, 3]_3$-codes in $\mathbb{F}_3\mathbb{Z}_6$ and in $\mathbb{F}_3 S_3$, $[8, 5, 3]_4$-codes in $\mathbb{F}_4 G$ for $G \in \{\mathbb{Z}_2^3, \mathbb{Z}_4 \oplus \mathbb{Z}_2, D_4\}$; $[12, 9, 3]_4$-codes in $\mathbb{F}_4 G$ for $G \in \{\mathbb{Z}_2^2 \oplus \mathbb{Z}_3, A_4\}$.

Non-associative loop algebras $A = \mathbb{F}_q G$ of small cardinality ($q \leqslant 5$, $|G| \leqslant 7$) have, with rare exceptions, a trivial lattice of left ideals: $0 < I_0 < \Delta < A$. So they do not contain any interesting loop-codes. The exceptional cases yield linearly optimal $[6, 4, 2]_q$- and $[6, 2, 4]_q$-codes for $q \in \{2, 3, 4\}$, and also a $[6, 3, 3]_3$-code. Moreover, the $[6, 2, 4]_q$-code found is an absolutely optimal [266, Ch. 17]. Below we give the Cayley tables of the non-associative loops that generate the $[6, 2, 4]_q$-code and the $[6, 3, 3]_3$-code, respectively:

0	1	2	3	4	5
1	0	3	2	5	4
2	3	4	5	0	1
3	2	5	4	1	0
4	5	0	1	3	2
5	4	1	0	2	3

0	1	2	3	4	5
1	0	3	2	5	4
2	3	4	5	0	1
3	4	5	0	1	2
4	5	0	1	2	3
5	2	1	4	3	0

The study of non-associative loop-codes of greater orders can give more interesting results.

The group algebras considered contain in particular the following well-known absolutely optimal linear codes: the simplex $[7, 3, 4]_2$-code in $\mathbb{F}_2\mathbb{Z}_7$; the $[7, 4, 3]_2$-Hamming-code in $\mathbb{F}_2\mathbb{Z}_7$; the $[8, 4, 4]_2$-Bauer-codes [168] in $\mathbb{F}_2(\mathbb{Z}_2^3)$, $\mathbb{F}_2(\mathbb{Z}_2 \times \mathbb{Z}_4)$, $\mathbb{F}_2 Q_8$, $\mathbb{F}_2 D_4$.

It is also interesting that using non-commutative groups one can construct linearly optimal codes that can not be realized using Abelian groups. Examples of such codes are: an $[8, 3, 5]$-code in $\mathbb{F}_4 Q_4$; $[10, 4, 6]$-codes in $\mathbb{F}_4 D_5$ and in $\mathbb{F}_5 D_5$; $[12, 8, 4]$- and $[12, 6, 6]$-codes in $\mathbb{F}_4 A_4$; a $[12, 6, 6]$-code in $\mathbb{F}_4 D_6$.

Many of the "champion codes" found in [87] are based on non-semisimple and non-commutative algebras. This makes the systematic investigation of codes contained in these algebras of substantial interest.

Acknowledgements

The author is sincerely grateful to professors V.P. Elizarov, E.V. Gorbatov, M. Hazewinkel, V.L. Kurakin, A.S. Kuzmin, V.T. Markov, C. Martínes, A.V. Mikhalev and V.N. Tsypyshev for their valuable suggestions during the preparation of this manuscript and for their help in writing some parts of it. I am also grateful for comprehensive support to the Center of New Information Technologies and to the Department of Higher Algebra of the M.V. Lomonosov Moscow State University, as well as to the Department of Higher Mathematics of the Russian State Social University. I would like to express my special thanks to the editor of the Handbook, prof. M. Hazewinkel, for his infinite patience in waiting for the manuscript of this Chapter and the personal help in transformation of the text into an understandable English one. Of course all the named persons do not bear any responsibility for errors and inaccuracies in this text.

References

[1] W. Adams, P. Loustaunau, An Introduction to Gröbner Bases, Grad. Stud. Math., vol. 3, Amer. Math. Soc., 1994.

[2] A.A. Albert, On nonassociative division algebras, Trans. Amer. Math. Soc. 72 (1952) 296–309.

[3] A.A. Albert, Finite noncommutative division algebras, Proc. Amer. Math. Soc. 9 (1958) 928–932.

[4] A.A. Albert, Finite division algebras and finite planes, in: Proc. Sympos. Appl. Math., vol. 10, Amer. Math. Soc., 1960, pp. 53–70.

[5] A.A. Albert, Generalized twisted fields, Pacific J. Math. 11 (1961) 1–8.

[6] A.N. Alekseichuk, V.P. Elizarov, Finite rings with big number of zero divisors, Diskret. Mat. 4 (2) (1992) 45–51 (in Russian); English transl. in: Discrete Math. Appl. 4 (1993) 51–57.

[7] R.B. Alien, On hypercomplex number systems belonging to an arbitrary domain of rationality, Trans. Amer. Math. Soc. 9 (1908) 203–218.

[8] Y. Al-Khamees, The intersection of distinct Galois subrings is not necessarily Galois, Compos. Math. 40 (3) (1980) 283–286.

[9] Y. Al-Khamees, The enumeration of finite principal completely primary rings, Abh. Math. Sem. Univ. Hamburg 51 (1981) 226–231.

[10] Y. Al-Khamees, Finite rings in which the multiplication of any two zero-divisors is zero, Arch. Math. 37 (2) (1981) 144–149.

[11] Y. Al-Khamees, On the structure of finite completely primary rings, J. Coll. Sci. King Saud Univ. 13 (1) (1982) 149–153.

[12] Y. Al-Khamees, The determination of the group of automorphisms of finite chain ring of characteristic p, Q. J. Math. Oxford (2) 42 (1991) 387–391.

[13] B. Amberg, L. Kazarin, On the central series of the adjoint group of a nilpotent p-algebra, Publ. Math. Debrecen 63 (3) (2003) 473–482.

[14] B. Amberg, L. Kazarin, On the adjoint group of a finite nilpotent p-algebra, (English) J. Math. Sci., (NY) 102 (3) (2000) 3979–3997.

[15] V.G. Antipkin, V.P. Elizarov, Rings of order p^3, Sibirsk. Math. Zh. 23 (4) (1982) 9–18 (in Russian); English transl. in: Siberian Math. J. 23 (4) (1982) 457–464.

[16] L.M. Arkhipov, Finite principal ideal rings, Mat. Zametki 12 (4) (1972) 373–379 (in Russian); English transl. in: Math. Notes 12 (1972) 656–659 (1973).

[17] E. Artin, Über einen Satz von Herrn J.H. Maclagan Wedderburn, Abh. math. Semin. Univ. Hamburg 5 (1927) 245–250.

[18] E. Artin, Zur Theorie der hyperkomlexen Zahlen, Abh. Math. Semin. Univ. Hamburg 5 (1927) 251–260.

[19] E. Artin, Zur Arithmetik hyperkomplexen Zahlen, Abh. Math. Semin. Univ. Hamburg 5 (1927) 261–289.

[20] E. Artin, C. Nesbitt, R. Thrall, Rings with Minimum Condition, Univ. Michigan Press, 1944.

[21] E. Artin, G. Whaples, The theory of simple rings, Amer. J. Math. 65 (1943) 87–107.

[22] K. Asano, Über verallgemeinerte Abelsche Gruppen mit hyperkomp-lexen Operatorenring und ihre Anwendungen, Japan J. Math. 15 (4) (1939) 231–253.

[23] K. Asano, Über Hauptidealringe mit Kettensatz, Osaka Math. J. 1 (1) (1949) 52–61.

[24] M.F. Atyah, I.G. MacDonald, Introduction to Commutative Algebra, Addison–Wesley, Reading, MA, 1969.

[25] G. Ayoub, On finite primary rings and their groups of units, Compos. Math. 21 (3) (1969) 247–252.

[26] G. Ayoub, On the group of units of certain rings, J. Number Theory 4 (1972) 383–403.

[27] G. Azumaya, A duality theory for injective modules (Theory of quasi-Frobenius modules), Amer. J. Math. 81 (1) (1959) 249–278.

[28] Ju.A. Bakhturin, Identities in Lee Algebras, Nauka, Moscow, 1985 (in Russian).

[29] Yu.A. Bakhturin, A.Yu. Ol'shanskij, Identical relations in finite Lie rings, Mat. Sb. 96 (138) (1975) 543–559 (in Russian); English transl. in: Math. USSR Sb. 25 (1975) 507–523.

[30] B. Baer, Inverses and zero divisors, Bull. Amer. Math. Soc. 48 (1942) 630–638.

[31] B. Baer, Kriterien für die Existenz eines Einselement in Ringen, Math. Z. 56 (1952) 1–17.

[32] R. Ballieu, Anneaux finis; systèmes hypercomplexes de rang deux sur un corps, Ann. Soc. Sci. Bruxelles 61 (1) (1947) 117–126.

[33] R. Ballieu, Anneaux finis; systèmes hypercomplexes de rang trois sur un corps commutatif, Ann. Soc. Sci. Bruxelles 61 (1) (1947) 222–227.

[34] R. Ballieu, Anneaux finis à module de type (p, p^2), Ann. Soc. Sci. Bruxelles 63 (1) (1949) 11–23.

[35] R. Ballieu, M.J. Schuind, Anneaux finis à module de type (p, p^r), Ann. Soc. Sci. Bruxelles 63 (1) (1949) 137–147.

[36] R. Ballieu, M.J. Schuind, Anneaux à module de type (p^m, p^{m+n}), Ann. Soc. Sci. Bruxelles 65 (1) (1951) 33–40.

[37] D. Bayer, M. Stillman, Computation of Hilbert functions, J. Symbolic Comput. 14 (1) (1992) 31–50.

[38] K. Baumgartner, Bemerkungen zum Isomorphieproblem der Ringe, Monatsh Math. 70 (1966) 299–308.

[39] R.A. Beaumont, Rings with additive group which is the direct sum of cyclic groups, Duke Math. J. 15 (2) (1948) 367–369.

[40] E.A. Behrens, Einreihige Ringe, Math. Z. 77 (3) (1961) 207–218.

[41] G.M. Bergman, Some examples in PI ring theory, Israel J. Math. 18 (3) (1974) 257–277.

[42] D.V., Belkin, About finite lattices of quasi-varietes of modules over finite rings, Preprint no. 11, Novosibirsk, 1995 (in Russian).

[43] H.E. Bell, Rings with finitely many subrings, Math. Ann. 182 (4) (1969) 314–318.

[44] M. Billis, Unique factorization in the integers modulo n, Amer. Math. Monthly 75 (5) (1968) 527.

[45] J.C. Binz, Endiche Ringe, Prax. Math. 12 (12) (1970) 325–330.

[46] J.F. Blake, Codes over certain rings, Inform. Control 20 (1972) 396–404.

[47] J.F. Blake, Codes over integer residue rings, Inform. Control 29 (4) (1972) 295–300.

[48] D.M. Bloom, List of 11 rings of order 4, Amer. Math. Monthly 71 (8) (1964) 918–919.

[49] A.H. Boers, L'anneau à quatre elements, Indag. Math. 28 (1) (1966) 14–21.

[50] D. Bollman, H. Ramirez, On the enumeration of matrices over finite commutative rings, Amer. Math. Monthly 76 (9) (1969) 1019–1023.

[51] V.M. Bondarenko, About similarity of matrices over a residue rings, Mat. Sb. Inst. Mat. AN USSR, Kiev 3 (1976) 275–277 (English transl.).

[52] Z.I. Borevich, I.R. Shafarevich, Number Theory, Academic Press, New York, 1966.

[53] N. Bourbaki, P.M. Cohn, J. Howie, Algebra II: Chapters 4–7, Springer, Berlin, 2003.

[54] M.P. Brameret, Treillis d'ideaux et structure d'anneaux, C. R. Acad. Sci. Paris 255 (1962) 1434–1435.

[55] M.P. Brameret, Treillis d'ideaux et structure d'anneaux, in: Semin. Dubreil–Pisot, 1962–1963, Fac. Sci. Paris 16 (1) (1967) 1–12.

[56] R. Brauer, On the nilpotency of the radical of a ring, Bull. Amer. Math. Soc. 48 (1942) 752–758.

[57] J.V. Brawely, L.A. Carlitz, A characterization of the $n \times n$-matrices over a finite fields, part I, Amer. Math. Monthly 80 (6) (1973) 670–672.

[58] J.V. Brawely, R.O. Gambl, Involutory matrices over finite commutative rings, Linear Algebra Appl. 21 (1978) 175–178.

[59] B. Buchberger, Ein Algorithmus zum auffinden der Basiselementen des Restklassenringes nach einem nulldimenschionalen Polynomideal, PhD thesis, Inst. University of Insbruck, Insbruck, Austria, 1965.

[60] L.R. Busarkina, Rings of first rank with nilpotent generators, Ural. Gos. Univ. Mat. Zap. 3 (3) (1962) 25–29 (in Russian).

[61] L.A. Carlitz, Functions and polynomials (mod p^n), Acta Arith. 24 (1) (1964) 67–78.

[62] E. Cartan, Les groupes bilinéaires des systèmes de nombres complexes, Ann. Fac. Sci. Univ. Toulouse B 12 (1898) 1–99.

[63] A. Cayley, On double algebras, Proc. London Math. Soc. 151 (1883) 185–197.

[64] P.V. Ceccherini, Some new results on certain finite structures, Lincei-Rend. Sci. Fis. Mat. Natur. 56 (1974) 840–855.

[65] A. Châtelet, Les groupes abéliens finis et les modules de points entier, Gauthier-Villars, Paris–Lille, 1925.

[66] C. Christensen, Rings with a few more zero divisors, Bull. Austral. Math. Soc. 5 (2) (1971) 271–274.

[67] C.J. Chikunji, On a class of finite rings, Comm. Algebra 27 (10) (1999) 5049–5081.

[68] C.J. Chikunji, Enumeration of finite rings with Jacobson radical of cube zero, http://arxiv.org/abs/math.RA/9905030/, 1999.

[69] C.J. Chikunji, On a class of rings of order p^5, Math. J. Okayama Univ. 45 (2003) 59–71.

[70] C.J. Chikunji, Using MATLAB to solve a classification problem in finite rings, in: 2-nd Internat. Conf. on the Teaching of Math., Greece, 2003, http://www.math.uoc.gr/ictm2/Proceedings/pap252.pdf.

[71] W.E. Clark, A coefficient ring for finite noncommutative rings, Proc. Amer. Math. Soc. 33 (1) (1972) 25–27.

[72] W.E. Clark, D.A. Drake, Finite chain rings, Abh. Math. Semin. Univ. Hamburg 39 (1973) 147–153.

[73] W.E. Clark, J.J. Liang, Enumeration of finite commutative chain rings, J. Algebra 27 (3) (1973) 445–453.

[74] H.L. Claasen, R.W.A. Goldbach, A field-like property of finite rings, Indag. Math. 3 (1992) 11–26.

[75] I.S. Cohen, On the structure and ideal theory of complete local rings, Trans. Amer. Math. Soc. 59 (1946) 54–106.

[76] P.M. Cohn, On the structure of the GL of a ring, Publ. Math. Inst. Hautes Études Sci. 30 (1966) 365–413.

[77] I. Constantinescu, W. Heise, A metric for codes over residue class rings, Problemy Peredachi Informacii 33 (3) (1997) 22–28; English transl: Probl. Inf. Transm. 33 (3) (1997) 208–213.

[78] I. Constantinescu, W. Heise, T. Honold, Monomial extensions of isometries between codes over $\mathbb{Z}_m$, in: Proc. of the 5th Internat. Workshop on Algebraic and Combinatorial Coding Theory, Unicorn Shumen, 1996, pp. 98–104.

[79] B. Corbas, Rings with few zero divisors, Math. Ann. 181 (1) (1969) 1–7.

[80] B. Corbas, Finite rings in which the product of any two zero divisors is zero, Arch. Math. 21 (5) (1970) 466–469.

[81] B. Corbas, P.D. Williams, Rings of order p^5. Part 1. Nonlocal rings, J. Algebra 231 (2000) 677–690.

[82] B. Corbas, P.D. Williams, Rings of order p^5. Part 2. Local rings, J. Algebra 231 (2000) 691–704.

[83] B. Corbas, P.D. Williams, Congruence classes in $M_3(\mathbb{F}_q)$ (q-odd), Discrete Math. 219 (2000) 27–47.

[84] B. Corbas, P.D. Williams, Congruence classes in $M_3(\mathbb{F}_q)$ (q-even), Discrete Math. 257 (2002) 15–27.

[85] M. Cordeo, G.P. Whene, A survey of finite semifields, Discrete Math. 208–209 (1999) 125–137.

[86] G.F. Cotti, Sul una classe die anelli cocritici, Riv. Mat. Univ. Parma 3 (1977–1978) 203–211.

[87] E. Couselo, S. Gonzalez, V. Markov, A. Nechaev, Loop-codes, Discrete Math. Appl. 14 (2) (2004) 163–172, VSP.

[88] I.G. Counell, A number theory problem concerning finite rings, Canad. Math. Bull. 7 (1964) 23–37.

[89] D. Cox, J. Little, D. O'Shea, Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra, Springer, Berlin, 1992.

[90] J. Dayer-Bennet, A note on finite regular rings, Bull. Amer. Math. Soc. 47 (1941) 784–787.

[91] P. Delsarte, An algebraic approach to the association schemes of coding theory, Philips Research (Rep. Suppl.) 10 (1973).

[92] J.B. Derr, J.F. Orr, P.S. Peck, Noncommutative rings of order p^4, J. Pure Appl. Algebra 97 (1994) 109–116.

[93] M.G. Desphand, Structure of right subdirectly irreducible rings, J. Algebra 17 (3) (1971) 317–325.

[94] M. Deuring, Algebren, Springer, Berlin, 1935.

[95] L.E. Dickson, On finite algebras, Gottingen Nachr. (1905) 358–393.

[96] L.E. Dickson, Linear algebras in which division is always uniquely possible, Trans. Amer. Math. Soc. 7 (1906) 370–390.

[97] L.E. Dickson, Algebras and Their Arithmetics, Chicago, 1923.

[98] L.E. Dickson, Duke Math. J. 1 (1935) 113–125.

[99] J.L. Dorroh, Concerning adjunctions to algebras, Bull. Amer. Soc. 38 (1) (1932) 85–88.

[100] N.H. Eggert, Quasi-regular groups of finite commutative nilpotent algebras, Pacific J. Math. 36 (1971) 631–634.

[101] K.E. Eldridge, Orders for finite noncommutative rings with unity, Amer. Math. Monthly 75 (5) (1968) 512–514.

[102] K.E. Eldridge, I. Fischer, DCC rings with a cyclic group of units, Duke Math. J. 34 (1967) 243–248.

[103] V.P. Elizarov, Non-nilpotent rings of order p^4, in: Proceedings of the Fifth All-Union Sympos. on Ring's, Algebra's and Module's Theory, Novosibirsk, 1982, pp. 51–52 (in Russian).

[104] V.P. Elizarov, Non-nilpotent finite rings, manuscript, VINITI, No. 1472-85, 37 pp. (in Russian).

[105] V.P. Elizarov, Systems of linear equations over a commutative ring, Uspekhi Mat. Nauk 48 (2) (1993) 181–182; English transl. in: Russian Math. Surveys 48 (2) (1993) 175–177.

[106] V.P. Elizarov, General solution of the homogeneous linear equations system over a commutative ring, Uspekhi Mat. Nauk 49 (1) (1994) 141–142; English transl. in: Russian Math. Surveys 49 (2) (1994) 153–154.

[107] V.P. Elizarov, Systems of linear equations over quasi-Frobenius rings, Fundam. Prikl. Mat. 1 (2) (1995) 535–539 (in Russian, English summary).

[108] V.P. Elizarov, Systems of linear equations over finite rings, in: Memoirs in Discrete Math., vol. 6, Fizmatlit, Moscow, 2002, pp. 535–539 (in Russian).

[109] V.P. Elizarov, Systems of linear equations over modules, Fundam. Prikl. Mat. 8 (4) (2002) 979–991 (in Russian, English summary).

[110] V.P. Elizarov, Necessary conditions for solvability of a system of linear equations over a ring, Diskret. Mat. 16 (2) (2004) 44–53; English. transl.: Discrete Math. Appl. 14 (2) (2004) 153–162.

[111] V.P. Elizarov, Finite Rings (Foundations of the Theory), Gelios–ARV, Moscow, 2006, 304 pp. (in Russian).

[112] V.P. Elizarov, Solvable and locally closed modules and rings, Diskret. Mat. 18 (1) (2006) 30–39; English transl. in: Discrete Math. Appl. 16 (1) (2006) 29–37.

[113] D.B. Erickson, Orders for finite noncommutative rings, Amer. Math. Monthly 73 (4) (1966) 376–377.

[114] G.J. Everett, Rings as groups with operators, Bull. Amer. Math. Soc. 45 (1939) 274–279.

[115] G.J. Everett, Vector spaces over rings, Bull. Amer. Math. Soc. 48 (1942) 312–316.

[116] C. Faith, Algebra II. Ring Theory, Springer, Berlin, 1976.

[117] H.K. Farahat, The multiplicative groups of a ring, Math. Z. 87 (3) (1965) 378–384.

[118] E.H. Feller, Properties of primary noncommutative rings, Trans. Amer. Math. Soc. 89 (1) (1958) 79–91.

[119] E.H. Feller, E.W. Swokowski, On ring extension for completely primary noncommutative rings, Trans. Amer. Math. Soc. 105 (2) (1962) 251–263.

[120] J.L. Fisher, Finite principal ideal rings, Canad. Math. Bull. 19 (3) (1976) 277–283.

[121] M.Ya. Finkelstein, Presentations of finite rings with identity, Math. Res. (Kishinev) 74 (1983) 131–139 (in Russian).

[122] H. Fitting, Theorie der Automorphismenringe Abelscher Gruppen und ihr Analogen bei nicht kommutativen Gruppen, Math. Ann. 107 (1933) 514–542.

[123] H. Fitting, Die Determinantenideale eines Modulus, Jahr. Deutsch. Math. Verein. 46 (9–12) (1936) 195–228.

[124] C.R. Fletcher, Unique factorization rings, Proc. Cambridge Philos. Soc. 65 (3) (1969) 580–583.

[125] C.R. Fletcher, The structure of unique factorization rings, Proc. Cambridge Philos. Soc. 67 (3) (1970) 535–540.

[126] C.R. Fletcher, Euclidean rings, J. London Math. Soc. 4 (1) (1971) 79–82.

[127] C.R. Fletcher, Equivalent conditions for unique factorization, Publ. Dept. Math. Lyon 8 (1) (1971) 13–22.

[128] C.R. Fletcher, Rings of small order, Gaz. Math. 64 (427) (1980) 9–22.

[129] C.R. Fletcher, M. Lidster, Constructing a ring without unique factorization, Gaz. Math. 62 (420) (1978) 104–109.

[130] W. Flor, J. Wiesenbauer, Zum Klassifikationsproblem endlicher Ringe, J. Osterreich Acad. Wiss. Math. Natur. Kl. Sitzungsber. II 183 (8–10) (1975) 309–320.

[131] J.B. Fountein, Nilpotent principal ideal rings, Proc. London Math. Soc. (3) 20 (2) (1970) 348–364.

[132] A. Fraenkel, Über die Teiler der Null und die Zerlegung von Ringen, J. Reine Angew. Math. 145 (1915) 139–176.

[133] A. Fraenkel, Über einfache Erweiterungen zerlegbarer Ringe, J. Reine Angew. Math. 151 (1921) 121–166.

[134] P.A. Freidman, Some finiteness conditions in associative rings, Ural. Gos. Univ. Mat. Zap. 3 (1) (1961) 77–84 (in Russian).

[135] P.A. Freidman, Rings whose commutative subrings are all Artinian, Izv. Vyssh. Uchebn. Zaved. Mat. 10 (137) (1973) 83–89 (in Russian).

[136] G. Frobenius, Theorie der hyperkomplexen Grossen. I, II, Sitz. Preuss. Akad. Wiss. 504–537 (1903) 634–645.

[137] S. Galovich, Unique factorization rings with zero divisors, Math. Mag. 51 (5) (1978) 276–283.

[138] N. Ganesan, Properties of rings with a finite number of zero divisors. I, Math. Ann. 157 (3) (1964) 215–217.

[139] N. Ganesan, Properties of rings with a finite number of zero divisors. II, Math. Ann. 161 (4) (1965) 241–246.

[140] G. Ganske, B. McDonald, Finite local rings, Rocky Mountain J. 3 (4) (1973) 512–540.

[141] T.G. Gazaryan, An example of non-isomorphic chain rings, Uspekhi Mat. Nauk 47 (3) (1992) 155–156 (in Russian); English transl. in: Russian Math. Surveys 47 (3) (1992).

[142] T.G. Gazaryan, Properties of matrices over a commutative ring with connection to the similarity relation, Diskret. Mat. 6 (3) (1994) 143–153 (in Russian); English transl. in: Discrete Math. Appl. 4 (1994) 455–466.

[143] T.G. Gazaryan, Similarity of involutory matrices over a local ring of characteristic 2^k, Diskret. Mat. 7 (4) (1995) 145–156 (in Russian); English transl. in: Discrete Math. Appl. 5 (6) (1995) 587–601.

[144] K.S. Ghent, A note on nilpotent algebras in four units, Bull. Amer. Math. Soc. 40 (1934) 331–338.

[145] G.K. Genov, P.N. Siderov, A basis of the identities of the fourth order matrix algebra over a finite field. I, Serdika Bulgar. Mat. 8 (1982) 313–323 (in Russian).

[146] G.K. Genov, P.N. Siderov, A basis of the identities of the fourth order matrix algebra over a finite field. II, Serdika Bulgar. Mat. 8 (1982) 351–366 (in Russian).

[147] G.K. Genov, A basis of identities of the algebra of third-order matrices over a finite field, Algebra Logika 20 (1981) 365–388 (in Russian); English transl. in: Algebra Logic 20 (1982) 241–257.

[148] R.W. Gilmer, Finite rings having a cyclic multiplicative group of units, Amer. J. Math. 85 (3) (1963) 447–452.

[149] R.W. Gilmer, A note on rings with only finitely many subrings, Scripta Math. 29 (1–2) (1973) 37–38.

[150] R.W. Gilmer, Zero divisors in commutative rings, Amer. Math. Monthly 93 (5) (1986) 382–387.

[151] R.W. Gilmer, J. Mott, Associative rings of order p^3, Proc. Japan Acad. 49 (10) (1973) 795–799.

[152] L.M. Gluskin, Ideals in rings and their multiplicative semigroups, Uspekhi Mat. Nauk 15 (4) (1960) 141–148 (in Russian); English transl. in: Amer. Math. Soc. Transl., (2) 27 (1963) 297–304.

[153] S. Gonzalez, V.T. Markov, C. Martines, A.A. Nechaev, I.F. Rua, Nonassociative Galois rings, Discrete Math. Appl. 12 (6) (2002) 591–606.

[154] S. Gonzalez, V.T. Markov, C. Martines, A.A. Nechaev, I.F. Rua, Coordinate sets of a generalized Galois rings, J. Algebra Appl. 3 (1) (2004) 31–48.

[155] S. Gonzalez, V.T. Markov, C. Martines, A.A. Nechaev, I.F. Rua, Cyclic generalized Galois rings, Comm. Algebra 33 (12) (2005) 1–12.

[156] S. Gonzalez, V.T. Markov, C. Martines, A.A. Nechaev, I.F. Rua, On cyclic top-associative generalized Galois rings, in: Lecture Notes in Computer Sci., vol. 2948, 2004, pp. 25–37.

[157] A.W. Goldie, Non-commutative principal ideal rings, Arch. Math. 13 (1962) 213–221.

[158] E.V. Gorbatov, Standard basis of a polynomial ideal over commutative Artinian chain ring, Diskret. Mat. 16 (1) (2004) 52–78 (in Russian); English transl. in: Discrete Math. Appl. 14 (1) (2004) 75–101.

[159] E.V. Gorbatov, Standard bases concordant with the norm and computations in ideals and polylinear recurring sequences, Fundam. Prikl. Mat. 10 (3) (2004) 23–71 (in Russian); English transl. in: J. Math. Sci. (NY) 139 (4) (2006) 6672–6707.

[160] E.V. Gorbatov, A.A. Nechaev, Cyclic families of polylinear recurrences over quasi-Frobenius modules, Uspekhi Mat. Nauk 57 (4) (2002) 167–168 (in Russian); English transl. in: Russian Math. Surveys 57 (4) (2002).

[161] B.O. Gorlov, Finite nilpotent algebras with metacyclic adjoint group, Ukrain Math. Zh. 47 (1995) 1426–1431.

[162] M. Greferath, S.E. Schmidt, Finite-ring combinatorics and MacWilliams' equivalence theorem, J. Combin. Theory Ser. A 92 (1) (2000) 17–28.

[163] M. Greferath, A. Nechaev, R. Wisbauer, Finite quasi-Frobenius modules and linear codes, J. Algebra Appl. 3 (3) (2004) 1–26.

[164] E.V. Guravlew, Local rings of order p^6 with 4-nilpotent radical, Sib. Elek. Mat. Izv. 3 (2006) 15–59, http://semr.math.nsc.ru.

[165] W.R. Hamilton, Lectures on Quaternions, Dublin, 1853.

[166] A.R. Hammons, P.V. Kumar, A.R. Calderbank, N.J.A. Sloane, P. Sole, The $\mathbb{Z}_4$-linearity of Kerdock, Preparata, Goethals and related codes, IEEE Trans. Inform. Theory 40 (2) (1994) 301–319.

[167] M. Hazewinkel, N. Gubareni, V. Kirichenko, Algebras, Rings and Modules, vol. 1, Math. Appl., vol. 575, Springer, Berlin, 2004, vol. 2, 2007.

[168] W. Heise, P. Quattrocci, Informations- und Codierungstheorie, Springer, Berlin, 1995.

[169] W. Heise, Th. Honold, A.A. Nechaev, Weighted modules and representations of codes, in: Proceedings of the Sixth International Workshop on Algebraic and Combinatorial Coding Theory (ACCT-VI), Pskov, Russia, September 6–12, 1998, Mezhd. Tsentr Nauchn. Tekhn. Inform., Moscow, 1998, pp. 123–129.

[170] K. Hensel, Eine neue Theorie der algebraischen, Zahlen. Math. Z. 2 (1918) 433–452.

[171] I.R. Hentzel, I.F. Rua, Primitivity of finite semifields with 64 and 81 elements, Internat. J. Algebra Comput. (2007).

[172] I.N. Herstein, A proof of a conjecture of Vandiver, Proc. Amer. Math. Soc. 1 (3) (1950) 370–371.

[173] I.N. Herstein, An elementary proof of a theorem of Jacobson, Duke Math. J. 21 (1954) 45–48.

[174] I.N. Herstein, Noncommutative Rings, Wiley, New York, 1968.

[175] Y. Hirano, On residually finite rings, Math. J. Okayama Univ. 35 (1993) (reprinted).

[176] J.H. Hodges, Idempotent matrices (mod p^a), Amer Math. Monthly 73 (1966) 277–278.

[177] S.J. Hong, A.M. Patel, A general class of maximal codes for computer application, IEEE Trans. Comput. 21 (12) (1972) 1322–1331.

[178] Th. Honold, I. Landjev, Linearly representable codes over chain rings, in: Proceedings of the Sixth International Workshop on Algebraic and Combinatorial Coding Theory (ACCT-VI), Pskov, Russia, September 6–12, 1998, Mezhd. Tsentr Nauchn. Tekhn. Inform., Moscow, 1998, pp. 135–141.

[179] T. Honold, Characterization of finite Frobenius rings, Arch. Math. (Basel) 76 (6) (2001) 406–415.

[180] Th. Honold, A.A. Nechaev, Weighted modules and representations of codes, Problemy Peredachi Informacii 35 (3) (1999) 18–39 (in Russian); Probl. Inf. Transm. 35 (3) (1999) 205–223.

[181] T.W. Hungerford, On the structure of principal ideal rings, Pacific J. Math. 25 (3) (1968) 543–547.

[182] X.-D. Hou, A.A. Nechaev, A construction of finite Frobenius rings and its application to partial difference set, J. Algebra 309 (1) (2007) 1–9.

[183] M. Istinger, H.K. Kaiser, A characterization of polynomially complete algebras, J. Algebra 56 (1979) 103–110.

[184] N. Jacobson, The Theory of Rings, Amer. Math. Soc. Math. Surveys, vol. 1, Amer. Math. Soc., New York, 1943, vi+150 pp.

[185] N. Jacobson, Structure theory for algebraic algebras of bounded degree, Ann. of Math. 46 (1945) 695–703.

[186] N. Jacobson, The radical and semisimplicity for arbitrary rings, Amer. J. Math. 67 (1945) 300–320.

[187] N. Jacobson, Structure of Rings, Amer. Math. Soc., Providence, RI, 1956.

[188] S.K. Jain, J. Luh, B. Zimmermann-Huisgen, Finite uniserial rings of prime characteristic, Comm. Algebra 16 (10) (1988) 2133–2135.

[189] G.J. Janusz, Separable algebras over commutative rings, Trans. Amer. Math. Soc. 122 (1966) 461–478.

[190] A.W. Jategaonkar, Left Principal Ideal Rings, Lecture Notes in Math., vol. 123, Springer, Berlin, 1970.

[191] R.E. Johnson, Principal right ideal rings, Canad. J. Math. 15 (2) (1963) 297–301.

[192] O.V. Kamlovskii, A.S. Kuzmin, Distributions of elements on cycles of linear recurring sequences over Galois rings, Russian Math. Surveys 53 (2) (1998) 392–393.

[193] F. Kasch, Moduln und Ringe, Teubner, 1977.

[194] A. Kertesz, Vorlesungen über Artinsche Ringe, Teubner, 1968.

[195] G. Keller, F.R. Olson, Counting polynomial functions (mod p^n), Duke Math. J. 35 (4) (1968) 835–838.

[196] I. Kheifetc, The extension theorem for isometries of linear codes over QF-modules, Fundam. Prikl. Mat. 7 (4) (2001) 1227–1236 (in Russian).

[197] E. Kircher, Some properties of certain finite algebras, Amer. J. Math. 39 (3) (1917) 273–280.

[198] V.V. Kirichenko, About quasi-Frobenius rings and Gorenstein orders, Tr. Mat. Inst. Steklov 148 (1978) 168–174; (in Russian); English transl. in: Proc. Steklov Inst. Math. 148 (1978) 171–177.

[199] W. Klingenberg, Lineare Gruppen über lokalen Ringen, Amer. J. Math. 83 (1961) 137–153.

[200] W. Klingenberg, Die Structur der lineare Gruppen über einem nichtkommutativen lokalen Ring, Arch. Math. 13 (1962) 73–81.

[201] J. Knopfmacher, Arithmetical properties of finite rings and algebras, and analytic number theory, Bull. Amer. Math. Soc. 76 (4) (1970) 830–833.

[202] J. Knopfmacher, Arithmetical properties of finite rings and algebras, and analytic number theory. I–VI, J. Reine Angew. Math. 252 (1972) 16–43; 254 (1972) 74–99, 259 (1973) 157–170; 270 (1974) 97–114; 271 (1974) 95–121; 277 (1975) 45–62.

[203] J. Knopfmacher, On the asymptotic enumeration of finite rings and modules, Arch. Math. 26 (6) (1975) 615–619.

[204] K. Kodaira, Über die Struktur des endlichen vollstandig primaren Ringes mit verschvindenden Radikalquadrat, Japan J. Math. 14 (1) (1937) 15–21.

[205] K. Koh, On "Properties of rings with a finite number of zero divisors", Math. Ann. 171 (1) (1967) 79–80.

[206] S.S. Korobkov, Lattice isomorphisms of finite reduced rings (in Russian, English summary) Izv. Ural. Gos. Univ. Mat. Mekh. 22 (4) (2002) 81–93.

[207] G. Köthe, Die Struktur der Ringe, deren Rcstklassen ring nach dem Radical vollstandig reduzibel ist, Math. Z. 32 (2) (1930) 161–181.

[208] G. Köthe, Über maximale nilpotente Unterringe und Nilringe, Math. Ann. 103 (2) (1930) 359–363.

[209] G. Köthe, Verallgemeinerte Abelsche Gruppen mit hyperkomplexen Operatoren Ring, Math. Z. 35 (1) (1934) 31–44.

[210] I. Kovacs, Infinite rings without infinite proper subrings, Publ. Math. Debrecen 4 (1–2) (1955) 104–107.

[211] L. Kronecker, Vorlesungen über Zahlentheorie, vol. 1, Teubner, Leipzig, 1901.

[212] W. Krull, Algebraische Theorie der Ringe. I, Math. Ann. 88 (1922) 80–122.

[213] W. Krull, Algebraische Theorie der Ringe. II, Math. Ann. 91 (1924) 1–46.

[214] W. Krull, Algebraische Theorie der Ringe. III, Math. Ann. 92 (1924) 183–213.

[215] W. Krull, Die verschiedenen Arten der Hauptidealringe, Sitz. Ber. Heidelbergen Akad. Wiss. Math.-Nat. Klasse 6 (1924) 3–16.

[216] W. Krull, Über verallgemeinerte endliche Abelsche Gruppen, Math. Z. 23 (1925) 161–196.

[217] W. Krull, Idealtheorie, Springer, 1935.

[218] W. Krull, Über den Galoisringen, Math. Ann. 185 (1) (1970) 31–33.

[219] R.L. Kruse, Identities satisfied by finite rings, J. Algebra 26 (1973) 298–318.

[220] R.L. Kruse, D.T. Price, On the subring structure of finite nilpotent rings, Pacific J. Math. 31 (1) (1969) 103–117.

[221] R.L. Kruse, D.T. Price, Nilpotent Rings, Gordon & Breach, New York, 1967.

[222] R.L. Kruse, D.T. Price, Enumerating finite rings, J. London Math. Soc. 2 (1) (1970) 149–159.

[223] V.L. Kurakin, The first coordinate sequence of a linear recurrence of maximal period over a Galois ring, Discrete Math. Appl. 4 (1994) 129–141.

[224] V.L. Kurakin, Exponent of congruence-subgroup of finite commutative chain ring, Algebra Logic 36 (6) (1997).

[225] V.L. Kurakin, Similarity invariants for matrices over a commutative Artinian chain ring, Mat. Zametki 80 (3) (2006) 403–412 (in Russian).

[226] V.L. Kurakin, A.S. Kuz'min, A.V. Mikhalev, A.A. Nechaev, Linear recurring sequences over rings and modules, in: Contemp. Math. and Its Appl. Thematic Surveys, vol. 10, Algebra 2, Moscow, 1994; J. Math. Sci. (NY) 76 (6) (1995) 2793–2915.

[227] V.L. Kurakin, A.S. Kuz'min, A.A. Nechaev, Pseudo-random and polylinear sequences, in: Memoires in Discrete Mathematics, vol. 1, TVP, Moscow, 1997, pp. 139–202 (in Russian).

[228] V.L. Kurakin, A.S. Kuz'min, A.A. Nechaev, Properties of linear and polylinear recurrences over Galois rings (I), in: Memoires in Discrete Mathematics, vol. 2, TVP, Moscow, 1998, pp. 191–222 (in Russian).

[229] V.L. Kurakin, A.S. Kuz'min, A.A. Nechaev, Structural, analytical and statistical properties of linear and polylinear recurrences, in: Memoires in Discrete Mathematics, vol. 3, Fizmatlit, Moscow, 2000, pp. 155–194 (in Russian).

[230] V.L. Kurakin, A.S. Kuz'min, A.A. Nechaev, Statistical properties of linear recurrences over Galois rings and quasi-Frobenius modules of the characteristic 4, in: Memoires in Discrete Mathematics, vol. 4, Fizmatlit, Moscow, 2001, pp. 91–128 (in Russian).

[231] V. Kurakin, A. Kuzmin, A. Nechaev, Codes and linear recurrences over Galois rings and QF-modules of characteristic 4, in: Proc. of the Sixth International Workshop on Algebraic and Combinatorial Coding Theory, ACCT-YI, Pskov, Russia, September 6–12, 1998, Mezhd. Tsentr Nauchn. Tekhn. Inform., Moscow, 1998, pp. 166–171.

[232] V.L. Kurakin, A.S. Kuzmin, V.T. Markov, A.V. Mikhalev, A.A. Nechaev, Linear codes and polylinear recurrences over finite rings and modules (a survey), in: M. Fossorier, et al. (Eds.), Applied Algebra, Algebraic Algorithms and Error Correcting Codes, Proc. 13th Internat. Symposium, AAECC-13, Honolulu, HI, USA, November 15–19, 1999, in: Lecture Notes in Comput. Sci., vol. 1719, Springer, Berlin, 1999, pp. 365–391.

[233] V.L. Kurakin, A.S. Kuzmin, V.T. Markov, A.V. Mikhalev, A.A. Nechaev, Codes and recurrences over finite rings and modules, Moscow Univ. Math. Bull. 54 (5) (1999) 15–28 (in English); transl. from Vestn. Moskov. Univ. Ser. 1 Mat. Mech. 5 (1999) 18–31.

[234] V.L. Kurakin, A.V. Mikhalev, A.A. Nechaev, Polylinear recurring sequences over a bimodule, in: D. Krob, et al. (Eds.), Formal Power Series and Algebraic Combinatorics. Proc. 12th International Conf., FPSAC'00, Moscow, Russia, June 26–30, 2000, Springer, Berlin, 2000, pp. 484–495.

[235] V.L. Kurakin, A.V. Mikhalev, A.A. Nechaev, V.N. Tsypyschev, Linear and polylinear recurring sequences over Abelian groups and modules, J. Math. Sci. (NY) 102 (6) (2000) 4598–4626.

[236] A.G. Kurosh, Higher Algebra, Mir, Moscow, 1980, Nauka, Moscow, 1975 (transl. from the second Russian edition).

[237] A.S. Kuz'min, A.A. Nechaev, Construction of noise-resistant codes by means of linear recurrences over Galois rings, Russian Math. Surveys 47 (5) (1992) 189–190; transl. from Uspekhi Mat. Nauk 47 (5) (1992) 183–184.

[238] A.S. Kuz'min, A.A. Nechaev, Linear recursive sequences over Galois rings, Russian Math. Surveys 48 (1) (1993) 171–172; transl. from Uspekhi Mat. Nauk. (1) 48 (289) (1993) 167–168.

[239] A.S. Kuz'min, A.A. Nechaev, Linear recurring sequences over Galois rings, Algebra Logic 34 (2) (1995) 87–100; transl. from Algebra Logika 34 (2) (1995) 169–189.

[240] A.S. Kuzmin, A.A. Nechaev, Trace-function on a Galois ring in coding theory, in: Lecture Notes in Comput. Sci., vol. 1255, Springer, Berlin, 1997, pp. 277–290.

[241] A. Kuzmin, A. Nechaev, Complete weight enumerators of generalized Kerdock code and related linear codes over Galois ring, Discrete Appl. Math. 111 (2001) 117–137.

[242] T.Y. Lam, A First Course in Noncommutative Rings, Grad. Texts in Math., vol. 131, Springer, 1991.

[243] T.Y. Lam, Lectures on Modules and Rings, Grad. Texts in Math., vol. 189, Springer, 1999.

[244] J. Lambek, Lectures on Rings and Modules, McGill Univ., Blaisdell, 1966, 279 pp.

[245] S. Lang, Algebraic Numbers, Columbia Univ. N.Y., Addison–Wesley, Reading, MA, 1964.

[246] S. Lang, Algebra, Addison–Wesley, Reading, MA, 1965.

[247] H. Lausch, W. Nöbauer, Algebra of Polynomials, North-Holland, Amsterdam, 1973.

[248] T.J. Laffey, On commutative subrings of infinite rings, Bull. London Math. Soc. 4 (1) (1972) 3–5.

[249] T.J. Laffey, Infinite rings with all proper subrings finite, Amer. Math. Monthly 81 (1974) 270–272.

[250] G.E. Leger, A note on some properties of finite rings, Proc. Amer. Math. Soc. 6 (6) (1955) 968–969.

[251] J. Levitzki, Über nilpotente Unterringe, Math. Ann. 105 (1931) 620–627.

[252] S. Ligh, Finite rings with central nilpotent elements, Tamkang J. Math. 16 (3) (1985) 45–46.

[253] P.Z. Lu, A criterion for annihilating ideals of LRS over Galois rings, Appl. Algebra Engrg. Comm. Comput. 11 (2) (2000).

[254] I.V. L'vov, Varieties of associative rings. I, Algebra Logic 12 (1973) 150–167; transl. from: Algebra Logika 12 (1973) 269–297.

[255] I.V. L'vov, Varieties generated by finite alternative rings, Algebra Logic 17 (1979) 195–198; transl. from Algebra Logika 17 (3) (1978) 282–286.

[256] C.C. MacDuffee, Matrices with elements in a principal ideal ring, Bull. Amer. Math. Soc. 39 (8) (1933) 564.

[257] D. MacHall, Commutativity in finite rings, Amer. Math. Monthly 83 (1) (1976) 30–32.

[258] Mathematics in the USSR for 40 Years. vol. 1, Fizmatlit, Moscow, 1959.

[259] N.H. MacCoy, Concerning matrices with elements in a commutative ring, Bull. Amer. Math. Soc. 45 (4) (1939) 280–284.

[260] N. MacCoy, D. Montgomery, A representation of generalized Boolean algebras, Math. J. 3 (1937) 455–459.

[261] K. MacCrimmon, A note on finite division rings, Proc. Amer. Math. Soc. 17 (1966) 1173–1177.

[262] B.R. MacDonald, Finite Rings with Identity, Dekker, New York, 1974.

[263] B.R. MacDonald, Similarity of matrices over Artinian principal ideal rings, Linear Algebra Appl. 21 (2) (1978) 153–162.

[264] K.R. MacLean, Commutative artinian principal ideal rings, Proc. London Math. Soc. (3) 26 (1973) 249–272.

[265] K.R. MacLean, Artinian principal ideal rings without identities, J. London Math. Soc. 12 (1) (1975) 53–58.

[266] F.J. MacWilliams, Combinatorial properties of elementary Abelian groups, PhD thesis, Radcliffe College, Cambridge MA, 1962.

[267] F.J. MacWilliams, N.J.A. Sloane, The Theory of Error-Correcting Codes, North-Holland, Amsterdam, 1977.

[268] D. Mainwaring, K.R. Pearson, Decomposability of finite rings, J. Austral. Math. Soc. A 28 (2) (1979) 136–138.

[269] Yu.N. Mal'tsev, The Structure of Associative Algebras, Altaj. Gos. Univ., Barnaul, 1994.

[270] Yu.N. Mal'tsev, E.N. Kuz'min, A basis for the identities of the algebra of second-order matrices over a finite field, Algebra Logic 17 (1978) 18–21; transl. from Algebra Logika 17 (1) (1978) 28–32.

[271] Yu.N. Mal'tsev, A.A. Nechaev, Critical rings and varieties of algebras, Algebra Logic 18 (1980) 210–214.

[272] Yu.N. Mal'tsev, Representation of finite rings by matrices over commutative ring, Math. Sb. (N.S.) 128 (170) (3) (1985) 383–402 (in Russian); English transl. in: Math. of the USSR Sb. 56 (1987) 379–402.

[273] V.T. Markov, A.A. Nechaev, Radicals of semiperfect rings related to idempotents, Fundam. Prikl. Mat. 6 (1) (2000) 293–298 (in Russian) available at http://mech.math.msu.su/fpm/eng/k00/k001/k00125h.htm.

[274] Yu.A. Medvedev, Identities of finite Jordan Φ-algebras, Algebra Logic 18 (1) (1980) 460–478; transl. from Algebra Logika 18 (1979) 723–748.

[275] F.E. Molin, Über Systeme höherer complexer Zahlen, Math. Ann. 41 (1893) 83–156.

[276] R.A. Melter, The number of invertible elements of a matrix ring over a finite p-ring, Mathematicae (RSR) 10 (4) (1976) 101–102.

[277] D.A. Mikhailov, A.A. Nechaev, Solving systems of polynomial equations over Galois–Eisenstein rings with the use of the canonical generating systems of polynomial ideals, Diskret. Mat. (in Russian); English transl. in: Discrete Math. Appl. 14 (1) (2004) 41–73.

[278] A.V. Mikhalev, A.A. Nechaev, Linear recurring sequences over modules, Acta Appl. Math. 42 (2) (1996) 161–202.

[279] M. Mizue, On the automorphisms of a certain class of finite rings, Kodai Math. Semin. Rep. 18 (4) (1996) 357–367.

[280] M. Nagata, Local Rings, Interscience Tracts in Pure Appl. Math., vol. 13, 1962, New York.

[281] T. Nakayama, On Frobenius algebras. I, Ann. of Math. 40 (1939) 611–633.

[282] T. Nakayama, Note on uni-serial and generalized uni-serial rings, Proc. Imp. Acad. Tokyo 16 (7) (1940) 285–289.

[283] A.A. Nechaev, On the structure of finite commutative rings with an identity, Mat. Zametki 10 (6) (1971) 679–688 (in Russian); English transl. in: Math. Notes 10 (1971) 840–845.

[284] A.A. Nechaev, Finite principal ideal rings, Math. Sb. 91 (3) (1973) 350–366 (in Russian); English transl. in: Math. USSR Sb. 20 (1973) 364–382.

[285] A.A. Nechaev, On a problem in the theory of rings of principal ideals, Mat. Zametki 15 (5) (1974) 757–763 (in Russian); English transl. in: Math. Notes 15 (1974) 453–457.

[286] A.A. Nechaev, Some radicals of finite rings connected with idempotents, in: Proc. of XIII All-Union Algebr. Sympos., vol. 2, Gomel, 1975, p. 326 (in Russian).

[287] A.A. Nechaev, Polynomial transformations of finite commutative principal ideal rings, Mat. Zametki 27 (6) (1980) 885–899 (in Russian); English transl. in: Math. Notes 27 (1980) 425–432.

[288] A.A. Nechaev, Criterion for completeness of systems of functions of p^n-valued logic containing operations of addition and multiplication modulo p^n, Metod. Diskret. Anal. 34 (1980) 74–87 (in Russian).

[289] A.A. Nechaev, Criterion for completeness of systems of functions of p^n-valued logic, containing operations of addition and multiplication modulo p^n, Metod. Diskret. Anal. 34 (1980) 74–87 (in Russian).

[290] A.A. Nechaev, Completeness criteria for systems of functions on a finite ring and quasi-Frobenius rings, Siberian Math. J. 23 (1983) 431–441; transl. from Sibirsk. Mat. Zh. 23 (3) (1982) 175–187.

[291] A.A. Nechaev, Similarity of matrices over a commutative Artinian local ring, J. Soviet. Math. 33 (1986) 1221–1237; transl. from Tr. Semin. im. I.G. Petrovskogo 9 (1983) 81–101 (in Russian).

[292] A.A. Nechaev, A characterization of critical finite principal ideal rings, Uspekhi Mat. Nauk 37 (5) (1982) 193–194 (in Russian); English transl. in: Russian Math. Surveys 37 (5) (1982) 184–185.

[293] A.A. Nechaev, Trace function in Galois rings and noise stable codes, in: Proc. of the V All-Union Sympos. on Theory of Rings, Algebras and Modules, Novosibirsk, 1982, p. 97 (in Russian).

[294] A.A. Nechaev, Kerdock code in a cyclic form, Diskret. Mat. 1 (4) (1989) 123–139 (in Russian); English transl. in: Discrete Math. Appl. 1 (4) (1991) 365–384.

[295] A.A. Nechaev, Linear recurring sequences over commutative rings, Diskret. Mat. 3 (4) (1991) 106–127 (in Russian); English transl. in: Discrete Math. Appl. 2 (6) (1992) 659–683.

[296] A.A. Nechaev, Cycle types of linear substitutions over finite commutative rings, Sb. Math. 78 (2) (1994) 283–311; transl. from Mat. Sb. 184 (3) (1993) 21–56.

[297] A.A. Nechaev, Linear recurrent sequences over quasi-Frobenius modules, Russian Math. Surveys 48 (3) (1993) 209–210; transl. from Uspekhi Mat. Nauk 48 (3) (1993) 197–198.

[298] A.A. Nechaev, Finite quasi-Frobenius modules. Applications to codes and linear recurrences, Fundam. Prikl. Mat. 1 (1) (1995) 229–254 (in Russian).

[299] A.A. Nechaev, Linear codes and multilinear recurrences over finite rings and quasi-Frobenius modules, Russian Acad. Sci. Dokl. Math. 52 (3) (1995) 404–407; transl. from Dokl. Akad. Nauk 345 (4) (1995) 451–454.

[300] A.A. Nechaev, Linear codes over modules and over spaces, MacWilliams identity, in: Proc. of the 1996 IEEE Internat. Sympos. Inform. Theory Appl., Victoria, BC, Canada, IEEE, 1996, pp. 35–38.

[301] A.A. Nechaev, Polylinear recurring sequences over modules and quasi-Frobenius modules, in: Y. Fong, et al. (Eds.), First Internat. Tainan–Moscow Algebra Workshop, Proc. Internat. Conference, Tainan, Taiwan, July 23–August 22, 1994, de Gruyter, Berlin, 1996, pp. 283–298.

[302] A.A. Nechaev, A.S. Kuzmin, Linearly presentable codes, in: Proceedings of the 1996 IEEE Internat. Sympos. Inform. Theory Appl., Victoria, BC, Canada, IEEE, 1996, pp. 31–34.

[303] A.A. Nechaev, Recurring sequences, in: D. Krob, et al. (Eds.), Formal Power Series and Algebraic Combinatorics, Proc. 12th International Conference, FPSAC'00, Moscow, Russia, June 26–30, 2000, Springer, Berlin, 2000, pp. 54–66.

[304] A.A. Nechaev, A.S. Kuzmin, V.T. Markov, Linear codes over finite rings and modules, Fundament. Prikl. Mat. 2 (3) (1996) 195–254 (in Russian); English transl. in: CNIT of Moscow State Univ., preprint, 1995-6-1, available at http://www.math.msu.su/~markov.

[305] A.A. Nechaev, Finite Frobenius modules in the theory of linear codes, in: Memoires in Discrete Mathematics, vol. 8, Fizmatlit, Moscow, 2004, pp. 187–215 (in Russian).

[306] A.A. Nechaev, V.N. Tzypyshev, Artinian bimodule with quasi-Frobenius canonical bimodule, in: Proc. Internat. Workshop devoted to 70th Anniversary of Scientific Algebraic Workshop of Moscow State University founded by O.J. Smidt in 1930, Department Moscow State Univ., 2000, pp. 39–40.

[307] A.A. Nechaev, D.A. Mikhailov, Canonical generating system of a monic polynomial ideal over a commutative artinian chain ring, Diskr. Mat. 13 (4) (2001) 3–42 (in Russian); English transl. in: Discrete Math. Appl. 11 (6) (2001) 545–546.

[308] A.A. Nechaev, Finite Frobenius bimodules, in: Internat. Algebra Conf. Dedicated to 250 Anniversary of Moscow University and 75 Anniversary of Department of High Algebra, Moscow State Univ., 2004, pp. 239–241.

[309] A.N. Oleksenko, The basis of identities of the matrix algebra of the second order over $\mathbb{Z}_{p^2}$, Fundam. Prikl. Mat. 6 (2) (2000) 501–531 (in Russian).

[310] A.N. Oleksenko, A basis for the identities of the algebra of second-order matrices over $\mathrm{GR}(p^{2n}, p^2)$, Izv. Altaj. Gos. Univ., Ser. Mat. Inform. Fiz. 2000 (1) (2000) 17–24 (in Russian).

[311] K.R. Pearson, J.E. Schneider, Rings with a cyclic group of units, J. Algebra 16 (2) (1970) 243–251.

[312] R.E. Peinado, On finite rings, Math. Mag. 40 (2) (1967) 83–85.

[313] B. Peirce, Linear associative algebras, Amer. J. Math. 4 (1881) 97–221.

[314] V. Peric, On rings with polynomial identity $x^n - x = 0$, Publ. Inst. Math. 33 (1983) 165–167.

[315] S.V. Polin, Über Identitäten endlicher Algebren, Sibirsk. Mat. Zh. 17 (1976) 1356–1366.

[316] S.V. Polin, Identities of finite algebras, Siberian Math. J. 17 (1977) 992–999. Online ordering link to full text.

[317] G. Pollak, Über die Struktur kommutativer Hauptidealringe, Acta Sci. Math. 22 (1–2) (1961) 62–74.

[318] J. Pomfret, Similarity of matrices over finite rings, Proc. Amer. Math. Soc. 37 (2) (1973) 421–422.

[319] B.A. Probert, Local rings whose maximal ideal is principal as a right ideal, Proc. London Math. Soc. (3) 19 (1969) 403–420.

[320] R. Raghavendran, Finite associative rings, Compos. Math. 21 (2) (1969) 195–220.

[321] R. Raghavendran, A class of finite rings, Compos. Math. 22 (1) (1970) 49–57.

[322] V.A. Ratinov, Semiperfect rings with commutative Jacobson radical, manuscript, VINITI Department, No. 3215-78, 1–46 (in Russian).

[323] V.A. Ratinov, The structure of semiperfect rings with commutative Jacobson radical, Mat. Sb. 110 (152) (3) (1979) 459–470 (in Russian); English transl. in: Math. Sb. 38 (1981) 427–436.

[324] V.A. Ratinov, Finite commutative rings with special types of groups of units, Math. Notes 32 (1983) 886–890 (in English); transl. from Mat. Zametki 32 (6) (1982) 799–807.

[325] V.A. Ratinov, Semiperfect rings with nilpotent adjoint group, Mat. Zametki 29 (2) (1981) 171–180 (in Russian); English transl. in: Math. Notes 29 (1–2) (1981) 90–94.

[326] V.A. Ratinov, Possible values of certain structural parameters of finite local commutative rings, Mat. Sb. 125 (12) (1984) 259–268 (in Russian); English transl. in: Sb. Math. 53 (1986) 261–270.

[327] L. Redei, Über die Ringe mit gegebenen Modul, Acta Math. Acad Sci. Hungar. 5 (1954) 27–28.

[328] L. Redei, Die einstufig nichtkommutativen endlichen Ringe, Acta Math. Acad. Sci. Hungar. 8 (3–4) (1957) 401–442.

[329] A.B. Remizov, Superstructure of the closed class of polynomials modulo k, Diskret. Mat. 1 (1) (1989) 3–15 (in Russian); English transl. in: Discrete Math. Appl. 1 (1) (1991) 9–22.

[330] I.G. Rosenberg, Polynomial functions over finite rings, Glas. Mat. 10 (30) (1975) 25–33.

[331] A. Rosenfeld, A note on two special types of rings, Scripta Math. 28 (1) (1967) 51–54.

[332] I.F. Rua, Primitive and non-primitive finite semifields, Comm. Algebra 32 (2) (2004) 793–803.

[333] A.S. Rybkin, Finite local rings of principal ideals, Mat. Zametki 28 (1) (1980) 3–16 (in Russian); English transl. in: Math. Notes 28 (1981) 465–472.

[334] R.E. Sabin, S.J. Lomonaco, Metacyclic error-correcting codes, Appl. Algebra Engrg. Comm. Comput. 6 (3) (1995) 191–210.

[335] R.E. Sabin, On determining all codes in semi-simple group rings, in: G. Cohen, et al. (Eds.), Proc. of the Applied Algebra, Algebraic Algorithms and Error-Correcting Codes, 10th Internat. Sympos., AAECC-10, San Juan de Puerto Rico, Puerto Rico, May 10–14, 1993, in: Lecture Notes in Comput. Sci., vol. 673, Springer, Berlin, 1993, pp. 279–290.

[336] R.E. Sabin, An ideal structure for some quasi-cyclic error-correcting codes, in: G.L. Mullen, et al. (Eds.), Finite Fields, Coding Theory, and Advances in Communications and Computing, Proc. of the Internat. Conference on Finite Fields, Coding Theory, and Advances in Communications and Computing, University of Nevada, Las Vegas, NV, August 7–10, 1991, in: Lecture Notes Pure Appl. Math., vol. 141, Dekker, New York, 1993, pp. 183–194.

[337] G. Scheffers, Zur Theorie der aus n Haupteinheiten ableiteren höheren complexen Zahlen, Leipzig Ber. Math.-Phys. Klass. (1889) 290–307.

[338] G. Scheffers, Zurückführung complexer Zahlensysteme auf typische Formen, Math. Ann. 39 (1891) 293–390.

[339] V.I. Schneidmüller, About rings with finite chains of subrings, Mat. Sb. 42 (5) (1935) 597–598 (in Russian).

[340] V.I. Schneidmüller, On rings with finite decreasing chains of subrings, Dokl. Acad. Sci. USSR 28 (4) (1940) 579–581 (in Russian).

[341] V.I. Schneidmüller, Infinite rings with finite decreasing chains of subrings, Mat. Sb. 27 (2) (1950) 219–228 (in Russian).

[342] G. Scorza, Le algebre doppie, Rend. R. Acad. Sci. Fis. Mat. Napoli (3) 28 (1922) 65–79.

[343] G. Scorza, Le algebre del 3^o ordine, Mem. Real. Acad. Sci. Fis. Mat. Napoli (2) 20 (13) (1935).

[344] G. Scorza, Le algebre del 4^o ordine, Mem. Real. Acad. Sci. Fis. Mat. Napoli (2) 20 (14) (1935).

[345] P. Shankar, On BCH codes over arbitrary integer rings, IEEE Trans. Inform. Theory 25 (4) (1979) 480–483.

[346] J.B. Shaw, Theory of linear associative algebra, Trans. Amer. Math. Soc. 4 (1903) 251–287.

[347] J.B. Shaw, On nilpotent algebras, Trans. Amer. Math. Soc. 4 (1903) 405–422.

[348] A.I. Shirshov, Some algorithmic problems for Lee algebras, Sibirsk. Mat. Zh. 3 (1962) 292–296 (in Russian).

[349] K. Shoda, Über die Automorphismen einer endlicher Abelscher Gruppe, Math. Ann. 100 (1928) 674–686.

[350] K. Shoda, Über die Einheitengruppe eines endlicher Ringes, Math. Ann. 102 (1930) 273–282.

[351] D. Singmaster, Rings of order four, Amer. Math. Monthly 71 (1964) 918–920.

[352] D. Singmaster, On polynomial functions (mod m), J. Number Theory 6 (1974) 345–352.

[353] L.A. Skornyakov, Left chain rings, in: Collection Dedicated to Memory of N.G. Chebotarev, Kazan Univ., 1964, pp. 75–88 (in Russian).

[354] L.A. Skornyakov, About left chain rings, Izv. Vyssh. Uchebn. Zaved. Mat. 4 (1966) 114–117 (in Russian).

[355] L.A. Skornyakov, About maximally-ideal covering of rings, Math. Notes 7 (3) (1970).

[356] R.E. Smithson, A note on finite Boolean rings, Math. Mag. 37 (5) (1964) 325–327.

[357] E. Snapper, Completely primary rings. I–IV, Ann. of Math. 52 (3) (1950) 666–693; 53 (1) (1951)125–142; 53 (3) (1951) 207–234; 55 (1) (1952) 46–64.

[358] Sono, Memoirs, College Sci. Kyoto 2 (1917).

[359] E. Spiegel, Codes over $\mathbb{Z}_m$, Inform. and Control 35 (1) (1977) 48–51.

[360] E. Spiegel, Codes over $\mathbb{Z}_m$, Inform. and Control 37 (1) (1978) 100–104, revised.

[361] R.P. Stanley, Zero square rings, Pacific J. Math. 30 (3) (1969) 811–824.

[362] E. Steinitz, Rechteckige Systeme und Moduln in algebraischen Zahlkorpern. I, Math. Ann. 71 (3) (1912) 328–354.

[363] W. Streb, Eine Kennzeichnung endlicher nilpotent Ringe, Elem. Math. 28 (3) (1973) 70–71.

[364] R.M. Sultanov, Certain properties of matrices with elements in non-commutative rings, Tr. Sect. Math. Azerb. SSR Sci. Acad. 11 (1946) 11–16 (in Russian).

[365] T. Sumiyahma, Note on maximal Galois subrings of finite local rings, Math. J. Okayama Univ. 21 (1) (1979) 31–32.

[366] D.A. Suprunenko, About conjugate matrices over residue rings, Dokl. Akad. Sci. BSSR 8 (1964) 693–695 (in Russian).

[367] A. Sychowicz, On embedding of finite rings into matrices, Acta Math. Hungar. 46 (3–4) (1985) 273–296.

[368] C. Szabó, V. Vértesi, The complexity of the word-problem for finite matrix rings, Proc. Amer. Math. Soc. 132 (12) (2004) 3689–3695.

[369] F.A. Szasz, Die Ringe ohne Linksideale, Bull. Stii. Bucuresti 1 (1950) 783–789.

[370] T. Szele, Ein Satz über die Strukture der endlichen Ringe, Acta Sci. Math. (Szeged) 11 (4) (1948) 246–250.

[371] T. Szele, On finiteness criterion for modules, Publ. Math. Debrecen 3 (1955) 253–256.

[372] T. Szele, Nilpotent Artinian rings, Publ. Math. Debrecen 4 (1955) 71–78.

[373] J. Szendrei, On the Jacobson radical of a ring, Publ. Math. Debrecen 4 (1–2) (1955) 93–97.

[374] K.-I. Tahara, H. Hirosi, On the circle group of finite nilpotent ring, Proc. of the “Group-Korea 1983”, pp. 161–179.

[375] G. Thierrin, On duo rings, Canad. Math. Bull. 3 (2) (1960) 167–172.

[376] B.R. Toskey, Rings on a direct sum of cyclic groups, Publ. Math. Debrecen 10 (1963) 93–95.

[377] A.I. Uzkow, Abstract foundation of Brandt’s theory of ideals, Mat. Sb. 6 (2) (1939) 253–281.

[378] A.I. Uzkov, Zur Idealtheorie der kommutativen Ringe I, Mat. Sb. 5 (3) (1939) 513–520.

[379] A.I. Uzkov, An algebraic lemma and normalizing Noether theorem, Mat. Sb. 22 (2) (1948) 349–350.

[380] A.I. Uzkov, On cyclic direct decomposability of modules over commutative rings, Mat. Sb. 62 (4) (1963) 469–475.

[381] Van der Waerden, Moderne Algebra. II, Springer, Berlin, 1931.

[382] H.S. Vandiver, Theory of finite algebras, Trans. Amer. Math. Soc. 13 (3) (1912) 293–304.

[383] Y.P. Vasiliev, Calculating methods of description of finite algebras, preprint, Acad. Sci. USSR, Siberian Branch, Irkutsk, 1984.

[384] J.W. Wamsley, On a condition for commutativity of rings, J. London Math. Soc. (2) 4 (1971) 331–332.

[385] W.C. Waterhouse, Rings with cyclic additive group, Amer. Math. Monthly 71 (4) (1964) 449–450.

[386] J.H.R. Wedderburn, A theorem on finite algebras, Trans. Amer. Math. Soc. 6 (1905) 349–352.

[387] J.H.R. Wedderburn, On hypercomplex numbers, Proc. London Math. Soc. (2) 6 (1908) 77–118.

[388] G.P. Wene, On the multiplicative structure of finite division rings, Aequationes Math. 41 (2–3) (1991) 222–233.

[389] G.P. Wene, Semifields of dimension $2n, n \geqslant 3$ over $GF(p^m)$ that have left primitive elements, Geom. Dedicata 41 (1) (1992) 1–3.

[390] J. Wiesenbauer, Über die endlichen Ringe mit gegebener additive Gruppe, Monatsh. Math. 78 (2) (1974) 164–173.

[391] J. Wiesenbauer, Über endlichen p-Ringe vom Rang Zwei, Math. Balkanica 130 (4) (1974) 723–725.

[392] J. Wiesenbauer, Anzahlsatze fur endliche p-Ringe vom Rang Zwei, in: Proc. of the Vienna Conference, Juni 21–24, 1984, in: Contribut. General Algebra, vol. 3, Wienn, 1985, pp. 391–396.

[393] R.S. Wilson, On the structure of finite rings. I, Compos. Math. 26 (1) (1973) 79–93.

[394] R.S. Wilson, On the structure of finite rings. II, Pacific J. Math. 51 (1) (1974) 317–325.

[395] R.S. Wilson, Representations of finite rings, Pacific J. Math. 53 (2) (1974) 643–649.

[396] R. Wisbauer, Grundlagen der Modul- und Ringtheorie, Fischer, München, 1988; translated in: Foundations of Module and Ring Theory, Gordon & Breach, Reading, MA, 1991.

[397] E. Witt, Über die kommutativitat endlicher Schiefkörper, Abh. Math. Semin. Univ. Hamburg 8 (1930) 413.

[398] J.A. Wood, Extension theorems for linear codes over finite rings, in: T. Mora, H. Mattson (Eds.), Applied Algebra, Algebraic Algorithms and Error-Correcting Codes, Springer, Berlin, 1997, pp. 329–340.

[399] J.A. Wood, Duality for modules over finite rings and applications to coding theory, Amer. J. Math. 121 (3) (1999) 555–575.

[400] J.A. Wood, Weight functions and the extension theorem for linear codes over finite rings, in: Finite Fields: Theory, Applications, and Algorithms, Waterloo, ON, 1997, in: Contemp. Math., vol. 225, Amer. Math. Soc., Providence, RI, 1999, pp. 231–243.

[401] P. Wu, A class of finite rings, J. Beijing Normal. Univ. (Natur. Sci.) 2 (1988) 8–11.

[402] A. Yaqub, Elementary proofs of the commutativity of p-rings, Amer. Math. Monthly 64 (1957) 253–254.

[403] O. Zariski, P. Samuel, Commutative Algebra. vols. 1, 2, van Nostrand, Princeton, NJ, 1958.

Section 4A
Lattices and Partially Ordered Sets

This page intentionally left blank

Algebraic and Categorical Aspects of Quantales

David Kruml and Jan Paseka

Department of Mathematics, Masaryk University, Janáčkovo nám. 2*a*, 602 00 *Brno, Czech Republic*
E-mail: kruml@math.muni.cz; paseka@math.muni.cz

Contents

HANDBOOK OF ALGEBRA, VOL. 5
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(07)05006-1

This page intentionally left blank

Preface

This text aims to be a short introduction to some of the basic notions in theory of quantales. We will try to give an up-to-date account of certain aspects on algebraic and categorical properties of quantales and quantale modules. We also add some recent developments involving representations and projectivity in quantales. Due to infinitesimal joins quantales do not form a variety but many of the methods of universal algebra can be still used.

The term *quantale* was suggested by C.J. Mulvey at the Oberwolfach Category Meeting (see [39]) as "a quantization" of the term *locale*. Locales (or *frames*, see [22,65,83] as primary sources of reference), form an order-theoretic counterpart of topological spaces and are therefore able to describe commutative C*-algebras. As is noted in [65], we can think of a locale as of a kind of space, more general than the classical one, allowing us to see topological phenomena in a new perspective. In the C*-algebra case, a commutative C*-algebra A is completely determined by its locale $\mathit{Idl}(A)$ of closed ideals which is order isomorphic to the lattice of open sets of its spectrum.

The main aim of C.J. Mulvey has been to find a substitute of locales which could play the same rôle for general C*-algebras to establish a generalized Gelfand–Naimark duality for all C*-algebras and study non-commutative topology.

Despite that, quantales can be considered as special instances of various concepts of algebra and category theory, namely

- complete multiplicative lattices,
- complete residuated lattices,
- semirings with infinitary sums,
- thin closed monoidal categories,
- monoids in complete semilattices, i.e. monoidal functors $\{*\} \to \mathcal{S}up$,

etc. Quantales are also applied in linear and other substructural logics and automaton theory. An important moment in the development of the theory of quantales was the realization that quantales give a semantics for propositional linear logic in the same way as Boolean algebras give a semantics for classical propositional logic (see [18]).

Quantales arise naturally as lattices of ideals, subgroups, or other suitable substructures of algebras, and then are called *spectra*. Since a quantale can be considered as a semiring, we can built a module theory and study representations of quantales. At least in the case of C*-algebras the construction of spectra yields a functor to quantales and a representation of a C*-algebra defines a representation of its spectrum. This is a good reason to consider representations of quantales as "points" of non-commutative spaces and consider *spatiality* of quantales. In contrast to locale theory, there is a proper class of *endomorphism quantales* which appear as targets of representations. All the endomorphism quantales are *simple* and spatiality of quantales can be equivalently studied by morphisms to simple quantales. In many applications, it may be useful to have a good notion of negation or pseudocomplement. Even that this is generally a non-trivial problem, we submit some positive solution for the case of endomorphism quantales.

The existing literature contains two monographs of K. Rosenthal [75,76] devoted to quantales and quantaloids and three survey papers on quantales [44,56,67].

1. Basic notions and examples

1.1. *Preliminaries*

DEFINITION 1.1.1. By a *sup-lattice* is meant a complete lattice, a *sup-lattice morphism* is a mapping preserving arbitrary joins (also called sup's).

If $f : S \to T$ is a sup-lattice morphism, the assignment

$$f(s) \leqslant t \quad \Leftrightarrow \quad s \leqslant f^{\dashv}(t),$$

explicitly

$$f^{\dashv}(t) = \bigvee \{ s \mid f(s) \leqslant t \},$$

defines a mapping $f^{\dashv} : T \to S$ preserving all meets. This mapping $f^{\dashv}$ is called the *adjoint* of f. If we denote by S^{op} the dual sup-lattice to S then $f^{\dashv} : T^{\mathrm{op}} \to S^{\mathrm{op}}$ is a sup-lattice morphism.

The top element of a sup-lattice is denoted by 1, the bottom element by 0.

DEFINITION 1.1.2. By a *quantale* is meant a sup-lattice Q equipped with an associative multiplication which distributes over joins

$$a\Bigl(\bigvee a_i\Bigr) = \bigvee (aa_i), \qquad \Bigl(\bigvee a_i\Bigr)a = \bigvee (a_i a)$$

for all $a, a_i \in Q$.

DEFINITION 1.1.3. Let Q be a quantale. An element $a \in Q$ is called:

- *idempotent* iff $aa = a$. We write $a \in \mathcal{I}(Q)$.
- *right-sided* (*left-sided*) iff $a1 \leqslant a$ ($1a \leqslant a$). We write $a \in \mathcal{R}(Q)$, respectively $a \in \mathcal{L}(Q)$.
- *strictly right-sided* (*strictly left-sided*) iff $a1 = a$ ($1a = a$).
- *two-sided* if it is both right-sided and left-sided. We write $a \in \mathcal{T}(Q)$.
- a *right* (*left*) *unit* iff $ba = b$ ($ab = b$) for any $b \in Q$.
- a *unit* if it is both a right and left unit.

DEFINITION 1.1.4. A quantale Q is called:

- *commutative* iff for every $a, b \in Q$: $ab = ba$.
- *idempotent* iff every $a \in Q$ is idempotent.
- (*strictly*) *right-sided*/*left-sided* iff every $a \in Q$ is (strictly) right-sided/left-sided.
- *right-idempotent*/*left-idempotent* iff every $a \in Q$ that is right-sided/left-sided is idempotent.
- *two-sided* iff every $a \in Q$ is two-sided.
- (*right*/*left*) *unital* iff Q has a (right/left) unit.

For a unital quantale Q, we shall denote e_Q the unit of Q.

A subset $T \subseteq Q$ is called a *subquantale* of Q if it is closed under all joins and multiplication in Q. It may be remarked that $\mathcal{R}(Q)$, $\mathcal{L}(Q)$ and $\mathcal{T}(Q)$ are subquantales of Q closed under arbitrary meets, and that $Q\mathcal{R}(Q) \subseteq \mathcal{R}(Q)$ and $\mathcal{L}(Q)Q \subseteq \mathcal{L}(Q)$.

DEFINITION 1.1.5. Let Q, K be a quantales. A *quantale morphism* is a sup-lattice morphism $f : Q \to K$ preserving multiplication. If Q and K are unital then a quantale morphism $f : Q \to K$ is said to be *unital* if it preserves the unit, i.e. $f(e_Q) = e_K$, where e_Q and e_K are the respective units of Q and K.

(Unital) quantales and (unital) quantale morphisms form a concrete category of sets with structure denoted by $\mathcal{Q}uant$ ($\mathcal{U}n\mathcal{Q}uant$).

An equivalence relation $\sim$ on Q is said to be a *congruence* if

$$a \sim b \quad \Rightarrow \quad ca \sim cb, ac \sim bc, \qquad a_i \sim b_i \quad \Rightarrow \quad \bigvee a_i \sim \bigvee b_i$$

for all $a, b, c, a_i, b_i \in Q$. One can easily check that congruences are exactly coset equivalences given by quantale morphisms.

LEMMA 1.1.6. *If $f : Q \to K$ is a surjective semigroup morphism and Q is a monoid with a unit e_Q, K is also a monoid with $e_K = f(e_Q)$ as a unit.*

PROOF. Let $b \in K$. Since f is a surjection, there exists $a \in Q$ such that $f(a) = b$. Then $b = f(a) = f(e_Q a) = f(e_Q)f(a) = f(e_Q)b$ and similarly for the unit on the right. □

DEFINITION 1.1.7. The actions $a\_$, $\_a$ for fixed $a \in Q$ determine sup-lattice endomorphisms on Q.[1] Their adjoints are denoted by $\_ \leftarrow a$, $a \to \_$, respectively. That is,

$$a \leqslant b \to c \quad \Leftrightarrow \quad ab \leqslant c \quad \Leftrightarrow \quad b \leqslant c \leftarrow a.$$

The operations $\to$, $\leftarrow$ are called *residuations*.

For any $a \in Q$, we have $a \leftarrow 0 \in \mathcal{R}(Q)$ and $a \to 0 \in \mathcal{L}(Q)$. A quantale Q is called *von Neumann* if

$$0 \leftarrow (r \to 0) = r \quad \text{for any } r \in \mathcal{R}(Q)$$

and

$$(0 \leftarrow l) \to 0 = l \quad \text{for any } l \in \mathcal{L}(Q).$$

In any von Neumann quantale Q we have $\mathcal{L}(Q) \simeq \mathcal{R}(Q)^{\mathrm{op}}$.

A quantale Q is said to be a *factor* if $\mathcal{T}(Q) = \{0, 1\}$.

A set $P \subseteq Q$ is said to be *residually closed* if $a \to p \in P$, $p \leftarrow a \in P$ for every $a \in Q$, $p \in P$.

A set $P \subseteq Q$ is said to be *separating* if for every $a, b \in Q$, $a \not\geqslant b$ there is an element $p \in P$ such that $a \leqslant p$, $b \not\leqslant p$. Note that then every element of Q is a meet of elements from P.

A nonempty set P, $1 \notin P$ is called *cyclic* if $\{a \to p \leftarrow b\colon a, b \in Q\} = P \cup \{1\}$ for every $p \in P$. In fact, any quantale having a cyclic set is non-trivial.

The next lemma can be proved by elementary calculations.

[1] Here $a\_$ is multiplication of the left $x \mapsto ax$, $x \in Q$ and similarly $\_a$ is multiplication on the right.

LEMMA 1.1.8. *(See [75].) Let Q be a quantale, $a, b, c, a_i \in Q$. Then*

$$
\begin{aligned}
&(b \to c) \leftarrow a = b \to (c \leftarrow a),\\
&a \to (b \to c) = (ab) \to c, \qquad (c \leftarrow a) \leftarrow b = c \leftarrow (ab),\\
&\Big(\bigvee a_i\Big) \to c = \bigwedge (a_i \to c), \qquad c \leftarrow \Big(\bigvee a_i\Big) = \bigwedge (c \leftarrow a_i).
\end{aligned}
$$

DEFINITION 1.1.9. By an *involutive quantale* (shortly **-quantale*) is meant a quantale Q with a unary operation of *involution* such that

$$a^{**} = a, \qquad (ab)^* = b^* a^*, \qquad \Big(\bigvee a_i\Big)^* = \bigvee a_i^*$$

for all $a, b, a_i \in Q$.

An element $a \in Q$ is called *Hermitean* if $a^* = a$, the set of Hermitean elements is denoted by $\mathcal{H}(Q)$.

A *-quantale Q is said to be an **-factor* if $\mathcal{T}(Q) \cap \mathcal{H}(Q) = \{0, 1\}$, i.e. if 0 and 1 are the only Hermitean two-sided elements of Q.

By an *involutive quantale morphism* (shortly **-morphism*) is meant a quantale morphism of *-quantales which also preserves the involution.

(Unital) *-quantales and (unital) *-morphisms form a concrete category of sets with structure denoted by $\mathcal{I}n\mathcal{Q}uant$ ($\mathcal{I}n\mathcal{U}n\mathcal{Q}uant$).

A quantale congruence $\sim$ on a *-quantale Q is said to be an *involutive congruence* if

$$a \sim b \quad \Rightarrow \quad a^* \sim b^*$$

for all $a, b \in Q$.

Note that $(a \to b)^* = b^* \leftarrow a^*$.

A set $P \subseteq Q$ is said to be **-separating* if $P \cup P^*$ is separating (where $P^* = \{p^*\colon p \in P\}$).

DEFINITION 1.1.10. Let Q be a quantale. By a left (right) Q-module is meant a sup-lattice M with an action $Q \times M \to M$ ($M \times Q \to M$) such that

$$
\begin{array}{c}
a\Big(\bigvee m_i\Big) = \bigvee (am_i),\\
\Big(\bigvee a_i\Big)m = \bigvee (a_i m),\\
a(bm) = (ab)m
\end{array}
\qquad
\left(
\begin{array}{c}
\Big(\bigvee m_i\Big)a = \bigvee (m_i a),\\
m\Big(\bigvee a_i\Big) = \bigvee (ma_i),\\
(mb)a = m(ba)
\end{array}
\right)
$$

for all $a, b, a_i \in Q, m, m_i \in M$.

A module M over a unital quantale Q is said to be unital if $e_Q m = m$ for all $m \in M$.

The adjoint of the module action is called a residuation as well and denoted by $\leftarrow$ ($\to$). One can easily verify similar properties of the residuation as in 1.1.8, from whose it follows that M^{op} with the residuation is a right (left) Q-module. M^{op} is called the *dual module* of M.

A mapping $f : M \to N$ of two left (right) Q-modules M, N is said to be a *Q-module morphism* if f is a sup-lattice morphism and $f(am) = af(m)$ ($f(ma) = f(m)a$) for every $a \in Q, m \in M$. Let us recall that the adjoint $f^{\dashv} : N^{\mathrm{op}} \to M^{\mathrm{op}}$ of a Q-module morphism $f : M \to N$ is a Q-module morphism of dual modules (cf. [1]). Two special cases of the module morphism are a submodule inclusion and a quotient mapping. The adjoint of inclusion is a quotient mapping and vice versa.

(Unital) Q-modules and Q-module morphisms form a concrete category of sets with structure denoted by $Q\text{-}\mathcal{M}od$ ($Q\text{-}\mathcal{U}n\mathcal{M}od$).

Throughout the paper we do not distinguish between left and right modules if it is not important or if the meaning is clear. We write 1_Q and 1_M for the quantale and module top, respectively, whenever a confusion is imminent.

We will close this section with an easy, but important lemma (see [54]).

LEMMA 1.1.11. *Let Q be a (non-unital) quantale, $Q[e] = \{a \vee \varepsilon \colon a \in Q, \varepsilon \in \{0, e\}\}$, e arbitrary such that $e \notin Q$. Then $Q[e]$ is both an unital quantale with unit e and a Q-module and we have a quantale embedding $i_e : Q \to Q[e]$. Moreover, if Q is an involutive quantale we have that $Q[e]$ is involutive as well.*

PROOF. Note that we shall identify the formal join $a \vee 0$ with a for any $a \in Q$, $0 \vee e$ with e. We may then define the supremum

$$\bigvee_i (a_i \vee \varepsilon_i) = \begin{cases} (\bigvee_i a_i) \vee e, & \text{if } \exists i \ \varepsilon_i = e, \\ \bigvee_i a_i, & \text{otherwise.} \end{cases}$$

Note that $Q[e]$ is isomorphic as a complete lattice with the Cartesian product of Q and a 2-element lattice.

Similarly, we may define the multiplication on $Q[e]$ as follows:

$$(a \vee \varepsilon') \cdot (b \vee \varepsilon'') = \begin{cases} a \cdot_Q b, & \text{if } \varepsilon' = \varepsilon'' = 0, \\ a \cdot_Q b \vee_Q b, & \text{if } \varepsilon' = e, \varepsilon'' = 0, \\ a \cdot_Q b \vee_Q a, & \text{if } \varepsilon' = 0, \varepsilon'' = e, \\ (a \cdot_Q b \vee_Q a \vee_Q b) \vee e, & \text{if } \varepsilon' = \varepsilon'' = e. \end{cases}$$

It is an easy task to check that $Q[e]$ is a unital quantale with the unit e. The embedding $i_e : Q \to Q[e]$ is defined as follows:

$$i_e(a) = a \vee 0 = a$$

for all $a \in Q$. Now assume that Q is involutive. Then we shall define the involution * on $Q[e]$ as follows:

$$(a \vee \varepsilon)^* := a^{*_Q} \vee \varepsilon$$

for all $a \in Q$ and $\varepsilon \in \{0, e\}$. Again, it is evident that * satisfies $x^{**} = x$, $x^* \cdot y^* = (y \cdot x)^*$ for all $x, y \in Q[e]$ and preserves arbitrary suprema. Moreover, the embedding $i_e : Q \to Q[e]$ preserves the involution. □

Note that the above construction corresponds to the extension of a non-unital C*-algebra to a unital C*-algebra (see [61]).

1.2. *Examples of quantales and quantale modules*

EXAMPLE 1.2.1 *(The power set of a semigroup).* Let $(S,\cdot)$ be a semigroup and $\mathcal{P}(S)$ the set of all its subsets. Then $\mathcal{P}(S)$ is a complete lattice and a multiplication can be defined on $\mathcal{P}(S)$ by $UV = \{u \cdot v\colon\ u \in U,\ v \in V\}$ for all $U, V \in \mathcal{P}(S)$. Evidently, the quantale $\mathcal{P}(S)$ is commutative (unital) iff S is commutative (a monoid).

In fact, this is the most general example giving rise to any quantale as will be shown below.

EXAMPLE 1.2.2 *(Formal languages).* Consider a (finite) set A and the free semigroup over A denoted by A^+. The elements of A are usually called letters (in this context) and the elements of A^+, called words, are usually written as finite sequences of letters using no parentheses and operation symbols. Admitting the empty sequence – the empty word ε, A^+ becomes a monoid denoted by A^*. Subsets of A^* are then called languages over an alphabet A. The set $\mathcal{P}(A^*)$ of all languages over A is a quantale again with the top element A^* and the unit $e = \{\varepsilon\}$.

EXAMPLE 1.2.3 *(Frames).* Any frame L is a unital, involutive, commutative and idempotent quantale with the multiplication given by the meet $\wedge$, with unit given by the top element 1 and $* = \mathrm{id}_L$.

EXAMPLE 1.2.4 *(Relations).* The set $\mathcal{R}el(X)$ of relations on a set X, ordered by inclusion, is a unital *-quantale with multiplication given by the composition of relations

$$T_1 \cdot T_2 = \bigl\{(x,z) \in X \times X\colon\ \exists y \in X (x,y) \in T_1 \text{ and } (y,z) \in T_2\bigr\},$$

with the equality relation providing the unit and the involution defined as the inverse relation

$$T^* = \bigl\{(x,y) \in X \times X\colon\ (y,x) \in T\bigr\}.$$

The right-sided elements of $\mathcal{R}el(X)$ are those relations R_A, for $A \subseteq X$, given by

$$R_A = (X - A) \times X.$$

Similarly, the left-sided elements of $\mathcal{R}el(X)$ are relations L_B, for $B \subseteq X$, given by

$$L_B = X \times B.$$

Note that $R_A^* = L_{X-A}$. Moreover, there is a well-known bijective correspondence between relations on X and sup-preserving maps from the power set $\mathcal{P}(X)$ to itself. Namely, for all $T \in \mathcal{R}el(X)$ and all $\phi : \mathcal{P}(X) \to \mathcal{P}(X)$, $T \mapsto \phi_T$, $\phi \mapsto R_\phi$ where

$$\phi_T(A) = \bigl\{x \in X\colon\ \exists y \in A \text{ such that } (x,y) \in T\bigr\},$$
$$R_\phi = \bigl\{(x,y) \in X \times X\colon\ x \in \phi\bigl(\{y\}\bigr)\bigr\}.$$

EXAMPLE 1.2.5 *(The lattice of all sup-lattice endomorphisms).* For any sup-lattice S, we will denote by $\mathcal{Q}(S)$ the sup-lattice of all sup-lattice endomorphisms $f : S \to S$ (with the pointwise ordering $f \leqslant g$ iff $f(x) \leqslant g(x)$ for all $x \in S$). Then $\mathcal{Q}(S)$ is a unital quantale; multiplication is composition and the unit is id_S. Moreover, $\mathcal{Q}(S)$ is a *-quantale provided that S is *a sup-lattice with duality*, i.e. with a unary operation $'$ such that

$$s'' = s,$$
$$\left(\bigvee_{i \in I} s_i\right)' = \bigwedge_{i \in I} s_i'$$

for all $s, s_i \in S$, $i \in I$. The involution * on $\mathcal{Q}(S)$ is then given by

$$\phi^*(s) = \left(\bigvee_{\phi(t) \leqslant s'} t\right)'$$

(see [62]). Following Example 1.2.4, $\mathcal{R}el(X) \cong \mathcal{Q}(\mathcal{P}(X))$ for any set X.

The right-sided elements of $\mathcal{Q}(S)$ are precisely mappings

$$\rho_s(t) = \begin{cases} 0, & t = 0, \\ s, & \text{otherwise,} \end{cases}$$

where $s \in S$. Similarly, the left-sided elements are precisely mappings

$$\lambda_s(t) = \begin{cases} 0, & t \leqslant s, \\ 1, & \text{otherwise.} \end{cases}$$

Thus $S \cong \mathcal{R}(\mathcal{Q}(S)) \cong \mathcal{L}(\mathcal{Q}(S))^{\mathrm{op}}$. If S is equipped with a duality then $\mathcal{R}(\mathcal{Q}(S)) \cong \mathcal{L}(\mathcal{Q}(S))$. Moreover, $\rho_s^* = \lambda_{s'}$ and $\lambda_s^* = \rho_{s'}$. Since

$$(\lambda_u \circ \rho_v)(t) = \begin{cases} 0, & \text{if } t = 0 \text{ or } v \leqslant u \\ 1, & \text{otherwise,} \end{cases}$$

we have that $\rho_s' = 0 \leftarrow \rho_s^*$, $\rho_{s'} = 0 \leftarrow \lambda_{s'}$, i.e. the sup-lattice isomorphism between S and $\mathcal{R}(\mathcal{Q}(S))$ preserves the duality.

It may be remarked that $\rho_s \circ \rho_s = \rho_s$ and $\lambda_s \circ \lambda_s = \lambda_s$, i.e. $\mathcal{Q}(S)$ is right-idempotent and left-idempotent.

The sup-lattice $\mathcal{Q}(S)$ can be described by means of a tensor product of sup-lattices

$$\mathcal{Q}(S) \cong \left(S \otimes S^{\mathrm{op}}\right)^{\mathrm{op}}$$

(see [23]). This isomorphism is given by

$$f \mapsto \bigvee_{f(t) \leqslant s} t \otimes s.$$

The elements $\rho_s \vee \lambda_t$ correspond to $t \otimes s$. If S is a sup-lattice with duality then the involution in $\mathcal{Q}(S)$ is induced by

$$(t \otimes s)^* = s' \otimes t'.$$

EXAMPLE 1.2.6 *(The lattice of left ideals of a ring).* Let R be a ring. A subset $A \subseteq R$ is a left ideal of R, if it is a subgroup of $(R, +)$ and for every $r \in R$, $a \in A$: $ra \in A$. Then the set of left ideals of a ring R denoted by $LIdl(R)$ forms a quantale with joins as ideals generated by the union of ideals and multiplication realized as the product of two ideals given by: $A \cdot B = \langle\{ab\colon\ a \in A,\ b \in B\}\rangle = \{a_1b_1 + \cdots + a_nb_n \mid a_i \in A,\ b_i \in B,\ 1 \leqslant i \leqslant n\}$. Similarly, the sets $RIdl(R)$ and $Idl(R)$ of right ideals or two-sided ideals of R are quantales as well. Obviously, all these three notions coincide when R is commutative.

Since a left ideal is closed under left multiplication by elements of R, $LIdl(R)$ is left-sided. It is left unital with $e_l = R$ with R also serving as the top element.

By the definition of residuation, $C \leftarrow A = \{b\colon\ A \cdot \{b\} \subseteq C\}$ and similarly $B \rightarrow C = \bigcup\{a\colon\ \{a\} \cdot B \subseteq C\}$.

EXAMPLE 1.2.7 *(Function modules).* Let X be a set and Q be a quantale. The set $Q^X = \{m : X \rightarrow Q\}$ equipped with the pointwise ordering and multiplication given as $(am)(x) = am(x)$ is evidently a Q-module. Its elements are sometimes called *vectors*.

It can be easily seen that Q^X is a unital Q-module if Q is a unital quantale. In that case, for every $x \in X$ we can define a map $\{x\} : X \rightarrow Q$ such that

$$\{x\}(y) = \begin{cases} e, & x = y, \\ 0, & \text{otherwise.} \end{cases}$$

These vectors are then called *unit vectors*. Since any $h\colon X \rightarrow Q$ can be uniquely expressed as $\bigvee(a_x\{x\})$ where $a_x = h(x)$, we can say that unit vectors $\{x\}$ form a basis of Q^X. Note that any constant map $x \mapsto a$ can be thus expressed as $\bigvee_{x \in X}(a\{x\}) = a \bigvee_{x \in X}\{x\}$.

1.3. *Notes on Section 1*

As indicated in the Preface, the study of ring-theoretical properties by means of complete lattices equipped with an associative multiplication was initiated by W. Krull [28] and developed by M. Ward and R.P. Dilworth [14,84–86]. Later on, S. Niefield and K.I. Rosenthal studied the ideal theory of rings along these lines, first in the commutative case and they began also the systematic study of the non-commutative case.

The theory of strictly right-sided idempotent quantales as a tool for studying non-commutative C*-algebras as proposed by Mulvey [38,39], was started by him and his student M. Nawaz [41], and F. Borceux, G. van den Bossche and J. Rosický [11]. The ideas of J. Rosický [77] on the topological and logical consequences of quantales influenced the consideration of involutive quantales. The need to introduce functoriality on the category of C*-algebras culminated in the description of the spectrum Max(A) of a C*-algebra A as the quantale of closed linear subspaces of A.

Investigations by C.J. Mulvey and J.W. Pelletier [42,43] elucidated the noncommutative topological structure of the spectrum Max A, which was shown to have quantal points that indeed correspond to the equivalence classes of irreducible representations of the C*-algebra A on a Hilbert space. Moreover, the spectrum Max(A) was shown to be a quantal space in an appropriate sense [43]. Independently, work by D. Kruml, J. Paseka, J.W. Pelletier, P. Resende, J. Rosický and others [29,32,51,56,62,63,70] in a number of combinations investigated other approaches to these concepts, examining in particular the properties of the spectral functor on the category of C*-algebras.

On the other hand, quantales have arisen in analysing the semantics of linear and other substructural logics [18,26,27,34,47,87].

The categorical approach to quantales starts with the work of A. Joyal and M. Tierney [23] (here commutative quantales are viewed as commutative monoids in the category of sup-lattices) and continues in the papers of R.P. Gylys [19,20] and in the monograph of K.I. Rosenthal [76].

Since quantales are mostly viewed as both a sup-lattice and a semigroup such that the multiplication distributes over joins one immediately obtains the notion of idempotency, unitality or commutativity. The notions of (strict) right(left)-sidedness are translated from the ideal theory of rings. The notion of a subquantale corresponds to the notion of a subalgebra or more general to a subobject. Morphisms in the category of (unital) quantales are both sup-lattice and (monoid) semigroup morphisms. Therefore congruences on quantales are both sup-lattice and semigroup congruences. Since a surjective semigroup morphism from a monoid to a semigroup (Lemma 1.1.6) preserves (and creates) congruences, the congruences on unital quantales coincide with the usual quantale congruences. A smooth passage from non-unital quantales to unital ones is provided by Lemma 1.1.11.

One of the remarkable properties of quantales is that the multiplication creates two operations of residuation $\rightarrow$ and $\leftarrow$. They give rise to a kind of pseudocomplementation on right-sided and left-sided elements. In the case of von Neumann quantales this pseudocomplementation induces a duality between right-sided and left-sided elements. The notions of factor, residually closed set, separating set and cyclic set play an important role in the characterization of simple objects in the category of quantales.

Involutive quantales (*-quantales) are quantales equipped with a semigroup operation of involution the distributes over joins. The morphisms in the category of (unital) involutive quantales are both sup-lattice and (unital) quantale morphisms that preserve the involution. Similarly, a *-congruence on an involutive quantale is a quantale congruence that preserves the involution.

A prominent role in the theory of quantales and their representations is played by quantale modules. They are sup-lattices equipped with an associative and join-distributive (in each variable separately) action of a quantale.

All the examples in Section 2 except Example 1.2.5 are taken from [75,76]. In the special case when S is an orthocomplemented sup-lattice and $s' = s^{\perp}$ for $s \in S$ (i.e. duality is given by orthocomplements), the calculations were done in C.M. Mulvey and J.W. Pelletier [40], the completely analogous general case is contained in [62,51].

2. Basic algebraic and categorical properties of quantales and quantale modules

2.1. *Free quantales and free quantale modules*

THEOREM 2.1.1. *The categories $\mathcal{Q}uant$, $\mathcal{U}n\mathcal{Q}uant$, $\mathcal{I}n\mathcal{Q}uant$, $\mathcal{I}n\mathcal{U}n\mathcal{Q}uant$ have free objects over $\mathcal{S}et$.*

PROOF. Let X be a set. Then $\mathcal{P}(X^+)$ is the free quantale over X and $\mathcal{P}(X^*)$ is the free unital quantale over X. Similarly, the free (unital) involutive quantale over X is the powerset of a free involutive semigroup (monoid) over X. We will now show that $\mathcal{P}(X^+)$ is a free quantale over X. The other cases go through the same way. By Example 1.2.1, we know that the powerset of a semigroup is a quantale. Now, let Q be a quantale and $f : X \to Q$ be a map. Define a map $i_X : X \to \mathcal{P}(X^+)$ as $i_X(x) = \{x\}$ and a map $\overline{f} : \mathcal{P}(X^+) \to Q$ as $\overline{f}(A) = \bigvee_{a\in A}(f(a_1)\cdot f(a_2)\cdot \dots \cdot f(a_n))$ where $a_1a_2\dots a_n = a \in A$.

We have to verify that $\overline{f}$ is a quantale morphism:

- $\bigvee_i \overline{f}(A_i) = \bigvee_i(\bigvee_{a\in A_i}(f(a_1)\cdot \dots \cdot f(a_n))) = \bigvee_{a\in \bigvee_i A_i}(f(a_1)\cdot \dots \cdot f(a_n)) = \overline{f}(\bigvee_i A_i)$,
- $\overline{f}(A)\cdot\overline{f}(B) = \bigvee_{a\in A}(f(a_1)\cdot \dots \cdot f(a_n))\cdot\bigvee_{b\in B}(f(b_1)\cdot \dots \cdot f(b_m)) = \bigvee_{a\in A,\ b\in B}(f(a_1)\cdot \dots \cdot f(a_n)\cdot f(b_1)\cdot \dots \cdot f(b_m)) = \bigvee_{c\in A\cdot B}(f(c_1)\cdot \dots \cdot f(c_k)) = \overline{f}(A\cdot B)$

Since $(\overline{f}\circ i_X)(x) = \overline{f}(\{x\}) = f(x)$, it is indeed the case that $\overline{f}\circ i_X = f$.

Let $h : \mathcal{P}(X^+) \to Q$ be another morphism that satisfies $h \circ i_X = f$. Then $h(A) = h(\bigvee_{a\in A}\{a\}) = h(\bigvee_{a\in A}\{a_1\dots a_n\}) = \bigvee_{a\in A}(h(\{a_1\dots a_n\}) = \bigvee_{a\in A}(h(\{a_1\})\cdot \dots \cdot h(\{a_n\})) = \bigvee_{a\in A}(f(a_1)\cdot \dots \cdot f(a_n)) = \overline{f}(A)$ and $\overline{f}$ is hence unique. □

THEOREM 2.1.2. *The categories Q-$\mathcal{M}od$, Q-$\mathcal{U}n\mathcal{M}od$ have free objects over $\mathcal{S}et$.*

PROOF. Let Q be a quantale. For a set X, let $Q[e]^X$ be the set of maps $h : X \to Q[e]$ equipped with the pointwise structure (see also Example 1.2.7). Evidently, $Q[e]^X$ is both a Q-module and a quantale. Define a map $\eta_X : X \to Q[e]^X$ as $\eta_X(x) = \{x\}$.

Then every map $f : X \to UM$ has a unique factorization $f = U\overline{f}\circ\eta_X$ where $\overline{f}(h) = \bigvee_{x\in X} h(x)f(x)$. Evidently, $\overline{f}$ is a Q-module morphism. Conversely, for any Q-module morphism $g : Q[e]^X \to M$ we have a unique map $\hat{g} : X \to UM$ such that $\hat{g} = g\circ\eta_X$.

Let Q be a unital quantale. Replacing $Q[e]$ by Q, we can show by the same arguments as above that Q^X is a free unital Q-module. □

2.2. *Algebraicity of quantales and quantale modules*

Recall that a category is called *algebraic* if it is monadic over $\mathcal{S}et$, and *equationally presentable* if its objects can be prescribed by (a proper class of) operations and equations.

The following propositions lists some important properties of algebraic and equationally presentable categories.

PROPOSITION 2.2.1. *Let $\mathcal{K}$ be an equationally presentable category. Then*

1. *$\mathcal{K}$ has all small limits, and they are constructed exactly as in $\mathcal{S}et$ (i.e. they are preserved by the forgetful functor U from $\mathcal{K}$ into $\mathcal{S}et$).*
2. *If the forgetful functor U has a left adjoint, then $\mathcal{K}$ is algebraic, and has all small colimits.*
3. *The monomorphisms in $\mathcal{K}$ are exactly the injective morphisms.*
4. *Epimorphisms in $\mathcal{K}$ need not be surjective; but the regular epis in $\mathcal{K}$ are exactly the surjective morphisms.*
5. *Every morphism in $\mathcal{K}$ can be factored (uniquely up to isomorphism) as a regular epimorphism followed by a monomorphism.*

PROOF. See [37, Chapter 1]. □

THEOREM 2.2.2. *All the categories $\mathcal{Q}uant$, $\mathcal{U}n\mathcal{Q}uant$, $\mathcal{I}n\mathcal{Q}uant$, $\mathcal{I}n\mathcal{U}n\mathcal{Q}uant$, Q-$\mathcal{M}od$ and Q-$\mathcal{U}n\mathcal{M}od$ are algebraic.*

PROOF. All above categories are clearly equationally presentable. From the preceding section we have corresponding free functors from $\mathcal{S}et$. By Proposition 2.2.1 we are done. □

2.3. *Congruences and nuclei in quantales and quantale modules*

DEFINITION 2.3.1. Let S be a sup-lattice, $j : S \to S$ an operator on S satisfying $s \leqslant j(s)$, and $s \leqslant t$ implies $j(s) \leqslant j(t)$ for all $s, t \in S$. We say that j is an *order prenucleus* on S. We put $S_j = \{s \in S\colon\ j(s) = s\}$. Evidently, S_j is a closure system in S, that is, closed under arbitrary meets in S. Write $\bigvee_j = j \circ \bigvee$. We let $\nu(j)$ be the associated closure operator, so that $\nu(j)(s) = \bigwedge\{t \in S_j\colon\ t \geqslant s\}$ is the smallest *order nucleus* (idempotent order prenucleus) greater than j. S_j is then a sup-lattice with the join $\bigvee_j$.

Now let $f : S \to T$ be a sup-lattice morphism. The composite $j = f^{\dashv} \circ f : S \to S$ is an order preserving map and it has a monad structure on it, i.e. it is an order nucleus.

The subsequent lemma explains the usefulness of prenuclei.

LEMMA 2.3.2. *Let j be an order prenucleus on a sup-lattice S, and let $f : S \to T$ a sup-lattice morphism such that $f \circ j = f$. Then $f \circ \nu(j) = f$.*

PROOF. Let $a \in S$ and let us set $W = \{x \in S\colon\ f(x) = f(a), x \geqslant a\}$. Evidently $a \in W$ because $f(a) = f(a)$ and W is a downset in $\uparrow a$ since f is order-preserving. Further, for $x \in W$, we have that $f(a) = f(x) = f(j(x))$, showing that W is j-stable. Finally, if $V \subseteq W$, we have that $f(a) = f(v)$ for all $v \in V$, hence $\bigvee V \in W$. It follows now that $a \leqslant s = \bigvee W \leqslant j(\bigvee W) \in W$. Consequently, $a \leqslant \nu(j)(a) \leqslant j(s) = s \in S_j$. Hence $f(\nu(j)(a)) = f(a)$. □

DEFINITION 2.3.3. Let Q be a quantale, $j : Q \to Q$ an order prenucleus on Q. We say that j is a *quantale prenucleus* on Q if $j(a)j(b) \leqslant j(ab)$ for all $a, b \in Q$. An idempotent quantale prenucleus is called (*quantale*) *nucleus* on Q. Similarly, let M be a Q-module, $k : M \to M$ an order prenucleus on M. We say that k is a *module prenucleus* on M if $ak(m) \leqslant k(am)$ for all $a \in Q$, $m \in M$. An idempotent module prenucleus is called *module nucleus* on Q.

PROPOSITION 2.3.4. *Let Q be a quantale, let M be a Q-module, let $j : Q \to Q$ a quantale prenucleus, and let $k : M \to M$ a module prenucleus. Then $\nu(j)$ is a nucleus and $\nu(k)$ is a module nucleus.*

PROOF. Let $a, b \in Q, m \in M$. The maps $f, g : Q \to Q_j$ defined by $f(x) = \nu(j)(ax)$ and $g(y) = \nu(j)(yb)$ are sup-preserving, $f(j(x)) = \nu(j)(aj(x)) \leqslant \nu(j)(j(a)j(x)) \leqslant \nu(j)(j(ax)) = f(x) \leqslant f(j(x))$ and similarly $g(j(y)) = g(y)$. By Lemma 2.3.2 we have that $f(\nu(j)(x)) = f(x)$ and $g(\nu(j)(y)) = g(y)$. Hence $a\nu(j)(x) \leqslant \nu(j)(a\nu(x)) = \nu(j)(ax)$ and $\nu(j)(y)b \leqslant \nu(j)(\nu(j)(y)b) = \nu(j)(yb)$. Consequently, $\nu(j)(a)\nu(j)(b) \leqslant \nu(j)(\nu(j)(a)b) \leqslant \nu(j)(\nu(j)(ab)) = \nu(j)(ab)$.

The module case follows by the analogous considerations applied to the map $h : M \to M_k$, h being defined by $h(m) = \nu(k)(am)$. □

DEFINITION 2.3.5. Let Q be a *-quantale, $j : Q \to Q$ a quantale prenucleus on Q. We say that j is an *involutive quantale prenucleus* on Q if $j(a)^* \leqslant j(a^*)$ for all $a \in Q$. An idempotent involutive quantale prenucleus is called (*quantale*) *involutive nucleus* on Q.

PROPOSITION 2.3.6. *Let Q be an involutive quantale, $j : Q \to Q$ an involutive quantale prenucleus. Then $\nu(j)$ is an involutive nucleus.*

PROOF. By 2.3.4, $\nu(j)$ is a quantale nucleus. Let $a \in Q$. The map $f : Q \to Q_j$ defined by $f(x) = \nu(j)(x^*)$ is sup-preserving, $f(j(x)) = \nu(j)(j(x)^*) \leqslant \nu(j)(j(x^*)) = f(x) \leqslant f(j(x))$. Consequently, $(\nu(j)(x))^* \leqslant \nu(j)((\nu(j)(x))^*) = \nu(j)(x^*)$. □

PROPOSITION 2.3.7. *Let $j : Q \to Q$ be a nucleus. Then for every $a, b \in Q$ we have $j(a \cdot b) = j(j(a) \cdot b) = j(a \cdot j(b)) = j(j(a) \cdot j(b))$.*

PROOF. Since j is a closure operator, $j(a \cdot b) \leqslant j(j(a) \cdot b) \leqslant j(j(a) \cdot j(b)) \leqslant j(j(a \cdot b)) = j(a \cdot b)$ and similarly $j(a \cdot b) \leqslant j(a \cdot j(b)) \leqslant j(j(a) \cdot j(b)) = j(a \cdot b)$. □

PROPOSITION 2.3.8. *If $j : Q \to Q$ is a nucleus, the set $Q_j = \{x \in Q\colon\ j(x) = x\}$ is a quantale with the multiplication given by $a \cdot_j b := j(ab)$ and $j : Q \to Q_j$ is then a quantale morphism.*

PROOF. We know that Q_j is a sup-lattice and $\cdot_j$ is clearly associative. It remains to prove the distributivity. Take $a \in Q_j$ and $b_i \in Q_j$, $i \in I$. Then $a \cdot_j \bigvee_j b_i = a \cdot_j j(\bigvee b_i) = j(a \bigvee b_i) = j(\bigvee(ab_i)) \leqslant j(\bigvee j(ab_i)) = j(\bigvee_j(a \cdot_j b_i)) = \bigvee_j(a \cdot_j b_i)$.

Since we have, for every $i \in I$, the inequality $a \cdot_j b_i \leqslant a \cdot_j \bigvee_j b_i$ we get $\bigvee_j(a \cdot_j b_i) \leqslant a \cdot_j \bigvee_j b_i$ and therefore we obtain the equality. □

Now, we shall establish a one-to-one correspondence between (order, module) nuclei and (sup-lattices, module) congruences.

DEFINITION 2.3.9. Let S, T be sup-lattices, $f : S \to T$ a sup-lattice morphism. Then the *kernel* of f, written $\ker(f)$, is defined by

$$\ker(f) = \bigl\{(a, b) \in S \times S\colon\ f(a) = f(b)\bigr\}.$$

PROPOSITION 2.3.10. *Let S, T be sup-lattices, $f : S \to T$ a sup-lattice morphism. Then* $\ker(f)$ *is a congruence of sup-lattices. Moreover, we have*

1. *If f is a morphism of quantales then* $\ker(f)$ *is a congruence of quantales.*
2. *If f is a morphism of quantale modules then* $\ker(f)$ *is a congruence of quantale modules.*

PROOF. Let $(a_i, b_i) \in \ker(f)$ for $i \in I$. Then $f(\bigvee_{i\in I} a_i) = \bigvee_{i\in I} f(a_i) = \bigvee_{i\in I} f(b_i) = f(\bigvee_{i\in I} b_i)$; hence $(\bigvee_{i\in I} a_i, \bigvee_{i\in I} b_i) \in \ker(f)$. Clearly $\ker(f)$ is an equivalence relation, so it follows that $\ker(f)$ is actually a congruence of sup-lattices.

Now, let f be a quantale morphism. Then $\ker(f)$ is both a congruence of semigroups and sup-lattices, i.e. a congruence of quantales.

Similarly, let f be a morphism of Q-modules for some quantale Q. Let $a \in Q, s, t \in S$, $(s, t) \in \ker(f)$. Then $f(as) = af(s) = af(t) = f(at)$. Therefore $(as, at) \in \ker(f)$, i.e. $\ker(f)$ is a congruence of quantale modules. □

PROPOSITION 2.3.11. *If S is a sup-lattice and $\sim$ a sup-lattice congruence on S, the factor set $S/{\sim}$ is a sup-lattice again and the projection $\pi : Q \to Q/{\sim}$ is therefore a sup-lattice morphism. Moreover, we have*

1. *If S is a quantale and $\sim$ a quantale congruence on S, then $S/{\sim}$ is a quantale and the projection $\pi : S \to S/{\sim}$ is therefore a quantale morphism. The quantale $S/{\sim}$ is then called a* quotient quantale *of S by the quantale congruence $\sim$.*
2. *If S is a quantale module and $\sim$ a module congruence on S, then $S/{\sim}$ is a quantale module and the projection $\pi : S \to S/{\sim}$ is therefore a quantale module morphism. The quantale module $S/{\sim}$ is then called a* quotient module *of S by the module congruence $\sim$.*

PROOF. We have to verify that joins defined on $S/{\sim}$ as $\bigvee_{i\in I}[s]_i = \left[\bigvee_{i\in I} s_i\right]$ do not depend on the choice of representatives.

Let $t_i \in S, t_i \sim s_i, i \in I$. Then $\bigvee_{i\in I} s_i \sim \bigvee_{i\in I} t_i$ and therefore $[\bigvee_{i\in I} s_i] = [\bigvee_{i\in I} t_i]$.

Similarly, for $\sim$ being a quantale congruence we have also to check that the multiplication defined by $[a] \cdot [b] = [ab]$ does not depend on the choice of representatives. Let $c, d \in S, a \sim c, b \sim d$. Then $ab \sim cd$. Hence $[a] \cdot [b] = [ab] = [cd] = [c] \cdot [d]$. In the same way we can prove the quantale module case. □

PROPOSITION 2.3.12. *Let S and T be sup-lattices, let $f : S \to T$ be a sup-lattice morphism, let $\sim$ be a sup-lattice congruence on S such that $\sim \subseteq \ker f$, and let π denote the projection $S \to S/{\sim}$. Then there exists a unique sup-lattice morphism $g : S/{\sim} \to T$ which satisfies $g \circ \pi = f$. Moreover, we have*

1. *If f is a quantale morphism and $\sim$ a quantale congruence on S, then g is a quantale morphism.*
2. *If f is a Q-module morphism, Q a quantale and $\sim$ a Q-module congruence on S, then g is a Q-module morphism.*

PROOF. First note that if $g \circ \pi = f$ then we must have $g([s]) = f(s)$. Since $\sim \subseteq \ker f$, g does not depend on the choice of representatives and thus it is a correctly defined map. Let $s_i \in S$, $i \in I$. Then $g(\bigvee_{i\in I}[s_i]) = g([\bigvee_{i\in I} s_i]) = f(\bigvee_{i\in I} s_i) = \bigvee_{i\in I} f(s_i) = \bigvee_{i\in I} g([s_i])$. Hence g is a sup-lattice morphism. If there is another morphism $h : Q/{\sim} \to$

R satisfying $h \circ \pi = f$, we easily see that $h([a]) = (h \circ \pi)(a) = f(a) = g([a])$. The remaining parts can be proved by the same procedure. □

PROPOSITION 2.3.13. *Let $f : Q \to R$ be a surjective quantale (Q-module) morphism. Then there exists a (Q-module) nucleus $j : Q \to Q$ such that $R \cong Q_j$.*

PROOF. Define $j = f^{\dashv} \circ f$. We know that j is an order nucleus and $f \circ j = f \circ f^{\dashv} \circ f = f$. Define a map $g : R \to Q_j$ by $g(r) = f^{\dashv}(r)$ for all $r \in R$. Evidently, $f \circ f^{\dashv} = \mathrm{id}_R$. We will show that g preserves joins: $g(\bigvee r_i) = g(\bigvee f(q_i)) = (g \circ f)(\bigvee q_i) = j(\bigvee q_i) = \bigvee j(q_i) = \bigvee (g \circ f)(q_i) = \bigvee g(r_i)$. Further, g is a surjective map because $g \circ f = f^{\dashv} \circ f = j : Q \to Q_j$ is a surjection that which implies g is a surjection. Let $g(r_1) = g(r_2)$. Then $r_1 = f(g(r_1)) = f(g(r_2)) = r_2$, i.e. g is injective and hence a sup-lattice isomorphism.

Let f be a quantale morphism. We have

$$\begin{aligned} f\big(j(a)j(b)\big) &= f\big(j(a)\big)f\big(j(b)\big) = \big(f \circ f^{\dashv} \circ f\big)(a)\big(f \circ f^{\dashv} \circ f\big)(b) \\ &= f(a)f(b) = f(ab). \end{aligned}$$

Then $f(j(a)j(b)) \leqslant f(ab)$ implies $j(a)j(b) \leqslant (f^{\dashv} \circ f)(ab) = j(ab)$.

Now, let f be a Q-module morphism. Then, for all $a \in Q$, $s \in S$, we have

$$f\big(aj(s)\big) = af\big(j(s)\big) = a\big(f \circ f^{\dashv} \circ f\big)(s) = af(s) = f(as).$$

Hence $f(aj(s)) \leqslant f(as)$ implies $aj(s) \leqslant (f^{\dashv} \circ f)(as) = j(as)$. □

THEOREM 2.3.14. (*Quantale representation theorem* [75].) *Let Q be a quantale. Then there is a semigroup S and a nucleus $j : \mathcal{P}(S) \to \mathcal{P}(S)$ such that $Q \cong \mathcal{P}(S)_j$.*

PROOF. We can take $S = Q$. Then the map $j : \mathcal{P}(Q) \to \mathcal{P}(Q)$ where $j(A) = {\downarrow}(\bigvee A)$ is certainly an order nucleus on Q and we will show that it is a nucleus too. For any $A, B \subseteq Q$, $c \leqslant \bigvee A$, $d \leqslant \bigvee B$ it follows that $cd \leqslant (\bigvee A)(\bigvee B) = \bigvee(AB)$. Thus $j(A)j(B) \subseteq j(AB)$. Evidently, $A \in \mathcal{P}(Q)_j \iff A = {\downarrow}a$ for some $a \in Q$.

This allows us to define a map $f : Q \to \mathcal{P}(Q)_j$ as $f(a) = {\downarrow}a$. It is clear that f is both injective and surjective. For $a, b \in Q$, $\bigvee({\downarrow}a{\downarrow}b)$ equals ab. Then $f(a) \cdot_j f(b) = {\downarrow}a \cdot_j {\downarrow}b = j({\downarrow}a{\downarrow}b) = {\downarrow}(ab) = f(ab)$. Clearly, f preserves arbitrary joins because $\bigvee f(a_i)$ is the smallest principal ideal containing all ${\downarrow}a_i$ which is, in fact, the principal ideal generated by their supremum: $f(\bigvee a_i)$. We have proved that f is a bijective quantale morphism and thus an isomorphism. □

Note that the sum $\bigsqcup_{i\in I} Q_i$ of quantales $Q_i, i \in I$ can be constructed as the free quantale over the semigroup sum $\overset{S}{\bigsqcup}_{i\in I} Q_i$ of Q_i with semigroup injections $\varepsilon_i : Q_i \to \overset{S}{\bigsqcup}_{j\in I} Q_j$ factored via the quantale congruence generated by the relation $\{(\varepsilon_i(\bigvee_\alpha x_\alpha), \{\varepsilon_i(x_\alpha)\colon x_\alpha, \alpha \in \Lambda\})\colon i \in I,\ \{x_\alpha\}_{\alpha\in\Lambda} \subseteq Q_i\}$. Similarly, a coequalizer k of quantale morphisms

$g, h : A \to B$ can be constructed as the quotient map $k : B \to B_{R_{g,h}}$, where $R_{g,h} = \{(g(a), h(a)) \colon a \in A\}$.

PROPOSITION 2.3.15. *Let M_i, $i \in I$, be Q-modules. Then $\prod_{i \in I} M_i \cong \sum_{i \in I} M_i$.*

PROOF. We will show that the product of Q-modules has the universal property of their sum.

Let N be a Q-module and let $f_i : M_i \to N$ be Q-module morphisms. Define $f = \bigvee (f_i \circ \pi_i) : \prod_{i \in I} M_i \to N$ which is again a Q-module morphism. For every $x \in M_i$, $i \in I$, define $m_{i,x} : I \to \bigcup M_i$ as

$$m_{i,x}(j) := \begin{cases} x, & i = j, \\ 0_{M_j}, & \text{otherwise.} \end{cases}$$

Evidently, $m_{i,x} \in \prod M_i$. Let us further define $\varepsilon_i : M_i \to \prod M_i$ as $\varepsilon_i(x) = m_{i,x}$. For any $i \in I$, ε_i is a Q-module morphism because $\varepsilon_i(\bigvee_j x_j) = m_{i, \bigvee_j x_j} = \bigvee_j m_{i,x_j} = \bigvee_j \varepsilon_i(x_j)$ and $\varepsilon_i(a \cdot x) = m_{i, a \cdot x} = a \cdot m_{i,x} = a \cdot \varepsilon_i(x)$.

Clearly, $f \circ \varepsilon_i = f_i$ and $\bigvee_j (\varepsilon_j \circ \pi_j) = \mathrm{id}_{\prod M_i}$.

Let us check that $\prod M_i$ together with the set of ε_i have the universal property of the sum. Suppose there exists $g : \prod M_i \to N$ such that $g \circ \varepsilon_i = f_i$ for any $i \in I$. We have $f = \bigvee (f_i \circ \pi_i) = \bigvee (g \circ \varepsilon_i \circ \pi_i) = g \circ \bigvee (\varepsilon_i \circ \pi_i) = g \circ \mathrm{id}_{\prod M_i} = g$. □

Note that the coequalizer k of Q-module morphisms $g, h : M \to N$ can be constructed as the quotient map $k : N \to N_{R_{g,h}}$, where $R_{g,h} = \{(g(m), h(m)) \colon m \in M\}$.

2.4. *Special morphisms of quantales and modules*

In this section we will consider special morphisms in the categories *Quant*, and *Q-$\mathcal{M}$od* and their subcategories.

PROPOSITION 2.4.1. *There are epimorphisms in the category Quant ($\mathcal{U}$n Quant, $\mathcal{I}$n Quant and $\mathcal{I}$n$\mathcal{U}$n Quant) that are not surjective.*

PROOF. Let A be a frame that is not a Boolean algebra. From Proposition 2.7 in [22] we know that there is an epimorphism of frames $c : A \to N(A)$ that is not surjective. Note that any frame morphism is an involutive unital quantale morphism. It is enough to show that c is also an epimorphism of (involutive, unital) quantales. Recall that $N(A)$ is generated by the elements of the form $c(a) \wedge u(b)$ where $u(b)$ and $c(b)$ are complementary in $N(A)$.

Now, let $f_1, f_2 : N(A) \to B$ are (involutive, unital) quantale morphisms such that $f_1 \circ c = f_2 \circ c$. Since both $f_1(N(A))$ and $f_2(N(A))$ are frames it is enough to show that they coincide. To prove this we have to show that $f_1(u(a)) = f_2(u(a))$ for all $a \in A$. We have that $f_1(u(a) \vee c(a)) = f_1(1_{N(A)}) = f_1(c(1)) = f_2(c(1)) = f_2(1_{N(A)}) = f_2(u(a) \vee c(a))$. Hence

$$\begin{aligned} f_1\big(u(a)\big) &= f_1\big(u(a)\big)\big(f_2\big(u(a) \vee c(a)\big)\big) \\ &= f_1\big(u(a)\big)f_2\big(u(a)\big) \vee f_1\big(u(a)\big)f_2\big(c(a)\big) \\ &= f_1\big(u(a)\big)f_2\big(u(a)\big) \vee f_1\big(u(a)\big)f_1\big(c(a)\big) \\ &= f_1\big(u(a)\big)f_2\big(u(a)\big) \vee f_1\big(u(a)c(a)\big) \\ &= f_1\big(u(a)\big)f_2\big(u(a)\big) \vee f_1(0_{N(A)}) = f_1\big(u(a)\big)f_2\big(u(a)\big) \\ &\leqslant f_1(1_{N(A)})f_2\big(u(a)\big) = f_2(1_{N(A)})f_2\big(u(a)\big) = f_2\big(u(a)\big). \end{aligned}$$

Similarly, $f_2(u(a)) \leqslant f_1(u(a))$. Altogether, $f_1 = f_2$. □

The case of epimorphisms in quantale modules is surprisingly nicer. Let us start with the following definition.

DEFINITION 2.4.2. A category $\mathcal{K}$ is said to satisfy the *amalgamation property* provided that for every monomorphism $m : A \to B$, whenever the square

$$\begin{tikzcd} A \arrow[r, "m"] \arrow[d, "m"'] & B \arrow[d, "p"] \\ B \arrow[r, "q"'] & C \end{tikzcd}$$

is a pushout, then it is also a pullback and p, q are monomorphisms.

LEMMA 2.4.3. *Let $\mathcal{K}$ be an algebraic category such that the regular monomorphisms in $\mathcal{K}$ are precisely the monomorphisms. Then the regular epimorphisms in $\mathcal{K}$ are precisely the epimorphisms.*

PROOF. Let $f : A \to B$ be a $\mathcal{K}$-epimorphism. Let us consider the regular epi-mono factorization $f = m \circ p$; $m : C \to B$ is a monomorphism and $p : A \to C$ is a regular epimorphism. Since f is an epimorphism so is m. Moreover, m is a regular monomorphism, i.e. it is an equalizer of morphisms $u, v : B \to D$. Since m is an epimorphism we have that $u = v$. But then the equalizer of u and v is the identity map id_B. Consequently, m is isomorphic to the identity on B and therefore it is an isomorphism. Finally, f is isomorphic to the regular epimorphism p and so is an regular epimorphism. □

PROPOSITION 2.4.4. *Let $\mathcal{K}$ be an algebraic category satisfying the amalgamation property. Then*

1. *The regular monomorphisms in the category $\mathcal{K}$ are precisely the monomorphisms.*
2. *The surjective morphisms in the category $\mathcal{K}$ are precisely the epimorphisms.*

PROOF. 1. Let $m : A \to B$ be a $\mathcal{K}$-monomorphism. Since $\mathcal{K}$ is algebraic it is cocomplete and therefore there exists a pushout

$$\begin{tikzcd} A \arrow[r, "m"] \arrow[d, "m"'] & B \arrow[d, "p"] \\ B \arrow[r, "q"'] & C \end{tikzcd}$$

which is also a pullback by the amalgamation property. Hence $m : A \to B$ is an equalizer of (p, q), i.e. m is a regular monomorphism. The other implication is evident.

2. From 2.2.1 we know that regular epis in $\mathcal{K}$ are exactly surjective morphisms and by 2.4.3 we know that the regular epimorphisms in $\mathcal{K}$ are precisely the epimorphisms. □

The next theorem was first proved in [46] and later on, by more simple arguments in [81]. We shall here give a shorter proof based on module prenuclei to demonstrate their strength.

THEOREM 2.4.5. *The category Q-$\mathcal{M}od$ satisfies the amalgamation property.*

PROOF. Let $m : A \to B$ be a Q-$\mathcal{M}od$-monomorphism. By the Theorems 2.2.2 and 2.2.1 one can assume that A is a Q-submodule of B and m is the inclusion map. Assume that the square

$$\begin{tikzcd}
A \arrow[r, "m"] \arrow[d, "m"'] & B \arrow[d, "p"] \\
B \arrow[r, "q"'] & C
\end{tikzcd}$$

is a pushout. Let $B \times B$ be the coproduct in Q-$\mathcal{M}od$ with the injections $i_1, i_2 : B \to B \times B$ defined by $i_1(b) = (b, 0)$ and $i_2(b) = (0, b)$. Let $d : B \times B \to D$ be a coequalizer of $i_1 \circ m$ and $i_2 \circ m$. By the construction of coequalizers d is a module nucleus corresponding to the module congruence R generated by the relation $S = \{((m(a), 0), (0, m(a))) \colon a \in A\}$. Evidently, d is induced by the module prenucleus $j : B \times B \to B \times B$ defined by $j(x, y) = (x, y) \vee (m^{\dashv}(y), m^{\dashv}(x))$ (indeed, $((x, y), (x, y) \vee (m^{\dashv}(y), m^{\dashv}(x))) \in R$ because $(m^{\dashv}(x), 0) \vee (0, m^{\dashv}(y)) \leqslant (x, y)$ implies $((x, y) \vee (m^{\dashv}(x), m^{\dashv}(y)), (x, y) \vee (m^{\dashv}(y), m^{\dashv}(x)) \in R$ and $j(m(a), 0) = j(m(a), m(a)) = j((0, m(a))$ for all $a \in A$). Evidently, $(x, y) \leqslant j(x, y)$, $(x, y) \leqslant (u, v)$ implies $j(x, y) \leqslant j(u, v)$ and $qj(x, y) = q((x, y) \vee (m^{\dashv}(y), m^{\dashv}(x))) = (qx, qy) \vee (qm^{\dashv}(y), qm^{\dashv}(x))$. Since $m(qm^{\dashv}(x)) = qm(m^{\dashv}(x)) \leqslant qx$ we have $qm^{\dashv}(x) \leqslant m^{\dashv}(qx)$. Hence $qj(x, y) \leqslant j(q(x, y)) = (qx, qy) \vee (m^{\dashv}(qy), qm^{\dashv}(qx))$. Moreover, $j \circ i_1(b), j \circ i_2(b) \in (B \times B)_j$ for all $b \in B$. This follows from the following computation: $j(j(i_1(b)) = j(j(b, 0)) = j((b, m^{\dashv}(b)) = (b \vee m^{\dashv}(m^{\dashv}(b)), m^{\dashv}(b) \vee m^{\dashv}(b)) = (b, m^{\dashv}(b)) = j(i_1(b))$ and similarly $j(j(i_2(b)) = (m^{\dashv}(b) \vee m^{\dashv}(b), b \vee m^{\dashv}(m^{\dashv}(b))) = (m^{\dashv}(b), b) = j(i_2(b))$. Then the diagram

$$\begin{tikzcd}
A \arrow[r, "m"] \arrow[d, "m"'] & B \arrow[d, "d \circ i_1"] \\
B \arrow[r, "d \circ i_2"'] & D
\end{tikzcd}$$

is a pushout and hence $C \simeq D$. Consequently, it is enough to check that the last diagram is a pullback. Recall that a pullback of the diagram

$$\begin{tikzcd}
 & B \arrow[d, "d \circ i_1"] \\
B \arrow[r, "d \circ i_2"'] & D
\end{tikzcd}$$

can be constructed as a submodule E of $B \times B$;

$$\begin{aligned} E &= \bigl\{(b_1, b_2) \in B \times B\colon\ d \circ i_1 \circ \pi_1(b_1, b_2) = d \circ i_2 \circ \pi_2(b_1, b_2)\bigr\} \\ &= \bigl\{(b_1, b_2) \in B \times B\colon\ d \circ i_1(b_1) = d \circ i_2(b_2)\bigr\} \\ &= \bigl\{(b_1, b_2) \in B \times B\colon\ j \circ i_1(b_1) = j \circ i_2(b_2)\bigr\} \\ &= \bigl\{(b_1, b_2) \in B \times B\colon\ \bigl(b_1, m^{\dashv}(b_1)\bigr) = \bigl(m^{\dashv}(b_2), b_2\bigr)\bigr\}. \end{aligned}$$

Since $b_1 = m^{\dashv}(b_2) \in A$ we have that $b_1 \in A$, therefore $b_1 = m^{\dashv}(b_1) = b_2 \in A$. Then $E = \{(b, b) \in B \times B\colon\ b \in A\} \simeq A$.

Let us show that $d \circ i_1$ and $d \circ i_2$ are monomorphisms. Let $d \circ i_1(b_1) = d \circ i_1(b_2)$. Then $j \circ i_1(b_1) = j \circ i_1(b_2)$ and hence $(b_1, m^{\dashv}(b_1)) = (b_2, m^{\dashv}(b_2))$, i.e. $b_1 = b_2$. The proof for $d \circ i_2$ is similar. □

THEOREM 2.4.6.
1. *The regular monomorphisms* (*regular epimorphisms*) *in the category Q-$\mathcal{M}od$* (*Q-$\mathcal{U}n\mathcal{M}od$*) *are precisely the monomorphisms* (*epimorphisms*).
2. *The categories $\mathcal{Q}uant$, $\mathcal{U}n\mathcal{Q}uant$, $\mathcal{I}n\mathcal{Q}uant$ and $\mathcal{I}n\mathcal{U}n\mathcal{Q}uant$ do not have the amalgamation property.*
3. *There are monomorphisms in the category*

   $\mathcal{Q}uant$ ($\mathcal{U}n\mathcal{Q}uant$, $\mathcal{I}n\mathcal{Q}uant$ and $\mathcal{I}n\mathcal{U}n\mathcal{Q}uant$)

   that are not regular monomorphisms.

PROOF.
1. This follows from Theorems 2.4.4 and 2.4.5.
2. This follows from Theorem 2.4.4 and Proposition 2.4.1.
3. This follows from Lemma 2.4.3 and Proposition 2.4.1. □

2.5. *Notes on Section 2*

The construction of both free (unital, involutive) quantales and free unital modules is an immediate translation of the analogous constructions in universal algebra. The only non-standard case is the construction of free quantale modules due to M. Ordelt [48].

The algebraicity of the above categories of quantales and quantale modules is an useful property that enables us immediately to construct limits, colimits and regular epi-mono factorizations.

The general theory of nuclei is contained in [75]. We added the part on prenuclei as we think they are a suitable technical tool for studying quotients of quantales and quantale modules. The standard proof of the Proposition 2.3.15 is based on the idea for modules over rings.

The study of epimorphisms in quantales and quantale modules is quite recent. In the category of (unital, involutive) quantales there are non-surjective epimorphism. The more pleasant case for quantale modules was established by C. Nkuimi-Jugnia [46] and later on by S.A. Solovyov [81]. This was done by proving the so-called amalgamation property.

3. Simple quantales and representations

3.1. *Simple quantales*

DEFINITION 3.1.1. A non-trivial (involutive) quantale Q is said to be (**-simple*) *simple* if any surjective (*-)morphism of (involutive) quantales is either an (*-)isomorphism or a constant morphism. Evidently, any simple involutive quantale is *-simple.

A quantale Q will be called *faithful* (*strongly faithful*) if, whenever

$$xr = yr \quad \text{and} \quad lx = ly \quad (lxr = lyr)$$

for all $r \in \mathcal{R}(Q)$ and $l \in \mathcal{L}(Q)$, then

$$x = y$$

for all $x, y \in Q$. Note that any strongly faithful quantale is faithful and a Cartesian product of faithful (strongly faithful) quantales is again faithful (strongly faithful).

We say that the multiplication of a quantale Q is *trivial* if $x \cdot y = 0$ for each $x, y \in Q$; it is equivalent to $1 \cdot 1 = 0$. It is evident that the only simple quantale with a trivial multiplication is the quantale $0_2 = \{0, 1\}$ (any surjective morphism of sup-lattices is in this case a surjective morphism of quantales). Analogously, the only *-simple quantale with a trivial multiplication is the quantale 0_2 with the identity involution.

THEOREM 3.1.2. (*See* [62,51,30].) *Let Q be an* (*involutive*) *quantale with a non-trivial multiplication. Then the following conditions are equivalent*:

1. *Q is simple* (**-simple*).
2. *Q is a faithful von Neumann* (**-*)*factor.*
3. *Q is a strongly faithful* (**-*)*factor.*
4. *Q has a separating* (**-separating*) *cyclic set.*

PROOF. (1) $\Rightarrow$ (2). Let $t \in \mathcal{T}(Q)$ ($t \in \mathcal{T}(Q) \cap \mathcal{H}(Q)$). The assignment $a \mapsto a \vee t$ defines a surjective (involutive) morphism $Q \to \uparrow t$. Since Q is simple, we have either $t = 0$ or $t = 1$, i.e. Q is a (*-)factor. From the additional condition $1 \cdot 1 \neq 0$ we conclude $1 \cdot 1 = 1$, because $1 \cdot 1$ is always two-sided (and Hermitean). Let us now define

$$a \sim b \quad \Leftrightarrow \quad \big(\forall l \in \mathcal{L}(Q), \forall r \in \mathcal{R}(Q)\big)(lar = lbr).$$

The relation $\sim$ is evidently an (involutive) quantale congruence. If $0 \sim 1$, then $0 = 1{\cdot}0{\cdot}1 = 1 \cdot 1 \cdot 1 = 1$. Thus $\sim$ is necessarily the diagonal $a \sim b \Leftrightarrow a = b$, and so Q is strongly faithful.

Now we shall prove that Q is von Neumann, i.e. for all $r \in \mathcal{R}(Q)$ and $l \in \mathcal{L}(Q)$, $0 \leftarrow (r \to 0) = r$ and $(0 \leftarrow l) \to 0 = l$.

We shall define, for $x, y \in Q$, $x\varrho y$ if $lx = ly$ for all $l \in \mathcal{L}(Q)$. Then ϱ is a non-trivial quantale congruence on Q. Hence $\varrho = \mathrm{id}_Q$.

It is enough to show that $0 \leftarrow (r \to 0)\varrho r$. Let $l \in \mathcal{L}(Q)$. Then $lr = 0$ or $lr = 1 \leqslant l(0 \leftarrow (r \to 0))$. If $lr = 0$ then $l \leqslant (r \to 0)$, $(r \to 0)(0 \leftarrow (r \to 0)) = 0$, i.e. $l(0 \leftarrow (r \to 0)) = 0$. Similarly for left-sided elements.

The involutive case proceeds analogously.

(2) ⇒ (3). Let $lar = lbr$ for all $l \in \mathcal{L}(Q)$ and $r \in \mathcal{R}(Q)$. Recall that $lar = 0$ iff $r \leqslant 0 \leftarrow (la)$. Hence $0 \leftarrow (la) = 0 \leftarrow (lb)$. Then $lb = (0 \leftarrow (lb)) \rightarrow 0 = (0 \leftarrow (la)) \rightarrow 0 = la$. Similarly, $ar = br$. Consequently, $a = b$, i.e. Q is strongly faithful.

(3) ⇒ (4). Put $P = \{r \rightarrow 0 \leftarrow l : r \in \mathcal{R}(Q), l \in \mathcal{L}(Q)\} - \{1\}$. $P \cup \{1\}$ is residually closed, because $b \rightarrow (r \rightarrow 0 \leftarrow l) \leftarrow a = (br) \rightarrow 0 \leftarrow (la)$. From the faithfulness we have that $1 \cdot 1 \cdot 1 \neq 1 \cdot 0 \cdot 1$. For $p \in P$ we have $1 \rightarrow p \leftarrow 1 \in \mathcal{T}(Q) = \{0, 1\}$. If $1 \rightarrow p \leftarrow 1 = 1$, then $1 = 1 \cdot 1 \cdot 1 \leqslant p$, thus $1 \rightarrow p \leftarrow 1 = 0$ and hence P is cyclic. Let $a \not\geqslant b$, then there exist $r \in \mathcal{R}(Q), l \in \mathcal{L}(Q)$ such that $lar \not\geqslant lbr$. Since Q is a factor and lar, lbr are two-sided, $lar = 0$ and $lbr = 1 \not\leqslant 0$. But it means that $a \leqslant r \rightarrow 0 \leftarrow l$ and $b \not\leqslant r \rightarrow 0 \leftarrow l$, i.e. P is separating.

In the involutive case we put either $P = \{r \rightarrow 0 \leftarrow l\colon r \in \mathcal{R}(Q), l \in \mathcal{L}(Q)\} - \{1\}$ if $\mathcal{T}(Q) = \{0, 1\}$ or $P = \{r \rightarrow t \leftarrow l\colon r \in \mathcal{R}(Q), l \in \mathcal{L}(Q)\} - \{1\}$ for some $t \in \mathcal{T}(Q) - \mathcal{H}(Q)$ otherwise. The remaining part of the proof can be done in a similar way as above.

(4) ⇒ (1). Let P be separating and cyclic. Then $0 = \bigwedge P = 1 \rightarrow p \leftarrow 1$ for every $p \in P$ and there are $a, b \in Q$ such that $p = b \rightarrow 0 \leftarrow a$. Let $a \in Q$ and suppose $1a1 \leqslant p$ for some $p \in P$. Then $a \leqslant 1 \rightarrow p \leftarrow 1 = 0$, that is $a \neq 0 \Rightarrow 1a1 = 1$. Further $0 \neq 1 = 1 \cdot 1 \cdot 1 \leqslant 1 \cdot 1$. Consider $a \not\geqslant b$ and the congruence $\sim$ generated by $a \sim b$. There exists a $p \in P$, $p = d \rightarrow 0 \leftarrow c$ such that $a \leqslant p, b \not\leqslant p$. Then $cad = 0$, $cbd \neq 0$, $cad \sim cbd$ and thus $0 = 1cad1 \sim 1cbd1 = 1$, i.e. there are only two congruences on Q, so that Q is simple.

In the involutive case we put $t = \bigwedge P = 1 \rightarrow p \leftarrow 1$ for every $p \in P$ and as above we show that any non-trivial principal involutive quantale congruence $\sim$ is $Q \times Q$ ($t \sim 1$ and $t^* \sim 1$ would imply that $0 = t \wedge t^* \geqslant tt^* \sim 11 = 1$). □

COROLLARY 3.1.3. *Let S be a sup-lattice. Then*

1. $\mathcal{Q}(S)$ *is simple.*
2. $\mathcal{Q}(S) \times \mathcal{Q}(S^{\mathrm{op}})$ *is *-simple.*

PROOF. If S is a trivial quantale, i.e. $S = \{0\}$ then $\mathcal{Q}(S) \simeq \{0\} \simeq \mathcal{Q}(S) \times \mathcal{Q}(S^{\mathrm{op}})$ is evidently simple and therefore *-simple. Now, let S be non-trivial. Then:

1. Evidently, $\rho_0 = \lambda_1 = 0_{\mathcal{Q}(S)}$ and $\rho_1 = \lambda_0 = 1_{\mathcal{Q}(S)}$ are the only two-sided elements.

Let us check that $\mathcal{Q}(S)$ is strongly faithful. Recall that, for any $f \in \mathcal{Q}(S)$,

$$(\lambda_u \circ f \circ \rho_v)(t) = \begin{cases} 0, & t = 0 \text{ or } f(v) \leqslant u, \\ 1, & \text{otherwise.} \end{cases}$$

Assume that $\alpha \not\geqslant \beta$ in $\mathcal{Q}(S)$, i.e. there is $x \in S$ for which $\alpha(x) \not\geqslant \beta(x)$. For $y = \alpha(x)$ we get $\alpha(x) \leqslant y$ and $\beta(x) \not\leqslant y$ and hence $0 = \lambda_y \circ \alpha \circ \rho_x \neq \lambda_y \circ \beta \circ \rho_x = 1$.

2. Note first, that the involution on $\mathcal{Q}(S) \times \mathcal{Q}(S^{\mathrm{op}})$ is defined as $(f, g)^* = (g^{\dashv}, f^{\dashv})$. From properties of adjoints it follows that $Q = \mathcal{Q}(S) \times \mathcal{Q}(S^{\mathrm{op}})$ is a *-quantale. We will show that it is *-simple. Indeed, $\mathcal{T}(Q) = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ and $(1, 0)^* = (0, 1)$, hence Q is a *-factor. Since $\mathcal{Q}(S)$ and $\mathcal{Q}(S^{\mathrm{op}})$ are strongly faithful we have that Q is strongly faithful. □

3.2. *Representations of quantales*

As in ring theory, quantale modules are related to *representations of quantales*, i.e. morphisms to a $\mathcal{Q}(S)$. Concretely, a representation $\mu : Q \to \mathcal{Q}(M)$ is determined by the action of elements of Q on a left Q-module M: $\mu(a)(m) = am$, and vice versa.

An important example arise from theory of C*-algebras. Given a C*-algebra A and a representation $\pi : A \to \mathcal{B}(H)$ to the algebra $\mathcal{B}(H)$ of the bounded operators on a Hilbert space H, there is a corresponding *-quantale morphism $\mu : \operatorname{Max} A \to \mathcal{Q}(L(H))$ where $L(H)$ is the a sup-lattice of closed subspaces of H. The morphism μ is strong, or unital, if and only if π is irreducible, or unital, respectively. Hence the strong and unital representations are of a special interest.

THEOREM 3.2.1. (*See* [54].) *Every quantale has a faithful representation, i.e. an embedding* $\mu : Q \to \mathcal{Q}(S)$.

PROOF. By 1.1.11 every quantale Q can be embedded into a unital quantale, thus we can assume that Q is unital. Then $\mu(a)(b) = ab$ defines a representation $\mu : Q \to \mathcal{Q}(Q)$ which is faithful because $\mu(a) = \mu(b)$ implies $a = \mu(a)(e) = \mu(b)(e) = b$. □

Let Q be a quantale, I, J sets and M a matrix over Q of type $I \times J$. Let as write M^i for the i-th row, M_j for the j-th column, and M^i_j for the (i, j)-th entry of M. For a matrix N of type $J \times K$ we define a product $P = MN$ of type $I \times K$ by

$$P^i_k = \bigvee_{j \in J} M^i_j N^j_k.$$

Matrices of the same type form a sup-lattice with pointwise joins. Thus matrices over Q form a quantaloid (a sup-enriched category) with the index sets as objects. The residuations can be computed pointwise as well, for example $M = N \to P$ (with types as before) is given by

$$M^i_j = \bigwedge_{k \in K} N^j_k \to P^i_k.$$

DEFINITION 3.2.2. Let P be a matrix of type $I \times J$ over a quantale Q. We define $Q_J \to P$ to be the sup-lattice of all column vectors u of type I obtained as $u = v \to P$ for some row vector v of type J. From $a \to u = a \to (v \to P) = (av) \to P$ and $(\bigvee v_{(k)}) \to P = \bigwedge v_{(k)} \to P$ it follows that the dual sup-lattice $(Q_J \to P)^{\mathrm{op}}$ is a left Q-module. In a similar way we obtain a right Q-module $(P \leftarrow Q^I)^{\mathrm{op}}$. Modules obtained in this way (or isomorphic to those) are called *matrical modules*.

THEOREM 3.2.3. (*See* [29].) *Every unital module is matrical.*

PROOF. Let M be a left Q-module. We set P to be the $M \times M$ matrix over Q with entries

$$P^m_n = n \to m = \bigvee \{a \in Q \mid an \leqslant m\}.$$

From the definition of P it follows that $a \to P_n = P_{an}$ and $\bigwedge P_{n_i} = P_{\bigvee n_i}$. Let E^n be the n-th row of the unit matrix of type $M \times M$. Then $E^n \to P = P_n$, hence columns of P form a submodule of $(Q_M \to P)^{\mathrm{op}}$. On the other hand, for every row vector u of type M we have

$$u \to P = \bigwedge_{m \in M} \{u_m \to P_m\} = P_{\bigvee_{m \in M} um},$$

hence every element of the matrix module is a column vector of P. Finally, for $m \nleqslant n$ we have $e \leqslant n \to n$, $e \nleqslant m \to n$, thus $P_m \neq P_n$, hence $m \mapsto P_m$ is an isomorphism. □

THEOREM 3.2.4. (*See* [29].) *Let P be a matrix of type $I \times J$. Then the left Q-module $(Q_J \to P)^{\mathrm{op}}$ is dual to the right Q-module $(P \leftarrow Q^I)^{\mathrm{op}}$.*

PROOF. The assignments $v \mapsto v \to P, u \mapsto P \leftarrow u$ define a Galois connection between row vectors of type J and column vectors of type I. The fix-points are precisely the elements of modules. For $a \in Q, u, u' \in Q_J \to P, v, v' \in P \leftarrow Q^I$ such that $u = v \to P, v = P \leftarrow u, u' = v' \to P, v' = P \leftarrow u'$ we have $u' \leqslant a \to u \Leftrightarrow u'a \leqslant u = v \to P \Leftrightarrow u'av \leqslant P \Leftrightarrow av \leqslant P \leftarrow u' = v' \Leftrightarrow v \leqslant v' \leftarrow a$ which establishes the module duality. □

DEFINITION 3.2.5. A quantale Q is called *spatial* if it has enough strong representations to separate elements.

An element $p \in Q$, $p \neq 1$ is called *prime* if

$$rl \leqslant p \quad \Rightarrow \quad r \leqslant p \text{ or } l \leqslant p$$

for all $r \in \mathcal{R}(Q), l \in \mathcal{L}(Q)$.

THEOREM 3.2.6. (*See* [29].) *A quantale Q is spatial if, and only if, everyone of its element is a meet of primes.*

PROOF. 1. Assume that Q is a spatial quantale. Then there is a strong embedding $\mu : Q \to \prod \mathcal{Q}(S_i)$. It is enough to show that each $\mathcal{Q}(S)$ has enough primes, products of quantales with enough primes have enough primes as well, and the adjoint of a strong morphism preserves primes.

1i. The right- and left-sided elements of $\mathcal{Q}(S)$ are of the form

$$\rho_x(z) = \begin{cases} x, & z \neq 0, \\ 0, & z = 0, \end{cases} \qquad \lambda_y(z) = \begin{cases} 1, & z \nleqslant y, \\ 0, & z \leqslant y, \end{cases}$$

respectively. Let $x \neq 1$, $y \neq 0$. We have $\rho_u \circ \lambda_v \leqslant \rho_x \vee \lambda_y$ if, and only if, $u \leqslant x$ or $y \leqslant v$, i.e. $\rho_u \leqslant \rho_u$ or $\lambda_v \leqslant \lambda_y$. This means that all elements of the form $\rho_x \vee \lambda_y$ are primes. From $\alpha \leqslant \rho_x \vee \lambda_y \Leftrightarrow \alpha(y) \leqslant x$ it follows that the primes $\rho_x \vee \lambda_y$ separate the elements of $\mathcal{Q}(S)$.

1ii. Let Q_i be a family of quantales, $Q = \prod Q_i$. Since multiplication in Q is given pointwise, the right- or left-sided elements have all components right- or left-sided, respectively.

It follows that when p is a prime in Q_i then $p' \in Q$ given by

$$p'_j = \begin{cases} p, & i = j, \\ 1, & i \neq j \end{cases}$$

is also a prime. These primes of Q clearly separate elements whenever the primes of each Q_i do so.

1iii. Finally, let $f : Q \to K$ be a strong morphism and p a prime of K. Then $f^{\dashv}(p) \neq 1$ and since f preserves right- and left-sided elements, we obtain $rl \leqslant f^{\dashv}(p) \Rightarrow f(r)f(l) = f(rl) \leqslant p \Rightarrow f(r) \leqslant p$ or $f(l) \leqslant p \Rightarrow r \leqslant f^{\dashv}(p)$ or $l \leqslant f^{\dashv}(p)$ for every $r \in \mathcal{R}(Q), l \in \mathcal{L}(Q)$. Hence $f^{\dashv}(p)$ is a prime.

2. Assume that every element of Q is a meet of primes. We will strongly embed Q into a unital quantale which still has enough primes. Then we will show that a module $(Q \to p)^{\mathrm{op}}$ with p prime induces a strong representation and that the representations separate elements.

2i. Let Q be a quantale with enough primes and $Q[e]$ its unitalization from 1.1.11. Consider the quotient of $Q[e]$ obtained by identifying 1 with $1 \vee e$. We will show that no other elements needs to be identified and hence $Q \to Q[e]/\{1 \sim 1 \vee e\}$ is a strong embedding. Let $a \in Q$. Then $(a1 \vee a)$ is clearly right-sided in Q and for every prime p of Q we have $a1 = (a1 \vee a)1 \leqslant p \Rightarrow a1 \vee a \leqslant p$, hence $a1 = a1 \vee a$ because primes separate elements. It means that the induced relation $a1 \sim a(1 \vee e) = a1 \vee e$ is already satisfied in Q. Similarly we prove $1a = 1a \vee a$. Further, 1 and $1 \vee e$ live on the top of $Q[e]$, thus we do not obtain anything new by joins, and multiplying the new elements of $Q[e]$ is also trivial: $(a \vee e)1 = 1, (a \vee e)(1 \vee e) = 1 \vee e$. This also explains that the right- and left-sided elements of $Q \to Q[e]/\{1 \sim 1 \vee e\}$ are only the old ones from Q and from $rl \leqslant p \Leftrightarrow rl \leqslant p \vee e$ for every $r, l, p \in Q$ we obtain that $p \vee e$ is a prime in $Q[e]/\{1 \sim 1 \vee e\}$ whenever p is a prime in Q. It follows that $Q[e]/\{1 \sim 1 \vee e\}$ has enough primes.

2ii. Let Q be a unital quantale with enough primes, p a prime, and a, b arbitrary elements. If $a1b \leqslant p$ then $a = ae \leqslant a1 \leqslant p$ or $b = eb \leqslant 1b \leqslant p$. In the later case we have $b \to p = (1b) \to p = 1$. If $b \nleqslant p$ then $1b \to p \leqslant p$ because $(1b) \to p$ is right-sided and p is prime. For every right-sided $r \leqslant p$ we have $r \leqslant 1 \to p$, hence $(1b) \to p = 1 \to p$. Consequently,

$$1 \to (b \to p) = (1b) \to p = \begin{cases} 1, & b \to p = 1, \\ 1 \to p, & \text{otherwise.} \end{cases}$$

But $1 \to p$ or 1 is the top or the bottom element, respectively, of the module $(Q \to p)^{\mathrm{op}}$ and the top element of $\mathcal{Q}((Q \to p)^{\mathrm{op}})$ acts in the same way, hence the induced representation $\mu_p : Q \to \mathcal{Q}((Q \to p)^{\mathrm{op}})$ is strong. Finally, for every $a \ngeqslant b$ we have a prime p such that $a \leqslant p$, $b \nleqslant p$ and $\mu_p(a)(e \to p) = a \to p \geqslant e$ while $\mu_p(b)(e \to p) = b \to p \ngeqslant e$, i.e. μ_p separates a, b. □

Proposition 3.2.7. (*See* [62].) *Let Q be a simple quantale with a non-trivial multiplication. Then the action on right-sided elements provides a strong embedding $\mu : Q \to \mathcal{Q}\mathcal{R}(Q)$.*

Proof. Let $a \neq b$. From 3.1.2 it follows that $lar \neq lbr$ for a suitable $r \in \mathcal{R}(Q), l \in \mathcal{L}(Q)$ and hence $ar \neq br$. Thus μ is an embedding. Let $r \in \mathcal{R}(Q)$. If $1r = 0$ then $lrs = 0$ for every $l \in \mathcal{L}(Q)$, $s \in \mathcal{R}(Q)$ and thus $r = 0$. Otherwise from the two-sidedness of $1r$ it follows that $1r = 1$. Hence

$$\mu(1)(r) = 1r = \begin{cases} 0, & r = 0, \\ 1, & \text{otherwise,} \end{cases}$$

i.e. μ is strong. □

Corollary 3.2.8. *A quantale is spatial if, and only if, it has enough strong morphisms to simple quantales with a non-trivial multiplication to separate elements.*

Now we will study spatiality of involutive quantales. Since the proofs of the following statements would be similar, we will point out only the extra difficulties of the involutive case.

Definition 3.2.9. A *-morphism $Q \to \mathcal{Q}(S) \times \mathcal{Q}(S^{\mathrm{op}})$ is called a **-representation*, a *-morphism $Q \to \mathcal{Q}(S)$ (with S self-dual) a *D-representation*.

A *-quantale is said to be **-spatial* (*D-spatial*) if it has enough *-representations (D-representations) to separate elements.

A matrix P of type $I \times J$ is called *prime* if at least one entry is different from 1 and

$$rl \leqslant P \quad \Rightarrow \quad r \leqslant P_j \text{ or } l \leqslant P^i$$

for all $i \in I$, $j \in J$, column vectors r formed by right-sided elements, and row vectors l formed by left-sided elements.

A matrix P is called *Hermitean* if $I = J$ and $(P^i_j)^* = P^j_i$ for every i, j.

Theorem 3.2.10. (*See* [29].) *A *-quantale Q is *-spatial if, and only if, every its element is a meet of primes.*

Proof. Recall that for every left Q-module M there is a dual right Q-module M^{op}. If Q is involutive, M^{op} can be regarded as a left Q-module with action $a \bullet m = m \leftarrow a^*$. Thus for the representation $\mu : Q \to \mathcal{Q}(M)$ we obtain a representation $\mu' : Q \to \mathcal{Q}(M^{\mathrm{op}})$. Since $\mu'(a) = \mu(a^*)^{\dashv}$, (μ, μ') induces a *-representation $Q \to \mathcal{Q}(M) \times \mathcal{Q}(M^{\mathrm{op}})$. The rest follows from Theorem 3.2.6. □

Theorem 3.2.11. (*See* [29].) *A *-quantale Q is D-spatial if, and only if, every element of it is a meet of primes contained in some Hermitean prime matrix.*

Proof. 1. Let P be a prime matrix of type I, J. Replacing elements by vectors and matrices of a suitable type in the proof of 3.2.6 we obtain that $(Q_J \to P)^{\mathrm{op}}$ is strong as well.

Assume that P is Hermitean. From $uv \leqslant P \Leftrightarrow v^*u^* \leqslant P$ it follows that $u \mapsto u^*$ defines an isomorphism between the dual modules $(Q_I \to P)^{\mathrm{op}}$ and $(P \leftarrow Q^I)^{\mathrm{op}}$ and hence $(Q_I \to P)^{\mathrm{op}}$ is self-dual with duality $m \mapsto m^* \to P$. Moreover, any element a of

Q acts on $(Q_I \to P)^{\mathrm{op}}$ in the same way as a^* on the dual $(P \leftarrow Q^I)^{\mathrm{op}}$, hence the induced representation $Q \to \mathcal{Q}(Q_I \to P)^{\mathrm{op}}$ is a D-representation.

2. On the other hand, if S is self-dual with duality $^\perp$, then any element $\rho_x \vee \lambda_y \neq 1$ is contained in the Hermitean matrix

$$P = \begin{pmatrix} \rho_x \vee \lambda_{x^\perp} & \rho_x \vee \lambda_y \\ \rho_{y^\perp} \vee \lambda_{x^\perp} & \rho_{y^\perp} \vee \lambda_y \end{pmatrix}.$$

P is prime because

$$\begin{pmatrix} \rho_u \\ \rho_v \end{pmatrix} (\lambda_w \quad \lambda_z) \leqslant P$$

implies that at least three of the following inequalities are true: $u \leqslant x$, $v \leqslant y^\perp$, $w \geqslant x^\perp$, $z \geqslant y$.

By a straightforward calculation we observe that if $f : Q \to K$ is a *-morphism (or strong morphism) then $f^\dashv$ preserves Hermitean (or prime) matrices. □

Involutive modifications of 3.2.7 and 3.2.8 are also possible. When Q is *-simple recall that there are two possibilities: either $\mathcal{T}(Q) = \{0, 1\}$ and then $\mu : Q \to \mathcal{Q}\mathcal{R}(Q)$ is a strong D-representation, or $\mathcal{T}(Q) = \{0, t, t^*, 1\}$ and then $\mu : Q \to \mathcal{Q}(\downarrow t \cap \mathcal{R}(Q)) \times \mathcal{Q}(\downarrow t^* \cap \mathcal{R}(Q))$ is a *-representation.

3.3. *Girard quantales*

Girard quantales can be considered as generalized Boolean algebras or discrete spaces and thus it is reasonable to find such a structure on endomorphism quantales $\mathcal{Q}(S)$. K.I. Rosenthal [75] used the Chu construction [7] and embedded a general quantale Q into a Girard quantale $Q \times Q^{\mathrm{op}}$. However, the embedding is not strong and does not respect the existing duality between right- and left-sided elements. We present another construction which provides a Girard structure between $\mathcal{Q}(S)$ and its dual and preserves von Neumann duality.

DEFINITION 3.3.1. An element $d \in Q$ is called *cyclic* if $ab \leqslant d \Leftrightarrow ba \leqslant d$ for all $a, b \in Q$.

d is called *dualizing* if $d \leftarrow (a \to d) = (d \leftarrow a) \to d = d$ for every $a \in Q$.

Quantale Q is said to be *Girard* if it admits a cyclic dualizing element. We write $a^\perp = a \to d = d \leftarrow a$.

EXAMPLE 3.3.2.

1. (See [76].) Rel X is a Girard quantale with $d = X \times X - \Delta_X$.
2. Let $M_n(\mathbb{C})$ be the C*-algebra of $n \times n$ complex matrices (that is $\mathcal{B}(\mathbb{C}^n)$). Then Max $M_n(\mathbb{C})$ is a Girard quantale with $d = \{A \mid \mathrm{Tr}\, A = 0\}$.
3. The interval $[0, 1]$ of reals with the usual order and with multiplication $ab = \max\{0, a + b - 1\}$ form a so called *Łukasiewicz quantale*. It is two-sided, commutative and Girard with $d = 0$.

PROPOSITION 3.3.3. *Every Girard quantale Q is von Neumann, $r \to 0 = r^{\perp}$ for every right-sided r, and $0 \leftarrow l = l^{\perp}$ for every left-sided l. That is, $^{\perp}$ extends the duality between $\mathcal{R}(Q)$ and $\mathcal{L}(Q)$.*

PROOF. Let r be right-sided and $r \leqslant d$. Then $r1 \leqslant d \Rightarrow r^{\perp} = 1 \Rightarrow r = 0$. Since ar is right-sided for every $a \in Q$, we get $ar \leqslant d \Leftrightarrow ar = 0$. Hence $r^{\perp} = r \to d = r \to 0$. The second equality is analogous. □

DEFINITION 3.3.4. Let M be a left Q-module and a right K-module. Then M is called *(Q, K)-bimodule* if $(am)b = a(mb)$ for all $a \in Q, m \in M, b \in K$. When $Q = K$ we will simply call M a Q-bimodule.

THEOREM 3.3.5. *Let S be a sup-lattice. On $S \otimes S^{\mathrm{op}}$ put*

$$(x \otimes y)(u \otimes v) = \begin{cases} 0, & u \leqslant y, \\ x \otimes v, & \textit{otherwise}. \end{cases} \tag{*}$$

Then

- *$S \otimes S^{\mathrm{op}}$ is a $\mathcal{Q}(S)$-bimodule,*
- *(*) provides a quantale structure on $S \otimes S^{\mathrm{op}}$,*
- *the so called* mix map *$\phi(x \otimes y) = \rho_x \circ \lambda_y$ provides a strong quantale and $\mathcal{Q}(S)$-bimodule morphism $\phi : S \otimes S^{\mathrm{op}} \to \mathcal{Q}(S)$,*
- *ϕ restricted to right- or left-sided elements is an isomorphism,*
- *ϕ preserves the involution when S is self-dual,*
- *$d = \bigvee_x x \otimes x \in S \otimes S^{\mathrm{op}}$ is a cyclic dualizing element with respect to the bimodule actions, i.e. $\alpha c \leqslant d \Leftrightarrow c\alpha \leqslant d, \alpha = d \leftarrow (\alpha \to d) = (d \leftarrow \alpha) \to d$, and $c = d \leftarrow (c \to d) = (d \leftarrow c) \to d$ for every $\alpha \in \mathcal{Q}(S), c \in S \otimes S^{\mathrm{op}}$ where $\to, \leftarrow$ are computed respectively to the actions,*
- *ϕ is self-dual, i.e. $\phi(\alpha^{\perp})^{\perp} = \phi^{\dashv}(\alpha)$.*

PROOF. S is clearly a left $\mathcal{Q}(S)$-module and S^{op} is its dual right $\mathcal{Q}(S)$-module. Thus $S \otimes S^{\mathrm{op}}$ is a $\mathcal{Q}(S)$-bimodule. The actions are given by $\alpha(x \otimes y)\beta = \alpha(x) \otimes \beta^{\dashv}(y)$. Notice that the product (*) is “bilinear” and behaves in the same way like the product $(\rho_x \circ \lambda_y) \circ (\rho_u \circ \lambda_v)$ in $\mathcal{Q}(S)$. Moreover, the elements $\bigvee_i \rho_{x_i} \circ \lambda_{y_i}$ form a subbimodule of $\mathcal{Q}(S)$ and $\alpha \circ \rho_x = \rho_{\alpha(x)}, \lambda_y \circ \beta = \lambda_{\beta^{\dashv}(y)}$ which agrees with the actions on $S \otimes S^{\mathrm{op}}$. Further $\phi(1 \otimes 0) = \rho_1 \vee \lambda_0 = 1$, i.e. ϕ is strong. We have $(x \otimes y) \leqslant (x \otimes y)(1 \otimes 0)$ iff $x \otimes y = 0$ or $y = 0$, hence right-sided elements of $S \otimes S^{\mathrm{op}}$ are of the form $x \otimes 0$ and $x \otimes 0 \mapsto \rho_x$ is an isomorphism. Similarly, $\mathcal{L}(S \otimes S^{\mathrm{op}}) \cong \mathcal{L}\mathcal{Q}(S)$. Finally, $\phi((x \otimes y)^*) = \phi(y' \otimes x') = \rho_{y'} \circ \lambda_{x'} = (\rho_x \circ \lambda_y)^* = \phi(x \otimes y)^*$. This finishes the first five points of the statement.

From $\alpha(x \otimes y) \leqslant d \Leftrightarrow \alpha(x) \leqslant y \Leftrightarrow \alpha \leqslant \rho_y \vee \lambda_x \Leftrightarrow x \leqslant \alpha^{\dashv}(y) \Leftrightarrow (x \otimes y)\alpha \leqslant d$ we have that $(x \otimes y) \to d = \rho_y \vee \lambda_x = d \leftarrow (x \otimes y)$. Similarly, $(\rho_y \vee \lambda_x) \to d = (\rho_y \to d) \wedge (\lambda_x \to d) = (1 \otimes y) \wedge (x \otimes 0) = (1 \wedge x) \otimes (y \vee 0) = x \otimes y$, etc. Extending this to the joins of $x \otimes y$ and meets of $\rho_y \vee \lambda_x$ we obtain the assertion.

Finally, $\phi((\lambda_x \vee \rho_y)^{\perp})^{\perp} = \phi(x \otimes y)^{\perp} = (\rho_x \circ \lambda_y)^{\perp} = (\rho_x \wedge \lambda_y)^{\perp} = \rho_x^{\perp} \vee \lambda_y^{\perp} = (0 \otimes x) \vee (y \otimes 1) = \phi^{\dashv}(\lambda_x \vee \rho_y)$ and this extends to all elements of $\mathcal{Q}(S)$. □

Using a classical result of G.N. Raney [66] (see also [21]) we characterize sup-lattices for which $\mathcal{Q}(S)$ is a Girard quantale.

THEOREM 3.3.6. *$\mathcal{Q}(S)$ is Girard if, and only if, S is completely distributive.*

PROOF. In our notation, id_S being *tight* means that it can be expressed as $\bigvee_i \rho_{x_i} \circ \lambda_{y_i}$ and by [66] this holds whenever S is completely distributive. But then the right- and left-sided elements generate the whole quantale $\mathcal{Q}(S)$ and hence $\phi : S \otimes S^{\mathrm{op}} \to \mathcal{Q}(S)$ is surjective. From self-duality it follows that ϕ is bijective and thus a quantale isomorphism. Hence $\mathcal{Q}(S)$ is a Girard quantale.

On the other hand, assume that $\mathcal{Q}(S)$ is Girard. Since every element is a meet of some $\rho_x \vee \lambda_y$ and $(\rho_x \vee \lambda_y)^{\perp} = \rho_y \wedge \lambda_x = \rho_y \circ \lambda_x$ by 3.3.3, every element is also a join of $\rho_y \circ \lambda_x$, in particular id_S is. □

3.4. *Notes on Section 3*

In the beginning, quantale points were studied in context of right-sided elements and appeared in works of F. Borceux, G. van den Bossche, C.J. Mulvey, M. Nawaz, and J. Rosický [11,41,77].

The idea of the endomorphism quantale $\mathcal{Q}(S)$ was introduced by C.J. Mulvey and J.W. Pelletier [40]. J.W. Pelletier and J. Rosický [62] characterized *-simple quantales and showed that $\mathcal{Q}(S)$ for self-dual S are *-simple. Then J. Paseka [51] adopted the results to the non-involutive case. The characterization by means of cyclic separating sets was added in [30].

Meanwhile C.J. Mulvey and J.W. Pelletier [42,43,63] continued their work on spectra of C*-algebras and introduced an alternative notion of quantale representation where the endomorphism quantales $\mathcal{Q}(S)$ are replaced by weak spectra of operator algebras $\mathrm{Max}_w\, \mathcal{B}(H)$ and the corresponding notion of spatiality is then different from our one. We refer the reader to the lucid summary [44].

J. Rosický [78] characterized right-sided spatial quantales and D. Kruml [29] generalized the result for arbitrary quantales and *-quantales.

The focus of applications of Girard quantales lies more in linear logic [18,87] and their affinity to von Neumann (and afterwards simple) quantales was not yet been studied, except for example of relational quantales due to K.I. Rosenthal [76]. It was mentioned in [56] that $\mathrm{Max}\, M_n(\mathbb{C})$ is not spatial and d is one of the maximal subspaces which is not prime. This is quite interesting because then spatialization of $\mathrm{Max}\, M_n(\mathbb{C})$ (i.e. taking the quotient determined by a strong representations) destroys its Girard structure.

4. Injectivity and projectivity in quantales

4.1. *Injective and projective quantales*

DEFINITION 4.1.1. A quantale E is called *injective* if for any two quantales Q, R, an injective morphism $f : Q \to R$ and any morphism $g : Q \to E$ there exists a morphism

$h : R \to E$ such that $h \circ f = g$.

$$\begin{tikzcd} Q \arrow[rr, "f"] \arrow[dr, "g"'] & & R \arrow[dl, dotted, "h"] \\ & E & \end{tikzcd}$$

LEMMA 4.1.2. *Every injective quantale is unital.*

PROOF. Suppose Q is an injective quantale. By Lemma 1.1.11 there exists a unital quantale R with an embedding $\iota : Q \to R$. Since Q is injective and id_Q is a quantale morphism, there exists a quantale morphism $f : R \to Q$ such that $f \circ \iota = \mathrm{id}_Q$ and f is thus a surjection. That means, by Lemma 1.1.6, that Q is unital. □

THEOREM 4.1.3. *There are no non-trivial injective quantales.*

PROOF. Let Q be an injective quantale; by Lemma 4.1.2 Q is unital. Note that the map $f : Q \to \mathcal{Q}(Q)$ defined as $f(a)(b) = a \cdot b$ is a quantale embedding. Since Q is injective there exists a quantale morphism $g : \mathcal{Q}(Q) \to Q$ satisfying $g \circ f = \mathrm{id}_Q$. Evidently, g is a surjective morphism from the simple quantale $\mathcal{Q}(Q)$ onto Q. Therefore Q is either a trivial quantale or isomorphic to $\mathcal{Q}(Q)$. But the latter is impossible because if Q has more than one element, f is not a surjection. □

Now, let us turn to the case of projective quantales. Recall that the following definition is equivalent with the usual one (see [12]).

DEFINITION 4.1.4. Let $\mathcal{K}$ be an algebraic category. An object P from $\mathcal{K}$ is said to be *projective* if for any objects Q, R from $\mathcal{K}$, a surjective morphism $f : Q \to R$ and any morphism $g : P \to R$ there exists a morphism $h : P \to Q$ that satisfies $f \circ h = g$.

$$\begin{tikzcd} Q \arrow[rr, "f"] & & R \\ & P \arrow[ul, dotted, "h"] \arrow[ur, "g"'] & \end{tikzcd}$$

An object Q from $\mathcal{K}$ is called a *retract* of an object R from $\mathcal{K}$ if there exist morphisms $u : Q \to R$ and $v : R \to Q$ such that $v \circ u = \mathrm{id}_Q$.

Clearly, u is an injection and v is a surjection.

The following is a well-known stuff from category theory (see e.g. [12]).

THEOREM 4.1.5. *Let $\mathcal{K}$ be an algebraic category. Then*

1. *Every free object is projective.*
2. *A retract of a projective object is also projective.*
3. *A projective object is a retract of a free object.*
4. *Let P_i, $i \in I$, be projective. Then $\sum_{i \in I} P_i$ is projective, too.*

Therefore, projective quantales are exactly the retracts of free objects.

To establish a description of projective quantales we need the following definition.

Definition 4.1.6. A sup-lattice L is said to be *completely distributive* (shortly CDL) if any $a \in L$ can be expressed as $a = \bigvee\{x\colon\ x \lll a\}$ where $x \lll a$ iff $a \leqslant \bigvee B \Rightarrow \exists b \in B : x \leqslant b$ for any subset $B \subseteq L$. In that case we say that x is *completely below* a.

Definition 4.1.7. A quantale Q is called a $\star$*-stable CDL* if it has the following properties:

1. Q is a CDL.
2. $a \lll b, c \lll d \Rightarrow ac \lll bd$.
3. $a \lll bc \Rightarrow \exists b', c' \in Q, b' \lll b, c' \lll c$ such that $a = b'c'$.

Q is called a *weakly* $\star$*-stable CDL* if conditions 1 and 2 hold and the condition 3 is weakened to:

4. $a \lll bc \Rightarrow \exists b', c' \in Q, b' \lll b, c' \lll c$ such that $a \leqslant b'c'$.

Note that the condition 4 in the definition of a weakly $\star$-stable CDL is superfluous – it clearly follows from the condition 1 ($a \lll bc = \bigvee\{b'c'\colon\ b' \lll b, c' \lll c\}$).

Li Yong-ming, Zhou Meng and Li Zhi-hui in 2002 in [35] asked the following question: is there a weakly $\star$-stable CDL which is not a $\star$-stable CDL? The answer is surprisingly easy.

Example 4.1.8. Let P_3 be the chain $\{0 < a < 1\}$. Define a multiplication on P_3 as follows:

$\cdot$	0	a	1
0	0	0	0
a	0	0	0
1	0	0	1

P_3 is a weakly $\star$-stable CDL, P_3 is not $\star$-stable since $a \lll 1 \cdot 1$ but a cannot be a product of two elements.

In fact, one has the following proposition that disproves the second part of the Theorem 1 in [35].

Proposition 4.1.9. *There is a finite $\star$-stable CDL that is not projective.*

Proof. Let us take the two-point Boolean algebra $\mathbf{2}$ and the free quantale $\mathcal{P}(\{a\}^+)$ over a one element set. Clearly, there is a surjective quantale morphism $f : \mathcal{P}(\{a\}^+) \to \mathbf{2}$ defined by $0 \mapsto 0$, $\{a\} \mapsto 1$.

Let us assume that $\mathbf{2}$ is projective. Then there is a quantale morphism $h : \mathbf{2} \to \mathcal{P}(\{a\}^+)$ such that the following diagram commutes:

$$\begin{tikzcd}
\mathcal{P}\big(\{a\}^+\big) \arrow[rr, "f"] & & \mathbf{2} \\
& \mathbf{2} \arrow[ul, dotted, "h"] \arrow[ur, "\mathrm{id}_{\mathbf{2}}"'] &
\end{tikzcd}$$

Consequently, $h(1) = h(1{\cdot}1) = h(1)h(1)$, i.e. $h(1)$ is an idempotent from $\mathcal{P}(\{a\}^+)$. But the only idempotent from $\mathcal{P}(\{a\}^+)$ is the bottom element. Therefore $1 = f \circ h(1) = f(0) = 0$, a contradiction. □

Now consider a quantale Q and the free quantale $\mathcal{P}(Q^+)$ over Q together with the inclusion map $\varphi : Q \to \mathcal{P}(Q^+)$. Take a map id_Q: by the definition of a free quantale there exists a unique morphism $\sigma : \mathcal{P}(Q^+) \to Q$ such that $\sigma \circ \varphi = \mathrm{id}_Q$. The map defined as $\sigma(A) = \bigvee_Q \{a_1 \cdot \cdots \cdot a_n \colon a_1 \ldots a_n = a \in A\}$ is a quantale morphism and satisfies this condition.

PROPOSITION 4.1.10. *A quantale Q is projective iff there exists a morphism $h : Q \to \mathcal{P}(Q^+)$ such that $\sigma \circ h = \mathrm{id}_Q$.*

PROOF. Since $\sigma \circ \varphi = \mathrm{id}_Q$, σ is a surjection. Then, by the projectivity of Q, such a morphism $h : Q \to \mathcal{P}(Q)$ satisfying the proposition exists.

$$\begin{tikzcd}
\mathcal{P}(Q^+) \arrow[rr, "\sigma"] & & Q \\
& Q \arrow[ul, dotted, "h"] \arrow[ur, "\mathrm{id}_Q"'] &
\end{tikzcd}$$

The converse implication follows from 4.1.5. Namely $\mathcal{P}(Q^+)$ is a free quantale and Q is a retract of $\mathcal{P}(Q^+)$. □

Note that a retract of a weakly $\star$-stable CDL is again a weakly $\star$-stable CDL (this can be easily verified or it follows directly from Example 4.2.7).

LEMMA 4.1.11. *The free quantale $\mathcal{P}(X^+)$ is a $\star$-stable CDL.*

PROOF. Since $\mathcal{P}(X^+)$ is a power set (hence every $A \in \mathcal{P}(X^+)$ can be expressed as a union of singletons that are certainly completely below A), it is a CDL.

From $\emptyset \neq A \lll B$ it follows that $A = \{b\}$ for some $b \in B$, so $A \lll B$, $C \lll D$, $A, C \neq \emptyset \Rightarrow AC = \{b\}\{d\} = \{bd\} \lll BD$. The cases $A = \emptyset$ or $C = \emptyset$ obviously hold.

If $\emptyset \neq A \lll BC \iff A = \{bc\} \subseteq BC \iff A = B'C'$ where $B' \lll B$ and $C' \lll C$. Again, when $A = \emptyset$, $A = \emptyset \cdot \emptyset$ and satisfies the condition. □

4.2. *Projective quantales: a general view*

Specifically, the setting here is the category $\mathcal{OS}gr$ of partially ordered semigroups in which we consider subcategories $\mathcal{K}$ containing the category *Quant* of quantales reflectively, subject to a very simple natural condition. The projectivity in question is then taken relative to the surjective quantale morphisms $h : L \to M$ for which the right adjoint $h^{\dashv} : M \to L$ belongs to $\mathcal{K}$, referred to as $\mathcal{K}$-flat projectivity introduced by Banaschewski [6] for frames.

The analog of the condition postulated by Banaschewski for frames, for the subcategory $\mathcal{K}$ of $\mathcal{OS}gr$, besides the assumption that *Quant* is reflective in $\mathcal{K}$, is as follows:

(C) For any $\varphi : A \to L$ in $\mathcal{K}$ where L is a quantale and A arbitrary, the corestriction of φ to any subquantale of L containing the image of φ also belongs to $\mathcal{K}$.

DEFINITION 4.2.1. We refer to this by saying that $\mathcal{K}$ is *corestrictive* over *Quant*.

Since $\mathcal{K}$ contains the category *Quant* reflectively we have, for any object A from $\mathcal{K}$, the universal map in $\mathcal{K}$ to quantales $\eta_A : A \to FA$ and, correspondingly, for any quantale L, the quantale morphism $\varepsilon_L : FL \to L$ such that $\varepsilon_L \circ \eta_L = \mathrm{id}_L$. Further, $\leqslant$ stands for the usual argumentwise partial order of maps between partially ordered sets, which is evidently preserved by the composition of maps in $\mathcal{K}$.

LEMMA 4.2.2.

1. *Each FA is generated by the image of η_A.*
2. $\mathrm{id}_{FL} \leqslant \eta_L \circ \varepsilon_L$ *for any quantale L.*
3. η_A *reflects order that is $f \circ \eta_A \leqslant g \circ \eta_A$ implies $f \leqslant g$ for any quantale morphisms $f, g : FA \to L$.*
4. *For any quantale L, if $h : L \to FL$ is a right inverse to $\varepsilon_L : FL \to L$ then $h \circ \varepsilon_L \leqslant$* id_{FL}.

PROOF. 1. Let $L \subseteq FA$ be the subquantale generated by $\mathrm{Im}(\eta_A)$, let $\varphi : A \to L$ be the corresponding corestriction of $\eta_A : A \to FA$, and $i : L \to FA$ be the identical subquantale embedding. Then by (C) we have a quantale morphism $h : FA \to L$ such that $h \circ \eta_A = \varphi$, hence $i \circ h \circ \eta_A = \eta_A$, and let therefore $i \circ h = \mathrm{id}_{FA}$ by the universality property of η_A. It follows that i is onto, showing $L = FA$.

2. Recall that by 1 each $b \in FL$ is the join of all $\eta_L(a) \leqslant b$ because η_L is a semigroup morphism. Now, for any such $a \in L$,

$$\eta_L(a) = \eta_L \circ (\varepsilon_L \circ \eta_L)(a) = (\eta_L \circ \varepsilon_L) \circ \eta_L(a) \leqslant (\eta_L \circ \varepsilon_L)(b)$$

since $\varepsilon_L \circ \eta_L = \mathrm{id}_L$. Hence $b \leqslant (\eta_L \circ \varepsilon_L)(b)$.

3. Let $f \circ \eta_A \leqslant g \circ \eta_A$. Then, for all $a \in A$, we have $f(\eta_A(a)) \leqslant g(\eta_A(a))$ hence $f(b) \leqslant g(b)$ for any $b \in FA$.

4. If $\varepsilon_L \circ h = \mathrm{id}_L$ then we have that $h = \mathrm{id}_{FL} \circ h \leqslant (\eta_L \circ \varepsilon_L) \circ h = \eta_L \circ (\varepsilon_L \circ h) = \eta_L$ by 2 and consequently

$$(h \circ \varepsilon_L) \circ \eta_L = h \circ \mathrm{id}_L = h \leqslant \eta_L = \mathrm{id}_{FL} \circ \eta_L$$

which implies the desired result by 3 since both $h \circ \varepsilon_L$ and id_{FL} are quantale morphisms. □

REMARK 4.2.3. As in [6], we have that $\varepsilon_L \circ \eta_L = \mathrm{id}_L$ and 2 from Lemma 4.2.2 imply that η_L is right adjoint to ε_L. Similarly, if $\varepsilon_L \circ h = \mathrm{id}_L$ for some $h : L \to FL$ then h is left adjoint to ε_L by 4 from Lemma 4.2.2 and consequently unique.

LEMMA 4.2.4. *Let A, B be in $\mathcal{K}$, let $g : A \to B$ be a $\mathcal{K}$-morphism, let K, L be quantales, and let $f : K \to L$ be a quantale morphism. Then the following diagrams*

$$\begin{tikzcd} A \arrow[r, "\eta_A"] \arrow[d, "g"'] & FA \arrow[d, "Fg"] \\ B \arrow[r, "\eta_B"] & FB \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} FK \arrow[r, "\varepsilon_K"] \arrow[d, "Ff"'] & K \arrow[d, "f"] \\ FL \arrow[r, "\varepsilon_L"] & L \end{tikzcd}$$

commute in their respective categories. Moreover, F preserves the partial order of maps and $F\eta_A$ is left adjoint to ε_{FA} in Quant.

PROOF. The commutativity of the first diagram follows from the fact that quantales are contained in $\mathcal{K}$ reflectively. For the second diagram, it is enough to check that $\varepsilon_L \circ Ff \circ \eta_K = f \circ \varepsilon_K \circ \eta_K$. By the naturality of the left side and the fact that $\varepsilon_K \circ \eta_K = \mathrm{id}_K$ we have that $f = \mathrm{id}_L \circ f = \varepsilon_L \circ \eta_L \circ f = \varepsilon_L \circ Ff \circ \eta_K$ and similarly $f \circ \varepsilon_K \circ \eta_K = f \circ \mathrm{id}_K = f$. Hence $\varepsilon_L \circ Ff = f \circ \varepsilon_K$.

For the last assertion, if $\varphi, \psi : A \to B$ are morphisms in $\mathcal{K}$ such that $\varphi \leqslant \psi$ then by naturality the following diagrams commute:

$$\begin{tikzcd} A \arrow[r, "\eta_A"] \arrow[d, "\varphi"'] & FA \arrow[d, "F\varphi"] \\ B \arrow[r, "\eta_B"] & FB \end{tikzcd} \qquad \begin{tikzcd} A \arrow[r, "\eta_A"] \arrow[d, "\psi"'] & FA \arrow[d, "F\psi"] \\ B \arrow[r, "\eta_B"] & FB \end{tikzcd}$$

Hence $(F\varphi) \circ \eta_A = \eta_B \circ \varphi \leqslant \eta_B \circ \psi = (F\psi) \circ \eta_A$. Therefore $F\varphi \leqslant F\psi$ by Lemma 4.2.2,3.

To check that $F\eta_A$ is left adjoint to ε_{FA} let us consider the following commutative diagram:

$$\begin{tikzcd} A \arrow[r, "\eta_A"] \arrow[d, "\eta_A"'] & FA \arrow[d, "F\eta_A"] & \\ FA \arrow[r, "\eta_{FA}"] & FFA \arrow[r, "\varepsilon_{FA}"] & FA \end{tikzcd}$$

Then $\mathrm{id}_{FA} \circ \eta_A = \varepsilon_{FA} \circ \eta_{FA} \circ \eta_A = \varepsilon_{FA} \circ F\eta_A \circ \eta_A$. Hence $\mathrm{id}_{FA} = \varepsilon_{FA} \circ F\eta_A$. The assertion then follows from Remark 4.2.3. □

The reflectiveness of *Quant* in $\mathcal{K}$ determines a binary relation on each quantale L as follows:

$$x \lhd a \quad \text{iff} \quad a \leqslant \varepsilon_L(b) \text{ implies } \eta_L(x) \leqslant b, \quad \text{for all } b \in FL.$$

LEMMA 4.2.5. *Let K, L be quantales. Then, for all $x, y, u, v \in K$ and for any quantale morphism $f : K \to L$, we have:*

1. *$x \leqslant y \lhd u \leqslant v$ implies $x \lhd v$.*
2. *If f is an isomorphism of quantales then $x \lhd u$ if and only if $f(x) \lhd f(u)$.*

PROOF. 1. Let $x \leqslant y \lhd u \leqslant v$. Assume that, for some $b \in FK$, $v \leqslant \varepsilon_K(b)$. Hence $u \leqslant \varepsilon_K(b)$, i.e. $\eta_K(x) \leqslant \eta_K(y) \leqslant b$.

2. We have the following commutative diagram:

$$\begin{tikzcd}
FK \arrow[r, "\varepsilon_K"] \arrow[d, "Ff"'] & K \arrow[r, "\eta_K"] \arrow[d, "f"] & FK \arrow[d, "Ff"'] \\
FL \arrow[r, "\varepsilon_L"] & L \arrow[r, "\eta_L"] & FL
\end{tikzcd}$$

Let $f(u) \leqslant \varepsilon_L(b)$ for some $b \in FL$. Evidently, $F(f)$ and $F(f^{-1})$ are mutually inverse isomorphisms of quantales. Hence $\varepsilon_L(b) = f \circ \varepsilon_K \circ F(f)^{-1}(b)$, i.e. $u \leqslant \varepsilon_K \circ F(f)^{-1}(b)$. This implies that $\eta_K(x) \leqslant F(f)^{-1}(b)$, i.e. $\eta_L \circ f(x) = F(f) \circ \eta_K(x) \leqslant b$. □

Now we may go on our result concerning $\mathcal{K}$-flat projectivity (for frames see again [6]).

THEOREM 4.2.6. *The following are equivalent for any quantale L.*

1. *L is $\mathcal{K}$-flat projective.*
2. *ε_L has a right inverse.*
3. *L is a retract of some FA, $A \in \mathcal{K}$.*
4. *For each $a \in L$, $a = \bigvee\{x \in L \mid x \lhd a\}$; further $x \cdot y \lhd a \cdot b$ whenever $x \lhd a$ and $y \lhd b$.*

PROOF. 1 ⇒ 2. By Remark 4.2.3, $(\varepsilon_L)^{\dashv} = \eta_L$ which is in $\mathcal{K}$ by its definition, and (ε_L) is onto since $\varepsilon_L \circ \eta_L = \mathrm{id}_L$. Hence, if L is $\mathcal{K}$-flat projective we can find $h : L \to FL$ such that $\varepsilon_L \circ h = \mathrm{id}_L$.

2 ⇒ 3. Evident.

3 ⇒ 1. Since a retract of a $\mathcal{K}$-flat projective object is $\mathcal{K}$-flat projective it is enough to show that each FA is $\mathcal{K}$-flat projective. Let us consider the diagram

$$\begin{tikzcd}
A \arrow[r, "\eta_A"] \arrow[d, "k^{\dashv}\circ f\circ\eta_A"'] & FA \arrow[d, "f"] \arrow[dl, dashed, "g"] \\
L \arrow[r, "k"] & M
\end{tikzcd}$$

with quantale morphisms k and f, k onto and $\mathcal{K}$-flat and f arbitrary.

Then $k^{\dashv} \circ f \circ \eta_A \in \mathcal{K}$ and hence we have a quantale morphism $g : FA \to L$ such that $g \circ \eta_A = k^{\dashv} \circ f \circ \eta_A$. Hence we have that $k \circ g \circ \eta_A = k \circ k^{\dashv} \circ f \circ \eta_A = f \circ \eta_A$ (k is onto and therefore $k \circ k^{\dashv} = \mathrm{id}_M$). This yields $k \circ g = f$ by Lemma 4.2.2(3).

2 ⇒ 4. Let us show that $x \lhd a$ iff $\eta_L(x) \leqslant h_L(a)$ for the given $h_L : L \to FL$ such that $\varepsilon_L \circ h_L = \mathrm{id}_L$. The direction (⇒) is immediate since $a = \varepsilon_L(h_L(a))$ and putting $b = h_L(a)$ we have that $\eta_L(x) \leqslant h_L(a)$. The other direction (⇐) follows from Lemma 4.2.2(4): if $a \leqslant \varepsilon_L(b)$ then we have that $h_L(a) \leqslant (h_L \circ \varepsilon_L)(b) \leqslant b$, and hence $\eta_L(x) \leqslant h_L(a)$ then implies $\eta_L(x) \leqslant b$.

Now, by Lemma 4.2.2(1), $h_L(a) = \bigvee\{\eta_L(x)\colon \eta_L(x) \leqslant h_L(a)\}$ and therefore, we have that

$$\begin{aligned}
a = \varepsilon_L\big(h(a)\big) &= \varepsilon_L\Big(\bigvee\big\{\eta_L(x)\colon \eta_L(x) \leqslant h_L(a)\big\}\Big) \\
&= \bigvee\big\{\varepsilon_L\big(\eta_L(x)\big)\colon x \lhd a\big\} = \bigvee\{x \in L\colon x \lhd a\}.
\end{aligned}$$

Further, if $x \lhd a$ and $y \lhd b$ then $\eta_L(x \cdot y) = \eta_L(x) \cdot \eta_L(y) \leqslant h_L(a) \cdot h_L(b) = h_L(a \cdot b)$ so that $x \cdot y \lhd a \cdot b$.

$4 \Rightarrow 2$. Let $h_L : L \to FL$ be the set map defined by $h_L(a) = \bigvee\{\eta_L(x)\colon\ x \lhd a\}$. Then we have that $\varepsilon_L \circ h_L = \mathrm{id}_L$ by the first part of 4 while

$$(h_L \circ \varepsilon_L)(b) = \bigvee\bigl\{\eta_L(x)\colon\ x \lhd \varepsilon_L(b)\bigr\} \leqslant b$$

since $x \lhd \varepsilon_L(b)$ implies $\eta_L(x) \leqslant b$, and hence $h_L \circ \varepsilon_L \leqslant \mathrm{id}_{FL}$.

Therefore, we have that h_L is a left adjoint to ε_L, and as such it preserves arbitrary joins. Further, for any $a, b \in L$,

$$\begin{aligned} h_L(a) \cdot h_L(b) &= \bigvee\bigl\{\eta_L(x)\colon\ x \lhd a\bigr\} \cdot \bigvee\bigl\{\eta_L(y)\colon\ y \lhd b\bigr\} \\ &= \bigvee\bigl\{\eta_L(x) \cdot \eta_L(y)\colon\ x \lhd a, y \lhd b\bigr\} \\ &\leqslant \bigvee\bigl\{\eta_L(x \cdot y)\colon\ x \cdot y \lhd a \cdot b\bigr\} \leqslant \bigvee\bigl\{\eta_L(z)\colon\ z \lhd a \cdot b\bigr\} \\ &= h_L(a \cdot b). \end{aligned}$$

To check the converse inequality, we have

$$\begin{aligned} h_L(a \cdot b) &= h\bigl(\varepsilon_L\bigl(h_L(a)\bigr) \cdot \varepsilon_L\bigl(h_L(b)\bigr)\bigr) = h_L\bigl(\varepsilon_L\bigl(h_L(a) \cdot h_L(b)\bigr)\bigr) \\ &= (h_L \circ \varepsilon_L)\bigl(h_L(a) \cdot h_L(b)\bigr) \leqslant \mathrm{id}_{FL}\bigl(h_L(a) \cdot h_L(b)\bigr) = h_L(a) \cdot h_L(b). \end{aligned}$$

All in all, this shows that h_L is a quantale morphism, right inverse to ε_L. □

EXAMPLE 4.2.7. Let $\mathcal{K} = \mathcal{OS}gr$. Then the reflection into quantales goes the following way. For any partially ordered semigroup A, let $\mathcal{D}(A)$ denote the down-set lattice of A. $\mathcal{D}(A)$ is a sup-lattice with respect to arbitrary unions. Multiplication on $\mathcal{D}(A)$ is defined as follows:

$$X \cdot Y = {\downarrow}\{xy\colon\ x \in X,\ y \in Y\} = \{z \in A\colon\ \exists x \in X,\ \exists y \in Y \text{ such that } z \leqslant xy\}.$$

Then $\mathcal{D}(A)$ becomes a quantale. Moreover, for any morphism $f : A \to B$ of partially ordered semigroups we have an induced quantale morphism $\mathcal{D}(f) : \mathcal{D}(A) \to \mathcal{D}(B)$ defined by the prescription $f(U) = {\downarrow}\{f(x)\colon\ x \in U\}$. From Lemma 4 in [35] we know that $\mathcal{D}$ is the reflection from $\mathcal{OS}gr$ into $\mathcal{Q}uant$. Moreover, the adjunction map $\varepsilon_L : \mathcal{D}(L) \to L$, for any quantale L, is the join map and for any partially ordered semigroup A, $x \in A$, and any downset U, $\eta_A(x) = {\downarrow} x \leqslant U$ iff $x \in U$, and hence the relation $\lhd$ now has the following concrete form: $x \lhd a$ iff $a \leqslant \bigvee U$ implies $x \in U$, for all $U \in \mathcal{D}(L)$. Therefore $\lhd = \lll$ and $\mathcal{OS}gr$-flat projective quantales are exactly weakly $\star$-stable CDLs, a result from [35].

COROLLARY 4.2.8. *If Q is a projective quantale, it is a weakly $\star$-stable CDL.*

PROOF. From Example 4.2.7 we know that any $\mathcal{OS}gr$-flat projective quantale is weakly $\star$-stable CDL and clearly any projective quantale is $\mathcal{OS}gr$-flat projective. □

4.3. *Notes on Section 4*

The study of injectivity and projectivity in quantales was initiated by Li Yongming, Zhou Meng and Li Zhihui [35]. They proved Lemma 4.1.2 that every injective quantale is unital and derived from it that there are no non-trivial injective quantales (Theorem 4.1.3).

They also attempted to characterize projective quantales. They introduced the notions of a (weakly) $\star$-stable CDL and proved that any free quantale is a $\star$-stable CDL (Theorem 4.1.11) and that any projective quantale is a weakly $\star$-stable CDL (Corollary 4.2.8). Unfortunately, there is an unreparable error in their proof that any $\star$-stable CDL is projective (see our counterexample 4.1.9).

The idea of describing general forms of projectivity in topological spaces and locales comes from the work of M.H. Escardó on properly injective spaces and injective locales over perfect embeddings [15,16]. This idea was developed by B. Banaschewski to the notion of $\mathcal{K}$-flat projectivity in frames.

The second author generalized Banaschewski's work in the manuscript [60]. The results in [35] concerning weakly $\star$-stable CDLs are then immediate corollaries of this generalization.

In the text we have cited only a few papers. Therefore it may be useful to specify suitable references for various relevant topics as in done below.

- Categorical and algebraic questions: [2,3,7,12,36,37].
- General lattice theory: [8,10,21,66,80].
- Logic of quantum mechanics: [9,25,64].
- Frames (basic references): [22,65,83].
- Quantales (basic references): [23,44,56,67,75,76].
- C*-algebras (basic references): [4,17,24,61].
- Frames (other references): [6,15,16].
- Quantales (initial sources): [14,28,38,39,84–86].
- Quantales (simple quantales and spatiality): [29–31,51,54,58,62,78].
- Quantale modules and quantaloids: [1,19,20,48,53,55,57,59,79,81].
- Quantales and C*-algebras: [11,32,33,40,42,43,63,77].
- Quantales (categorical questions): [35,46,60].
- Quantales (linear and other substructural logics, computer science): [18,26,27,34,47,68, 69,82,87].
- Quantales (quantum logic): [5,13,71–74].
- Quantales (separation axioms): [49,50].
- Quantales (other references): [41,45,52,70].

Acknowledgements

The first author was supported by the Ministry of Education of the Czech Republic under the project MSM0021622409 and the second author was supported by the Grant Agency of the Czech Republic under the grant No. 201/06/0664.

References

[1] S. Abramsky, S. Vickers, Quantales, observational logic and process semantics, Math. Structures Comput. Sci. 3 (1993) 161–227.

[2] J. Adámek, Theory of Mathematical Structures, Reidel, Dordrecht, 1983.

[3] J. Adámek, Abstract and Concrete Categories, Wiley, New York, 1990.

[4] C.A. Akemann, A Gelfand representation theory of C*-algebras, Pacific J. Math. 6 (1970) 305–317.

[5] H. Amira, B. Coecke, I. Stubbe, How quantales emerge by introducing induction within the operational approach, Helv. Phys. Acta 71 (1998) 554–572.

[6] B. Banaschewski, Projective frames: A general view, Cah. Topol. Géom. Différ. Catég. XLVI (2005) 301–312.

[7] M. Barr, *-Autonomous Categories, Lecture Notes in Math., vol. 752, Springer-Verlag, Berlin, 1979.

[8] G. Birkhoff, Lattice Theory, Amer. Math. Soc., New York, Providence, RI, 1940.

[9] G. Birkhoff, J. von Neumann, The logic of quantum mechanics, Ann. of Math. 37 (1936) 823–843.

[10] T.S. Blyth, M.F. Janowitz, Residuation Theory, Internat. Ser. Monogr. Pure Appl. Math., vol. 102, Pergamon, New York, 1972.

[11] F. Borceux, J. Rosický, G. van den Bossche, Quantales and C^*-algebras, J. London Math. Soc. 40 (1989) 398–404.

[12] F. Borceux, Handbook of Categorical Algebra 1, Basic Category Theory, Cambridge, Cambridge Univ. Press, 1994.

[13] B. Coecke, I. Stubbe, Operational resolutions and state transitions in a categorical setting, Found. Phys. Lett. 12 (1999) 29–49.

[14] R.P. Dilworth, Non-commutative residuated lattices, Trans. Amer. Math. Soc. 46 (1939) 426–444.

[15] M.H. Escardo, Properly injective spaces and function spaces, Topology Appl. 89 (1–2) (1998) 75–120.

[16] M.H. Escardo, Injective locales over perfect embeddings and algebras of the upper powerlocale monad, Appl. Gen. Topology 4 (1) (2003) 193–200.

[17] R. Giles, H. Kummer, A non-commutative generalization of topology, Indiana Univ. Math. J. 21 (1) (1971) 91–102.

[18] J.-Y. Girard, Linear logic, Theoret. Comput. Sci. 50 (1997) 1–102.

[19] R.P. Gylys, Quantal sets and sheaves over quantales, Lithuanian Math. J. 34 (1) (1994) 8–29.

[20] R.P. Gylys, Sheaves on quantaloids, Lithuanian Math. J. 40 (2) (2000) 105–134.

[21] D.A. Higgs, K.A. Rowe, Nuclearity in the category of complete semilattices, J. Pure Appl. Algebra 57 (1989) 67–78.

[22] P.T. Johnstone, Stone Spaces, Cambridge Univ. Press, Cambridge, 1982.

[23] A. Joyal, M. Tierney, An extension of the Galois theory of Grothendieck, Mem. Amer. Math. Soc. 309 (1984).

[24] R. Kadison, J. Ringrose, Fundamentals of the Theory of Operator Algebras, vols. I, II: Advanced Theory, Academic Press, New York, 1997.

[25] G. Kalmbach, Orthomodular Lattices, Academic Press, London, 1983.

[26] N. Kamide, Quantized linear logic, involutive quantales and strong negation, Studia Logica 77 (3) (2004) 355–384.

[27] N. Kamide, On a logic of involutive quantales, MLQ Math. Log. Q. 51 (6) (2005) 579–585.

[28] W. Krull, Axiomatische Begründung der allgemeinen Idealtheorie, Sitz. Phys.-Med. Soc. Erlangen 56 (1924) 47–63.

[29] D. Kruml, Spatial quantales, Appl. Categ. Structures 10 (2002) 49–62.

[30] D. Kruml, J. Paseka, On simple and semisimple quantales, in: Topology Atlas Invited Contributions, Topology Atlas, Toronto, 2002, pp. 1–13.

[31] D. Kruml, Distributive quantales, Appl. Categ. Struct. 11 (2003) 561–566.

[32] D. Kruml, J.W. Pelletier, P. Resende, J. Rosický, On quantales and spectra of C*-algebras, Appl. Categ. Struct. 11 (2003) 543–560.

[33] D. Kruml, P. Resende, On quantales that classify C*-algebras, Cah. Topol. Géom. Différ. Catég. XLV (4) (2004) 287–296.

[34] J. Lambek, Some lattice models of bilinear logic, Algebra Universalis 34 (1995) 541–550.

[35] Li Yongming, Zhou Meng, Li Zhihui, Projective and injective objects in the category of quantales, J. Pure Appl. Algebra 176 (3) (2002) 249–258.

[36] S. Mac Lane, Categories for the Working Mathematician, Springer-Verlag, New York, 1998.

[37] E.G. Manes, Algebraic Theories, Springer-Verlag, New York, 1980.

[38] C.J. Mulvey, Are there sheaves in the foundations of physics, lecture, Oxford Univ., 1975.

[39] C.J. Mulvey, Rend. Circ. Mat. Palermo (2) Suppl. 12 (1986) 99–104.

[40] C.J. Mulvey, J.W. Pelletier, A quantisation of the calculus of relations, Canad. Math. Soc. Conf. Proc. 13 (1992) 345–360.

[41] C.J. Mulvey, M. Nawaz, Quantales: Quantal sets, in: Theory Decis. Lib. Ser. B Math. Statist. Methods, vol. 32, Kluwer Acad. Publ., Dordrecht, 1995, pp. 159–217.

[42] C.J. Mulvey, J.W. Pelletier, On the quantisation of points, J. Pure Appl. Algebra 159 (2001) 231–295.

[43] C.J. Mulvey, J.W. Pelletier, On the quantisation of spaces, J. Pure Appl. Algebra 175 (2002) 289–325.

[44] C.J. Mulvey, Quantales, in: Encyclopaedia of Mathematics, Third Suppl., Kluwer, Dordrecht, 2002, pp. 312–314.

[45] C.J. Mulvey, P. Resende, A noncommutative theory of Penrose tilings, Internat. J. Theoret. Phys. 44 (6) (2005) 655–689.

[46] C. Nkuimi-Jugnia, Amalgamation property and epimorphisms in the category of modules over a quantale, Rap. Séminaire 303, Département de Mathématique UCL, 2000, 18 p.

[47] H. Ono, in: K. Došen, P. Schroeder-Heister (Eds.), Semantics for Substructural Logics. Substructural Logics, Oxford Univ. Press, Oxford, 1993.

[48] M. Ordelt, Modules over a quantale, Diploma thesis, Masaryk University Brno (2004) (in Czech).

[49] J. Paseka, Regular and normal quantales, Arch. Math. 22 (4) (1986) 203–210.

[50] J. Paseka, Conjunctivity in quantales, Arch. Math. (Brno) 24 (4) (1988) 173–180.

[51] J. Paseka, Simple quantales, in: Proceedings of the Eight Prague Topological Symposium 1996, Topology Atlas, 1997, pp. 314–328.

[52] J. Paseka, On some duality for orthoposets, Internat. J. Theoret. Phys. 37 (1) (1998) 155–161.

[53] J. Paseka, Hilbert Q-modules and nuclear ideals in the category of $\bigvee$-semilattices with a duality, in: Proceedings of the Eight Conference on Category Theory and Computer Science CTCS'99, Elsevier Sci., Amsterdam, 1999, pp. 1–19.

[54] J. Paseka, D. Kruml, Embeddings of quantales into simple quantales, J. Pure Appl. Algebra 148 (2000) 209–216.

[55] J. Paseka, A note on Girard bimodules, Internat. J. Theoret. Phys. 39 (2000) 805–811.

[56] J. Paseka, J. Rosický, Quantales, in: B. Coecke, D. Moore, A. Wilce (Eds.), Current Research in Operational Quantum Logic: Algebras, Categories and Languages, vol. 111, Kluwer Academic, 2000, pp. 245–261.

[57] J. Paseka, A note on nuclei of quantale modules, Cah. Topol. Géom. Différ. Catég. XLIII (1) (2002) 19–34.

[58] J. Paseka, Scott-open distributive filters and prime elements of quantales, in: Proceedings of the Klagenfurt Conference 2003 on General Algebra (AAA 66), in: Contrib. General Algebra, vol. 15, Verlag Johannes Heyn, Klagenfurt, 2004, pp. 87–98.

[59] J. Paseka, Characterization of Morita equivalence pairs of quantales, Internat. J. Theoret. Phys. 44 (2005) 875–883.

[60] J. Paseka, Projective quantales: a general view, Internat. J. Theoret. Phys., in press.

[61] G.K. Pedersen, C*-Algebras and Their Automorphism Groups, Academic Press, London, 1979.

[62] J.W. Pelletier, J. Rosický, Simple involutive quantales, J. Algebra 195 (1997) 367–386.

[63] J.W. Pelletier, Von Neumann algebras and Hilbert quantales, Appl. Categ. Struct. 5 (1997) 249–264.

[64] C. Piron, Foundations of Quantum Physics, Benjamin, London, 1976.

[65] A. Pultr, Frames, in: M. Hazewinkel (Ed.), Handbook of Algebra, vol. 3, Elsevier, Amsterdam, 2003, pp. 791–858.

[66] G.N. Raney, Tight Galois connections and complete distributivity, Trans. Amer. Math. Soc. 97 (1960) 418–426.

[67] P. Resende, Quantales and observational semantics, in: B. Coecke, D. Moore, A. Wilce (Eds.), Current Research in Operational Quantum Logic: Algebras, Categories and Languages, vol. 111, Kluwer Academic Publ., 2000, pp. 263–288.

[68] P. Resende, Quantales, finite observations and strong bisimulation, Theoret. Comput. Sci. 254 (2001) 95–149.

[69] P. Resende, S. Vickers, Localic sup-lattices and tropological systems, Theoret. Comput. Sci. 305 (1–3) (2003) 311–346.
[70] P. Resende, Sup-lattice 2-forms and quantales, J. Algebra 276 (2004) 143–167.
[71] L. Román, B. Rumbos, A characterization of nuclei in orthomodular and quantic lattices, J. Pure Appl. Algebra 73 (1991) 155–163.
[72] L. Román, B. Rumbos, Quantum logic revisited, Found. Phys. 21 (6) (1991) 727–734.
[73] L. Román, Quantum logic and linear logic, Internat. J. Theoret. Phys. 33 (6) (1994) 1163–1172.
[74] L. Román, Orthomodular lattices and quantales, Internat. J. Theoret. Phys. 44 (7) (2005) 783–791.
[75] K.J. Rosenthal, Quantales and Their Applications, Pitman Res. Notes Math. Ser., vol. 234, Longman, 1990.
[76] K.I. Rosenthal, The Theory of Quantaloids, Pitman Res. Notes Math. Ser., vol. 348, Longman, 1996.
[77] J. Rosický, Multiplicative lattices and C^*-algebras, Cah. Topol. Géom. Différ. Categ. 30 (1989) 95–110.
[78] J. Rosický, Characterizing spatial quantales, Algebra Universalis 34 (1995) 175–178.
[79] J. Rosický, Quantaloids for concurrency, Appl. Categ. Structures 9 (2001) 329–338.
[80] Z. Shmuely, The structure of Galois connections, Pacific J. Math. 54 (1974) 209–225.
[81] S.A. Solovyov, On the category Q-Mod, Preprint, 2006, 1–22.
[82] C.J. van Alten, The finite model property for knotted extensions of propositional linear logic, Algebra Universalis 48 (3) (2004) 253–271.
[83] S. Vickers, Topology via Logic, Cambridge Univ. Press, Cambridge, 1989.
[84] M. Ward, Residuations in structures over which a multiplication is defined, Duke Math. J. 3 (1937) 627–636.
[85] M. Ward, Structure residuation, Ann. of Math. 39 (1938) 558–568.
[86] M. Ward, R.P. Dilworth, Residuated lattices, Trans. Amer. Math. Soc. 45 (1939) 335–354.
[87] D.N. Yetter, Quantales and (noncommutative) linear logic, J. Symbolic Logic 55 (1) (1990) 41–64.

Section 4H
Hopf Algebras and Related Structures

Hopf Algebras in Renormalisation

Dominique Manchon

Laboratoire de mathématiques (*CNRS-UMR* 6620), *Université Blaise Pascal*, *F*63177 *Aubière cedex*, *France*
E-mail: *manchon@math.univ-bpclermont.fr*

Contents

HANDBOOK OF ALGEBRA, VOL. 5
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(07)05007-3

1. Renormalisation in physics

Systems in interaction are most common in physics. When parameters (such as mass, electric charge, acceleration, etc.) characterising the system are considered, it is crucial to distinguish between the *bare parameters*, which are the values they would take if the interaction were switched off, and the actually observed parameters. Renormalisation can be defined as any procedure able to transform the bare parameters into the actually observed ones (i.e. with interaction taken into account), which will therefore be called *renormalised*. Consider (from [CK1]) the following example: the initial acceleration of a spherical balloon is given by

$$g = \frac{m_0 - M}{m_0 + \frac{M}{2}} g_0 \tag{1}$$

where $g_0 \simeq 9.81\ \mathrm{m\,s^{-2}}$ is the gravity acceleration at the surface of the Earth, m_0 is the mass of the balloon, and M is the mass of the volume of the air occupied by it. Note that this acceleration decreases from g_0 to $-2g_0$ when the interaction (represented here by the air mass M) increases from 0 to $+\infty$. The total force $F = mg$ acting on the balloon is the sum of the gravity force $F_0 = m_0 g_0$ and Archimedes' force $-Mg_0$. The bare parameters (i.e. in the absence of air) are thus m_0, F_0, g_0 (mass, force and acceleration, respectively), whereas the renormalised parameters are:

$$m = m_0 + \frac{M}{2}, \qquad F = \left(1 - \frac{M}{m_0}\right) F_0, \qquad g = \frac{m_0 - M}{m_0 + \frac{M}{2}} g_0. \tag{2}$$

In perturbative quantum field theory an extra difficulty arises: the bare parameters are usually infinite, reflecting the fact that the idealised "isolated system" definitely cannot exist, and in particular cannot be observed. Typically bare parameters are given by divergent integrals[1] such as

$$\int_{\mathbb{R}^4} \frac{1}{1 + \|p\|^2}\, dp. \tag{3}$$

One must then subtract another infinite quantity to the bare parameter to recover the renormalised parameter, which is finite, as this one can be actually measured! Such a process takes place in two steps:

(1) a *regularisation procedure*, which replaces the bare infinite parameter by a function of one variable z which tends to infinity when z tends to some z_0.

(2) the *renormalisation procedure* itself, of combinatorial nature, which extracts an appropriate finite part from the function above when z tends to z_0. When this procedure can be carried out, the theory is called *renormalisable*.

There is usually considerable freedom in the choice of a regularisation procedure. Let us mention, among many others, the *cut-off regularisation*, which amounts to consider integrals like (3) over a ball of radius z (with $z_0 = +\infty$), and *dimensional regularisation* which consists, roughly speaking, in "integrating over a space of complex dimension z",

[1] To be precise, the physical parameters of interest are given by a *series* each term of which is a divergent integral. We do not approach here the question of convergence of this series once each term has been renormalised.

with $z_0 = d$, the actual space dimension of the physical situation (for example $d = 4$ for Minkowski space–time). In this case the function which appears is meromorphic in z with a pole at z_0 [C,HV,Sp]. The "complex-dimensional space" involved has been recently given a rigorous meaning in terms of type II spectral triples by A. Connes and M. Marcolli [CM2].

The renormalisation procedure is an algorithm of combinatorial nature, the *BPHZ algorithm* (after N. Bogoliubov, O. Parasiuk, K. Hepp and W. Zimmermann [BP,H,Z]). The combinatorial objects involved are *Feynman graphs*: to each graph[2] corresponds (by *Feynman rules*) a quantity to be renormalised, and an integer (the *loop number*) is associated to the graph. The initial data for this algorithm are determined by the choice of a *renormalisation scheme*, which consists in choosing the finite part for the "simplest" quantities, corresponding to graphs with loop number $L = 1$. For example we can simply remove the pole part at z_0 of a meromorphic function and then consider the value at z_0. This is the *minimal subtraction scheme*. The renormalisation of the "more complicated" quantities are then given recursively with respect to the loop number.

D. Kreimer discovered in the late nineties [K1,CK,CK1,CK2], that the BPHZ recursion procedure involved is based on a Hopf algebra structure on Feynman graphs. This major breakthrough has been followed by important developments both in physics and in mathematics: on the one hand Hopf algebras reveal themselves as a prominent tool in renormalisation as well as in other aspects of quantum field theory [Br,BS], and on the other hand renormalisation techniques used for a long time by physicists can now be given a meaning in the abstract context of connected filtered or graded Hopf algebras, with also purely mathematical applications, e.g. in combinatorics [ABS,EG1,EG2,EGK1,EGK2,EGGV,EGK3, EGM,EM1,FG,JR] or in number theory [GZ,MP2].

2. Background material on algebras and coalgebras

We recall some well-known definitions and facts about algebras, modules, coalgebras and comodules (see e.g. [J,Sw]), leading to the coradical filtration of a coalgebra.

2.1. *Algebras and modules*

This paragraph, devoted to the Jacobson radical of an algebra, is largely borrowed from [J]. The main result is Corollary 3 which states that the Jacobson radical of a finite-dimensional algebra is the intersection of its maximal two-sided ideals.

2.1.1. *Basic definitions* A k-algebra is by definition a k-vector space A together with a bilinear map $m : A \otimes A \to A$ which is *associative*. The associativity is expressed by the commutativity of the following diagram:

$$\begin{tikzcd}
A \otimes A \otimes A \arrow[r, "m\otimes I"] \arrow[d, "I\otimes m"] & A \otimes A \arrow[d, "m"] \\
A \otimes A \arrow[r, "m"] & A
\end{tikzcd}$$

[2] Together with external momenta, cf. Section 7.5 below.

The algebra A is *unital* if moreover there is a unit $\mathbf{1}$ in it. This is expressed by the commutativity of the following diagram:

$$\begin{tikzcd}
k \otimes A \arrow[r, "u\otimes I"] \arrow[dr, "\sim"] & A \otimes A \arrow[d, "m"] & A \otimes k \arrow[l, "I\otimes u"'] \arrow[dl, "\sim"'] \\
& A &
\end{tikzcd}$$

where u is the map from k to A defined by $u(\lambda) = \lambda\mathbf{1}$. The algebra A is *commutative* if $m \circ \tau = m$, where $\tau : A \otimes A \to A \otimes A$ is the *flip*, defined by $\tau(a \otimes b) = b \otimes a$. A subspace $J \subset A$ is called a *subalgebra* (resp. a *left ideal, right ideal, two-sided ideal*) of A if $m(J \otimes J)$ (resp. $m(J \otimes A)$, $m(A \otimes J)$, $m(J \otimes A + A \otimes J)$ is included in J.

2.1.2. *Algebras and tensor product* To any vector space V we can associate its *tensor algebra* $T(V)$. As a vector space it is defined by

$$T(V) = \bigoplus_{k\geqslant 0} V^{\otimes k},$$

with $V^{\otimes 0} = k$ and $V^{\otimes k+1} := V \otimes V^{\otimes k}$. The product is given by the *concatenation*:

$$m(v_1 \otimes \cdots \otimes v_p,\ v_{p+1} \otimes \cdots \otimes v_{p+q}) = v_1 \otimes \cdots \otimes v_{p+q}.$$

The embedding of $k = V^{\otimes 0}$ into $T(V)$ gives the unit map u. The tensor algebra $T(V)$ is also called the *free* (*unital*) *algebra generated by* V. This algebra is characterised by the following universal property: for any linear map φ from V to an algebra A there is a unique algebra morphism $\widetilde{\varphi}$ from $T(V)$ to A extending φ.

Let A and B be unital k-algebras. We put a unital algebra structure on $A \otimes B$ in the following way:

$$(a_1 \otimes b_1).(a_2 \otimes b_2) = a_1a_2 \otimes b_1b_2.$$

The unit element $\mathbf{1}_{A\otimes B}$ is given by $\mathbf{1}_A \otimes \mathbf{1}_B$, and the associativity is clear. This multiplication is given by

$$m_{A\otimes B} = (m_A \otimes m_B) \circ \tau_{23},$$

where $\tau_{23} : A \otimes B \otimes A \otimes B \to A \otimes A \otimes B \otimes B$ is defined by the flip of the two middle factors:

$$\tau_{23}(a_1 \otimes b_1 \otimes a_2 \otimes b_2) = a_1 \otimes a_2 \otimes b_1 \otimes b_2.$$

2.1.3. *Modules* Let A be any unital algebra. A *left A-module* is a k-vector space M together with a map $\alpha : A \otimes M \to M$ such that the following diagrams commute:

$$\begin{tikzcd}
A \otimes A \otimes M \arrow[r, "m\otimes I"] \arrow[d, "I\otimes\alpha"] & A \otimes M \arrow[d, "\alpha"] \\
A \otimes M \arrow[r, "\alpha"] & M
\end{tikzcd}
\qquad
\begin{tikzcd}
k \otimes M \arrow[r, "u\otimes I"] \arrow[dr, "\sim"] & A \otimes M \arrow[d, "\alpha"] \\
& A
\end{tikzcd}$$

The map α is called the action of the algebra A on M. For any $a \in A$ and $m \in M$ we usually denote by $a.m$ the action $\alpha(a \otimes m)$ of a on m. The two diagrams above express the identities:

$$(a.b).m = a.(b.m), \qquad \mathbf{1}.m = m$$

for any $a, b \in A$ and $m \in M$. The *right A-module* are defined similarly, replacing $A \otimes M$ with $M \otimes A$ (details are left to the reader). A linear subspace N of a left A-module M is called a *submodule* if $\alpha(A \otimes N) \subset N$. The intersection of all left submodules of M containing a subset P is called the *left submodule generated by P*. A left module M is *simple* if it does not contain any submodule different from $\{0\}$ or M itself. If a left module M can be written as a direct sum of simple modules, M is said to be *semi-simple*.

PROPOSITION 1. *For any left maximal ideal J of an algebra A the quotient A/J is a simple left A-module, and conversely any simple left A-module is isomorphic to a simple left A-module of this form.*

PROOF. The first assertion is immediate. Conversely, let M a simple left module and let $m \in M - \{0\}$. Let J_m be the annihilator of m. This is a left ideal of A, and by simplicity of M the map:

$$\begin{aligned} \phi_m : A &\to M \\ a &\mapsto a.m \end{aligned}$$

gives rise to an morphism of left A-modules from A/J_m onto M. It is injective by definition of J_m, and surjectivity comes from the simplicity of the module M. So ϕ_m is an isomorphism. The left ideal J_m is maximal, which proves the proposition. □

Now let M be an A-module. We denote by A'_M the algebra of endomorphisms of M as an A-module, and we denote by A''_M the algebra of endomorphisms of M as an A'_M-module. Clearly any $a \in A$ gives rise to an element of A''_M. Following N. Jacobson [J, Chapter 4.3] we shall give a proof of an important *density theorem*:

THEOREM 1. *Let M be a semi-simple A-module, and let $x_1, \ldots, x_n$ be a finite collection of elements of M. Then for any $a'' \in A''_M$ there exists an element $a \in A$ such that $ax_i = a''x_i$ for any $i = 1, \ldots, n$.*

PROOF. First notice that any A-submodule N of M is an A''_M-submodule. To see this write (thanks to semi-simplicity) $M = N \oplus T$ where T is another A-submodule of M. The projection e on N with respect to this decomposition is an element of A'_M. For any $a'' \in A''_M$ we have then:

$$a''(N) = a'' \circ e(M) = e \circ a''(M) \subset N.$$

Consider then for any fixed positive integer n the semi-simple module M^n, the direct sum of n copies of M. The algebra A'_{M^n} coincides with the algebra of $n \times n$ matrices over A'_M, and the diagonal matrices over A''_M form a subalgebra of A''_{M^n}, and thus realise an embedding of A''_M into A''_{M^n}. Now let $x = (x_1, \dots, x_n) \in M^n$. Then $N = A.x$ is an A-submodule of M^n. So it is an A''_{M^n}-submodule, hence an A''_M-submodule via the diagonal embedding above. Then for any $a'' \in A''_M$ there exists $a \in A$ such that $a''x = ax$, which proves the theorem. □

COROLLARY 1. *On a semi-simple finite-dimensional module M the natural map from A into A''_M is surjective.*

2.1.4. *The Jacobson radical* Let A be a k-algebra. The *radical* $\operatorname{rad} M$ of a left module is by definition the intersection of all maximal submodules of M. When the module M is the algebra A itself, the radical $\operatorname{rad} A$ is the intersection of all maximal left ideals, and is called the *Jacobson radical* of the algebra A. An alternative definition of the Jacobson radical is the following: a *primitive ideal* is the annihilator of a simple module. In view of Proposition 1, any primitive ideal is the annihilator of A/J where J is a maximal left ideal. Of course a primitive ideal is two-sided.

LEMMA 1. *Any primitive ideal is an intersection of maximal left ideals.*

PROOF. Any primitive ideal J is by definition the annihilator of a simple module M. The annihilator J_m of any $m \in M - \{0\}$ is then a maximal left ideal containing J, and it is clear that we have:

$$J = \bigcap_{m \in M - \{0\}} J_m.$$

□

PROPOSITION 2. *The Jacobson radical of A is the intersection of its primitive ideals.*

PROOF. Let us call P the intersection of all primitive ideals of A. By Lemma 1 and Proposition 1, P is indeed the intersection of all maximal left ideals. □

LEMMA 2 *(Nakayama's lemma). Let M a finitely generated A-module, and let N, L two submodules of M such that $M = L + N$ and $N \subset \operatorname{rad} M$. Then $L = M$.*

PROOF. Suppose that L is strictly contained in M. As M is finitely generated there exists a maximal nontrivial submodule $\widetilde{L}$ containing L. It contains N as well by definition of $\operatorname{rad} M$. Then $\widetilde{L}$ contains $L + N$, so $L + N$ cannot be equal to M. □

COROLLARY 2. *The Jacobson radical of a finite-dimensional algebra is nilpotent.*

PROOF. Let A a finite-dimensional algebra with radical R. Observe first that for any A-module M we have the inclusion:

$$R.M \subset \operatorname{rad} M.$$

Indeed any maximal submodule N of M contains $J.M$ where J is a primitive ideal, namely the annihilator of M/N. Hence any maximal submodule of M contains $R.M$. Suppose now that A is finite-dimensional and that J is an ideal of A such that $R.J = J$. A fortiori $\operatorname{rad} J = J$. Applying Nakayama's Lemma 2 to $M = J$ and $L = \{0\}$ we get $J = \{0\}$. We immediately deduce from this fact that for any positive integer n, R^n either contains strictly R^{n+1} or is equal to $\{0\}$. As A is finite-dimensional R^n is indeed equal to $\{0\}$ for some n. □

REMARK 1. Although the definition of the Jacobson ideal is not symmetric (because we used left ideals and left modules), the Jacobson ideal itself is a symmetric notion: in other words the Jacobson radicals of algebras A and A^{opp} coincide. In order to see this one can show that $\operatorname{Rad} A$ is the biggest two-sided ideal J such that $1-x$ is invertible for any $x \in J$ [B, §6.3]. This definition is indeed symmetric.

2.1.5. *Maximal two-sided ideals* It is easily seen that any maximal two-sided ideal in an algebra A is primitive. The converse is false in general: for example, in the enveloping algebra of the non-trivial two-dimensional Lie algebra, the ideal $\{0\}$ is primitive but not maximal [Di]. However we have the following result.

PROPOSITION 3. *Any finite-codimensional primitive ideal is maximal.*

Before giving a proof of this result, we need the following definitions: an algebra A is *simple* if it does not contain any proper two-sided ideal. A *division algebra* is an algebra such that any nonzero element is inversible.

LEMMA 3. *Let D be a division algebra. Then the algebra $M_n(D)$ of square $n \times n$ matrices over D is simple.*

PROOF. Let us consider for any $i, j \in \{1, \dots, n\}$ the elementary matrix e_{ij} with vanishing entries except the one on the i-th row and j-th column which is equal to the unit of D. Denote by I_i the left ideal of $M_n(D)$ consisting of matrices such that all columns vanish except the i-th column. These left ideals are all simple and isomorphic as $M_n(D)$-modules. Now let I a nonzero two-sided ideal of $M_n(D)$. Let X a nonzero element of I, and x_{kl} a nonzero entry of the matrix X. Then for any $i \in \{0, \dots, n\}$ the product $e_{ik}X$ belongs to $I_i \cap I$ and is different from 0. Then $I_i \cap I \neq \{0\}$. As I_i is simple as a left module for each i that means that I contains all the left ideals I_i, and hence $I = M_n(D)$. □

There is a converse to this result, namely any simple *Artinian* algebra (a fortiori any simple finite-dimensional algebra) is isomorphic to $M_n(D)$ where D is a division algebra. This is a particular case of the *Wedderburn–Artin theorem*, which gives a complete description of *semi-simple algebras* [DF, Chapter 1], [DK, Chapter 2].

PROOF OF PROPOSITION 3. A finite-dimensional primitive ideal I is the annihilator of a simple finite-dimensional module M. By simplicity of M the algebra $D = A'_M$ is a division algebra (Schur's lemma). The action of A on M yields (thanks to Corollary 1) a surjective algebra morphism from A onto A''_M, and hence an algebra isomorphism from A/J

onto A''_M. But A''_M is a matrix algebra over D, and then is simple (according to Lemma 3). So A/J is a simple algebra, which amounts to say that J is maximal as a two-sided ideal. □

COROLLARY 3. *In a finite-dimensional algebra, primitive ideals and maximal two-sided ideals coincide. In particular the Jacobson radical is the intersection of all maximal two-sided ideals in this case.*

2.2. *Coalgebras and comodules; the coradical filtration*

This section is mostly borrowed from M.E. Sweedler's book [Sw], particularly Chapters 1, 2, 8 and 9. Another good reference is [Ab].

2.2.1. *Coalgebras* Coalgebras are objects which are somehow dual to algebras: the axioms for coalgebras are derived from axioms for algebras by reversing the arrows of the corresponding diagrams: a k-coalgebra is by definition a k-vector space C together with a bilinear map $\Delta : C \to C \otimes C$ which is *co-associative*. The co-associativity is expressed by the commutativity of the following diagram:

$$\begin{tikzcd}
C \otimes C \otimes C & C \otimes C \arrow[l, "\Delta \otimes I"'] \\
C \otimes C \arrow[u, "I \otimes \Delta"] & C \arrow[l, "\Delta"] \arrow[u, "\Delta"']
\end{tikzcd}$$

A coalgebra C is *co-unital* if moreover there is a co-unit ε such that the following diagram commutes:

$$\begin{tikzcd}
k \otimes C & C \otimes C \arrow[l, "\varepsilon \otimes I"'] \arrow[r, "I \otimes \varepsilon"] & C \otimes k \\
& C \arrow[ul, "\sim"] \arrow[u, "\Delta"] \arrow[ur, "\sim"'] &
\end{tikzcd}$$

A subspace $J \subset C$ is called a *subcoalgebra* (resp. a *left coideal, right coideal, two-sided coideal*) of C if $\Delta(J)$ is contained in $J \otimes J$ (resp. $C \otimes J$, $J \otimes C$, $J \otimes C + C \otimes J$). The duality alluded to above can be made more precise:

PROPOSITION 4.

(i) *The linear dual C^* of a co-unital coalgebra C is a unital algebra, with product (resp. unit map) the transpose of the coproduct (resp. of the co-unity).*

(ii) *Let J a linear subspace of C. Denote by $J^\perp$ the orthogonal of J in C^*. Then:*

- *J is a two-sided coideal if and only if $J^\perp$ is a subalgebra of C^*.*
- *J is a left coideal if and only if $J^\perp$ is a left ideal of C^*.*
- *J is a right coideal if and only if $J^\perp$ is a right ideal of C^*.*
- *J is a subcoalgebra if and only if $J^\perp$ is a two-sided ideal of C^*.*

PROOF. For any subspace K of C^* we shall denote by $K^\perp$ the subspace of those elements of C on which any element of K vanishes. It coincides with the intersection of the orthogonal of K with C, via the canonical embedding $C \hookrightarrow C^{**}$. So we have for any linear subspaces $J \subset C$ and $K \subset C^*$:

$$J^{\perp\perp} = J, \qquad K^{\perp\perp} \supset K.$$

Suppose that J is a two-sided coideal. Take any ξ, η in $J^\perp$. For any $x \in J$ we have:

$$\langle \xi\eta, x\rangle = \langle \xi \otimes \eta, \Delta x\rangle = 0,$$

as $\Delta x \subset J \otimes C + C \otimes J$. So $J^\perp$ is a subalgebra of C^*. Conversely if $J^\perp$ is a subalgebra then:

$$\Delta J \subset (J^\perp \otimes J^\perp)^\perp = J \otimes C + C \otimes J,$$

which proves the first assertion. We leave it to the reader as an exercise to prove the three other assertions along the same lines. □

Dually we have the following:

PROPOSITION 5. *Let K a linear subspace of C^*. Then*:

- *$K^\perp$ is a two-sided coideal if and only if K is a subalgebra of C^*.*
- *$K^\perp$ is a left coideal if and only if K is a left ideal of C^*.*
- *$K^\perp$ is a right coideal if and only if K is a right ideal of C^*.*
- *$K^\perp$ is a subcoalgebra if and only if K is a two-sided ideal of C^*.*

PROOF. The linear dual $(C \otimes C)^*$ naturally contains the tensor product $C^* \otimes C^*$. Take as a multiplication the restriction of ${}^t\Delta$ to $C^* \otimes C^*$:

$$m = {}^t\Delta : C^* \otimes C^* \to C^*,$$

and put $u = {}^t\varepsilon : k \to C^*$. It is easily seen, by just reverting the arrows of the corresponding diagrams, that coassociativity of Δ implies associativity of m, and that the co-unit property for ε implies that u is a unit. □

The coalgebra C is *cocommutative* if $\tau \circ \Delta = \Delta$, where $\tau : C \otimes C \to C \otimes C$ is the flip. It will be convenient to use *Sweedler's notation*:

$$\Delta x = \sum_{(x)} x_1 \otimes x_2.$$

Cocommutativity expresses itself then as

$$\sum_{(x)} x_1 \otimes x_2 = \sum_{(x)} x_2 \otimes x_1.$$

Coassociativity reads in Sweedler's notation:

$$(\Delta \otimes I) \circ \Delta(x) = \sum_{(x)} x_{1:1} \otimes x_{1:2} \otimes x_2 \sum_{(x)} x_1 \otimes x_{2:1} \otimes x_{2:2} = (I \otimes \Delta) \circ \Delta(x).$$

We shall sometimes write the iterated coproduct as

$$\sum_{(x)} x_1 \otimes x_2 \otimes x_3.$$

Sometimes we shall even mix the two ways of using Sweedler's notation for the iterated coproduct, in case we want to keep partially track of how we have constructed it [DNR]. For example,

$$\begin{aligned}\Delta_3(x) &= (I\otimes\Delta\otimes I)\circ(\Delta\otimes I)\circ\Delta(x)\\ &= (I\otimes\Delta\otimes I)\Bigg(\sum_{(x)} x_1\otimes x_2\otimes x_3\Bigg)\\ &= \sum_{(x)} x_1\otimes x_{2:1}\otimes x_{2:2}\otimes x_3.\end{aligned}$$

To any vector space V we can associate its *tensor coalgebra* $T^c(V)$. It is isomorphic to $T(V)$ as a vector space. The coproduct is given by the *deconcatenation* or cuts:

$$\Delta(v_1\otimes\cdots\otimes v_n)=\sum_{p=0}^{n}(v_1\otimes\cdots\otimes v_p)\otimes(v_{p+1}\otimes\cdots\otimes v_n).$$

The co-unit is given by the natural projection of $T^c(V)$ onto k.

Let C and D be unital k-coalgebras. We put a co-unital coalgebra structure on $C\otimes D$ in the following way: the comultiplication is given by

$$\Delta_{C\otimes D}=\tau_{23}\circ(\Delta_C\otimes\Delta_D),$$

where τ_{23} is again the flip of the two middle factors, and the co-unity is given by $\varepsilon_{C\otimes D}=\varepsilon_C\otimes\varepsilon_D$.

2.2.2. *Comodules* Let C be any co-unital coalgebra. A *left C-comodule* is a k-vector space M together with a map $\Phi: M\to C\otimes M$ such that the following diagrams commute:

$$\begin{tikzcd} C\otimes C\otimes M & C\otimes M \arrow[l, "\Delta\otimes I"'] \\ C\otimes M \arrow[u, "I\otimes\Phi"'] & M \arrow[l, "\Phi"'] \arrow[u, "\Phi"'] \end{tikzcd} \qquad \begin{tikzcd} k\otimes M & C\otimes M \arrow[l, "\varepsilon\otimes I"'] \\ & C \arrow[ul, "\sim"'] \arrow[u, "\Phi"'] \end{tikzcd}$$

The notion of *right C-comodule* is defined similarly. A linear subspace N of a left C-comodule M is called a *subcomodule* if $\Phi(C)\subset C\otimes N$. The intersection of all left subcomodules of M containing a subset P is again a subcomodule, called the *left subcomodule generated by P*. It will be convenient to use again Sweedler's notation:

$$\Phi(m)=\sum_{(m)} m_1\otimes m_0,$$

with $m_0\in M$ and $m_1\in C$. The comodule property reads in Sweedler's notation:

$$\begin{aligned}(\Phi\otimes I)\circ\Phi(m) &= \sum_{(x)} m_{1:1}\otimes m_{1:2}\otimes m_0 \sum_{(m)} m_1\otimes m_{0:1}\otimes m_{0:0}\\ &= (I\otimes\Delta)\circ\Phi(m).\end{aligned}$$

We shall sometimes write the iterated coproduct as:

$$\sum_{(m)} m_1 \otimes m_2 \otimes m_0.$$

For a right comodule we have a similar behaviour with m_0 on the left. The notion of comodule is dual to the notion of module in the sense that any left (resp. right) C-comodule M admits a right (resp. left) C^*-module structure. To be precise suppose for the moment that Φ is any linear map from M to $M \otimes C$, and define $\alpha_\Phi : C^* \otimes M \to M$ as the following composition:

$$\begin{tikzcd} C^* \otimes M \arrow[r, "I\otimes\Phi"] & C^* \otimes M \otimes C \arrow[r, "\tau\otimes I"] & M \otimes C^* \otimes C \arrow[r, "I\otimes<-,->"] & M \otimes k \arrow[r, "\sim"] & M. \end{tikzcd}$$

Then:

PROPOSITION 6. (M, Φ) *is a right* (*resp. left*) C*-comodule if and only is* (M, α_Φ) *is a left* (*resp. right*) C^**-module.*

For a proof, see, e.g. Sweedler [Sw, Section 2.1]. Note that the duality property is not perfect: while the linear dual of a coalgebra is always an algebra, the linear dual of an algebra is not in general a coalgebra. However the *restricted dual* A° of an algebra A is a coalgebra. It is defined as the space of linear forms on A vanishing on some finite-codimensional ideal. Along the same lines, for a coalgebra C the only left C^*-modules related to a right C-comodule structure via Proposition 6 are the *rational* left C^*-modules, i.e. those left modules such that the linear map:

$$\begin{aligned} \rho : M &\to \mathcal{L}(C^*, M) \\ m &\mapsto (x \mapsto x.m) \end{aligned}$$

has image included in $M \otimes C$ via the embedding:

$$\begin{aligned} j : M \otimes C &\to \mathcal{L}(C^*, M) \\ m \otimes c &\mapsto \big(x \mapsto \langle x, c\rangle m\big), \end{aligned}$$

see [Sw] for details. We come now to the fundamental theorem of comodule structure theory.

THEOREM 2. *Let M be a left comodule over a coalgebra C. For any element $m \in M$ the subcomodule generated by m is finite-dimensional.*

PROOF. There is a finite collection $(c_i)_{i=1,\dots,s}$ of linearly independent elements of C and a collection $(m_i)_{i=1,\dots,s}$ of elements of M such that

$$\Phi(m) = \sum_{i=1}^{s} m_i \otimes c_i.$$

Let N be the linear subspace of M generated by $m_1, \dots, m_s$. Let us show that N is a left subcomodule of M. First note that thanks to the co-unit axiom we have:

$$m = (I \otimes \varepsilon) \circ \Phi(m) = \sum_{j=1}^{s} \varepsilon(c_j) m_j,$$

hence $m \in N$. On the other hand, considering the linear forms $(f_i)_{i=1,\dots,s}$ on C such that $f_i(c_j) = \delta_i^j$ we have:

$$\begin{aligned}
\Phi(m_i) &= (I \otimes I \otimes f_i)\big(\Phi(m_i) \otimes c_i\big) \\
&= (I \otimes I \otimes f_i)\left(\sum_{j=1}^{s} \Phi(m_j) \otimes c_j\right) \\
&= (I \otimes I \otimes f_i) \circ (\Phi \otimes I) \circ \Phi(m) \\
&= (I \otimes I \otimes f_i) \circ (I \otimes \Delta) \circ \Phi(m) \\
&= \sum_{j=1}^{s} m_j \otimes (I \otimes f_i)(\Delta c_j),
\end{aligned}$$

whence $\Phi(m_i) \in N \otimes C$, which proves the theorem. □

COROLLARY 4. *Let M be a left comodule over a coalgebra C. Then any left subcomodule of M generated by a finite set is finite-dimensional.*

PROOF. Remark that if $P = \{m_1, \dots, m_n\}$, the left subcomodule generated by P is the sum of the left comodules generated by the m_j's, and then apply Theorem 2. □

2.2.3. *Structure of coalgebras* Let C be a coalgebra. Any intersection of subcoalgebras is a subcoalgebra. To see this consider any family $(D_\alpha)_{\alpha \in \Lambda}$ of subcoalgebras of C. Then $I := \sum_\alpha D_\alpha^\perp$ is a two-sided ideal of C^*, as a sum of two-sided ideals. Hence $I^\perp$ is a subcoalgebra according to Proposition 5. But $I^\perp$ is indeed the intersection of the subcoalgebras D_α. In particular, the intersection of all subcoalgebras containing a given subset P of C will be called the subcoalgebra generated by P. We can now state the fundamental theorem of coalgebra theory.

THEOREM 3. *Let C be a coalgebra. Then the subcoalgebra generated by one single element x is finite-dimensional.*

PROOF. The coalgebra C is a left comodule over itself. Let N be the left subcomodule generated by x. According to Theorem 2, N is finite-dimensional. Then $N^\perp$ has finite codimension, equal to $\dim N$. It is a left ideal thanks to Proposition 4. The quotient space $E = C^*/N^\perp$ is a finite-dimensional left module over $C^\perp$. Let K be the annihilator of this left module. As kernel of the associated representation $\rho : C^* \to \operatorname{End} E$ it has clearly finite codimension, and it is a two-sided ideal.

Now $K^\perp$ is a subcoalgebra according to Proposition 5. Moreover it is finite-dimensional, as $\dim K^\perp = \operatorname{codim} K^{\perp\perp} \leqslant \operatorname{codim} K$. Finally $K \subset N^\perp$ implies that $N^{\perp\perp} \subset K^\perp$.

A fortiori $N \subset K^{\perp}$, so x belongs to $K^{\perp}$. The subcoalgebra generated by x is then included in the finite-dimensional subcoalgebra $K^{\perp}$, which proves the theorem. □

A coalgebra C is said to be *irreducible* if two nonzero subcoalgebras of C have always nonzero intersection. A *simple* coalgebra is a coalgebra which does not contain any proper subcoalgebra. A coalgebra C will be called *pointed* if any simple subcoalgebra of C is one-dimensional.

LEMMA 4. *Any coalgebra C contains a simple subcoalgebra.*

PROOF. According to Theorem 3 we may suppose that C is finite-dimensional, and the lemma is immediate in this case. □

PROPOSITION 7. *A coalgebra C is irreducible if and only if it contains a unique simple subcoalgebra.*

PROOF. Suppose C irreducible, and suppose that D_1 and D_2 are two simple subcoalgebras. The intersection $D_1 \cap D_2$ is nonzero, and hence, by simplicity, $D_1 = D_2$. Conversely suppose that E is the only simple subcoalgebra of C, and let D any subcoalgebra. According to Lemma 4 we have $E \subset D$, hence E is included in any intersection of subcoalgebras, which proves that C is irreducible. □

LEMMA 5. *Let $(C_\alpha)_{\alpha\in\Lambda}$ a family of subcoalgebras of a coalgebra C such that C is the direct sum of the C_α's. Then for any subcoalgebra D we have*:

$$D = \bigoplus_{\alpha\in\Lambda} D \cap C_\alpha.$$

PROOF. The sum is indeed direct and included in D. To prove the reverse inclusion, pick any y in D and decompose it inside C:

$$y = \sum_{\alpha\in\Lambda} y_\alpha$$

with $y_\alpha \in C_\alpha$ (finite sum). Consider for any $\gamma \in \Lambda$ the linear form f_γ defined by

$$\begin{aligned} f_{\gamma|C_\alpha} &= \varepsilon_{|C_\alpha} \quad && \text{if } \alpha = \gamma \\ &= 0 && \text{if } \alpha \neq \gamma. \end{aligned}$$

Then $f_\gamma(y) = \varepsilon(y_\gamma)$. Now we have:

$$\begin{aligned} (I \otimes f_\gamma) \circ \Delta(y) &= \sum_\alpha (I \otimes f_\gamma) \circ \Delta(y_\alpha) \\ &= \sum_\alpha \sum_{(y_\alpha)} (y_\alpha)_1 f_\gamma\big((y_\alpha)_2\big) \\ &= \sum_{(y_\gamma)} (y_\gamma)_1 \varepsilon\big((y_\gamma)_2\big) \end{aligned}$$

$$= (I \otimes \varepsilon) \circ \Delta(y_\gamma)$$
$$= y_\gamma.$$

This shows that y_γ is in D, which proves the lemma. □

Let us define the *coradical* of a coalgebra C as the sum R of its simple subcoalgebras. As indicated by the terminology this notion is dual to the notion of Jacobson radical of an algebra: cf. Proposition 9 below.

PROPOSITION 8. *Let R be the coradical of a coalgebra C. Then for any subcoalgebra D the coradical R_D of D is equal to $R \cap D$.*

PROOF. Any simple subcoalgebra of D is a simple subcoalgebra of C, so $R_D \subset R \cap D$. Conversely by Lemma 5, $R \cap D$ is a direct sum of simple subcoalgebras of D, hence $R \cap D \subset R_D$. □

PROPOSITION 9. *If C is a finite-dimensional coalgebra with coradical R, then $R^\perp$ is the Jacobson radical of the algebra C^*.*

PROOF. If S is a simple subcoalgebra of C it is clear from dimension considerations that $S^\perp$ is a maximal two-sided ideal of C^*. Conversely any maximal two-sided ideal of C^* is the orthogonal of a simple subcoalgebra, so $R^\perp$ is indeed the intersection of all maximal two-sided ideals of C^*. Finally (Corollary 3), maximal two-sided and primitive ideals of a finite-dimensional algebra coincide. □

2.2.4. *The wedge* Let C be a coalgebra and X, Y two linear subspaces of C. We define:

$$X \wedge Y = \{x \in C,\ \Delta x \in X \otimes C + C \otimes Y\}.$$

We define as well inductively:

$$\textstyle\bigwedge^0 X = \{0\}, \qquad \bigwedge^n X = \left(\bigwedge^{n-1} X\right) \wedge X.$$

The alternative definition in terms of the algebra structure on C^* is often more manageable, and its verification is straightforward:

$$X \wedge Y = (X^\perp Y^\perp)^\perp.$$

Here is a first application of this definition:

PROPOSITION 10.

(i) *The wedge is associative*: $(X \wedge Y) \wedge Z = X \wedge (Y \wedge Z)$.
(ii) *If X is a left coideal then* $\{0\} \wedge X = X$.
(iii) *If X is a left coideal and Y is a right coideal, then $X \wedge Y$ is a subcoalgebra of C.*
(iv) *The wedge of two subcoalgebras is a subcoalgebra.*
(v) *If $X \subset X'$ and $Y \subset Y'$, then $X \wedge Y \subset X' \wedge Y'$.*

PROOF. According to the definition we have:

$$(X \wedge Y) \wedge Z = X \wedge (Y \wedge Z) = \big(X^{\perp}Y^{\perp}Z^{\perp}\big)^{\perp}.$$

Let X (resp. Y) be a left (resp. right) coideal. According to Proposition 4, $X^{\perp}$ is a left ideal of C^* and $Y^{\perp}$ is a right ideal. The product $(X^{\perp}Y^{\perp})$ is then a two-sided ideal. Second and third assertions follow by applying Proposition 4 again, and by noticing that:

$$\{0\} \wedge X = (C^*.X^{\perp})^{\perp} = X^{\perp\perp} = X.$$

Assertion (iv) is an immediate consequence of (iii), and (v) is clear. □

The wedge admits the following comodule version: let M be a right C-comodule with coaction $\Phi : M \to M \otimes C$. Let N be a subspace of M and X be a subspace of C. We define:

$$N \wedge X = \{x \in M,\ \Phi x \in N \otimes C + M \otimes X\}.$$

One can check that if X is a right coideal the $N \wedge X$ is a subcomodule, and if N is a subcomodule then $N \subset N \wedge X$.

PROPOSITION 11. *Let R be the coradical of a coalgebra C, and let M be a right C-comodule. Then for $\{0\} \subset M$ we have*:

$$\bigcup_n \{0\} \wedge \big(\textstyle\bigwedge^n R\big) = M.$$

PROOF. Let $x \in M$, and let N be a finite-dimensional subcomodule containing x (which exists thanks to Theorem 2). If Φ denotes the coaction, then $\Phi(N) \subset N \otimes X$ where X is a finite-dimensional subspace of C. Let D be a finite-dimensional subcoalgebra containing X (which exists thanks to Theorem 3). It is clear that N is a right D-comodule.

Applying Proposition 8, the coradical of D is $R_0 = R \cap D$. Proposition 9 says that $R_0^{\perp}$ is the Jacobson radical of D^*, which is nilpotent by Corollary 2: there exists a positive integer n such that $(R_0^{\perp})^n = \{0\}$. Dualizing we get that $\bigwedge_D^n R_0 = D$, where the subscript D serves to indicate with respect to which coalgebra the wedge operation is performed. Clearly we have:

$$\textstyle\bigwedge_D^n R_0 \subset \bigwedge_C^n R_0 \subset \bigwedge_C^n R,$$

the second inclusion coming from assertion 5) of Proposition 10. We have then:

$$D \subset \textstyle\bigwedge^n R$$

(we have dropped the subscript "C" here), hence the inclusions:

$$N \subset \{0\} \wedge D \subset \{0\} \wedge \big(\textstyle\bigwedge^n R\big).$$

The initial element x belongs then to $\{0\} \wedge (\bigwedge^n R)$ for some n, which proves the assertion. □

2.2.5. *The coradical filtration* Let C be a coalgebra with coradical R. We consider for any integer $i \geqslant 0$:

$$C^i = \bigwedge^{i+1} R.$$

The following proposition is an immediate consequence of Propositions 10 and 11:

PROPOSITION 12. *$(C^i)_{i\geqslant 0}$ is an increasing sequence of subcoalgebras of C, and we have*:

$$C = \bigcup_{i\geqslant 0} C^i.$$

The coalgebra C is then endowed with an increasing filtration by subcoalgebras: its *coradical filtration*.

PROPOSITION 13. *The coproduct is compatible with the coradical filtration, in the sense that the following inclusion holds*:

$$\Delta(C^n) \subset \sum_{i=0}^{n} C^i \otimes C^{n-i}.$$

PROOF. For any $i \in \{0, \ldots, n+1\}$ we have:

$$\bigwedge^{n+1} R = \big(\bigwedge^i R\big) \wedge \big(\bigwedge^{n-i+1} R\big).$$

This is immediate for $i = 1, \ldots, n$, and comes from Proposition 10, assertion 2) for $i = 0$ and $i = n + 1$. We have then (setting $C^{-1} = \{0\}$):

$$\Delta C^n \subset \bigcap_{i=0}^{n+1} (C \otimes C^{n-i} + C^{i-1} \otimes C).$$

The right-hand side (RHS) is contained in $C^n \otimes C^n$. Choose any supplementary subspace D_i of C^{i-1} inside C^i: so $C^0 = D_0$, and $C^i = D_0 \oplus \cdots \oplus D_i$. We have:

$$\begin{aligned}(RHS) &= \bigcap_{i=0}^{n+1} \bigoplus_{r\leqslant i-1 \text{ or } s\leqslant n-i} D_r \otimes D_s \\ &= \bigoplus_{r+s\leqslant n} D_r \otimes D_s \\ &= \sum_{i=0}^{n} C^i \otimes C^{n-i},\end{aligned}$$

which proves the result. □

2.2.6. *Convolution product* Let A be an algebra and C be a coalgebra (over the same field k). Then there is an associative product on the space $\mathcal{L}(C, A)$ of linear maps from C to A called the *convolution product*. It is given by

$$\varphi * \psi = m_A \circ (\varphi \otimes \psi) \circ \Delta_C .$$

In Sweedler's notation it reads:

$$\varphi * \psi(x) = \sum_{(x)} \varphi(x_1)\psi(x_2).$$

The associativity is a direct consequence involving both associativity of A and coassociativity of C.

3. Bialgebras and Hopf algebras

3.1. *Basic definitions*

A (unital and co-unital) *bialgebra* is a vector space $\mathcal{H}$ endowed with a structure of unital algebra (m, u) and a structure of co-unital coalgebra (Δ, ε) which are compatible. The compatibility requirement is that Δ is an algebra morphism (or equivalently that m is a coalgebra morphism), ε is an algebra morphism and u is a coalgebra morphism. It is expressed by the commutativity of the three following diagrams:

$$\begin{tikzcd}
\mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H} \arrow[rr, "\tau_{23}"] & & \mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H} \arrow[d, "m\otimes m"] \\
\mathcal{H}\otimes\mathcal{H} \arrow[u, "\Delta\otimes\Delta"] \arrow[r, "m"'] & \mathcal{H} \arrow[r, "\Delta"'] & \mathcal{H}\otimes\mathcal{H}
\end{tikzcd}$$

$$\begin{tikzcd}
\mathcal{H}\otimes\mathcal{H} \arrow[r, "\varepsilon\otimes\varepsilon"] \arrow[d, "m"] & k\otimes k \arrow[d, "\sim"] & & \mathcal{H}\otimes\mathcal{H} & k\otimes k \arrow[l, "u\otimes u"'] \\
\mathcal{H} \arrow[r, "\varepsilon"] & k & & \mathcal{H} \arrow[u, "\Delta"] & k \arrow[l, "u"'] \arrow[u, "\sim"']
\end{tikzcd}$$

A *Hopf algebra* is a bialgebra $\mathcal{H}$ together with a linear map $S : \mathcal{H} \to \mathcal{H}$ called the *antipode*, such that the following diagram commutes:

$$\begin{tikzcd}
 & \mathcal{H}\otimes\mathcal{H} \arrow[rr, "S\otimes I"] & & \mathcal{H}\otimes\mathcal{H} \arrow[dr, "m"] & \\
\mathcal{H} \arrow[ur, "\Delta"] \arrow[rr, "\varepsilon"] \arrow[dr, "\Delta"'] & & k \arrow[rr, "u"] & & \mathcal{H} \\
 & \mathcal{H}\otimes\mathcal{H} \arrow[rr, "I\otimes S"] & & \mathcal{H}\otimes\mathcal{H} \arrow[ur, "m"'] &
\end{tikzcd}$$

In Sweedler's notation it reads:

$$\sum_{(x)} S(x_1)x_2 = \sum_{(x)} x_1 S(x_2) = (u \circ \varepsilon)(x).$$

In other words the antipode is an inverse of the identity I for the convolution product on $\mathcal{L}(H, H)$. The unit for the convolution is the map $u \circ \varepsilon$.

A *primitive element* in a bialgebra $\mathcal{H}$ is an element x such that $\Delta x = x \otimes 1 + 1 \otimes x$. A *grouplike element* is a nonzero element x such that $\Delta x = x \otimes x$. Note that grouplike elements make sense in any coalgebra.

A *bi-ideal* in a bialgebra $\mathcal{H}$ is a two-sided ideal which is also a two-sided co-ideal. A *Hopf ideal* in a Hopf algebra $\mathcal{H}$ is a bi-ideal J such that $S(J) \subset J$.

3.2. *Some simple examples*

3.2.1. *The Hopf algebra of a group* Let G be a group, and let kG be the group algebra (over the field k). It is by definition the vector space freely generated by the elements of G: the product of G extends uniquely to a bilinear map from $kG \times kG$ into kG, hence a multiplication $m : kG \otimes kG \to kG$, which is associative. The neutral element of G gives the unit for m. The space kG is also endowed with a co-unital coalgebra structure, given by

$$\Delta\Big(\sum \lambda_i g_i\Big) = \sum \lambda_i . g_i \otimes g_i \quad \text{and} \quad \varepsilon\Big(\sum \lambda_i g_i\Big) = \sum \lambda_i .$$

This defines the *coalgebra of the set* G: it does not take into account the extra group structure on G, as the algebra structure does.

PROPOSITION 14. *The vector space kG endowed with the algebra and coalgebra structures defined above is a Hopf algebra. The antipode is given by*

$$S(g) = g^{-1}, \quad g \in G.$$

PROOF. The compatibility of the product and the coproduct is an immediate consequence of the following computation: for any $g, h \in G$ we have

$$\Delta(gh) = gh \otimes gh = (g \otimes g)(h \otimes h) = \Delta g . \Delta h.$$

Now $m(S \otimes I)\Delta(g) = g^{-1}g = e$ and similarly for $m(I \otimes S)\Delta(g)$. But $e = u \circ \varepsilon(g)$ for any $g \in G$, so the map S is indeed the antipode. □

REMARK 2. If G were only a semigroup, the same construction would lead to a bialgebra structure on kG: the Hopf algebra structure (i.e. the existence of an antipode) reflects the group structure (the existence of the inverse). We have $S^2 = I$ in this case, but involutivity of the antipode is not true for general Hopf algebras.

3.2.2. *Tensor algebras* There is a natural structure of cocommutative Hopf algebra on the tensor algebra $T(V)$ of any vector space V. Namely we define the coproduct Δ as the unique algebra morphism[3] from $T(V)$ into $T(V) \otimes T(V)$ such that:

$$\Delta(1) = 1 \otimes 1, \qquad \Delta(x) = x \otimes 1 + 1 \otimes x, \ x \in V.$$

[3] For the existence of this unique algebra morphism, see 2.1.2 above.

We define the co-unit as the algebra morphism such that $\varepsilon(1) = 1$ and $\varepsilon_{|V} = 0$. This endows $T(V)$ with a cocommutative bialgebra structure. We claim that the principal anti-automorphism

$$S(x_1 \otimes \cdots \otimes x_n) = (-1)^n x_n \otimes \cdots \otimes x_1$$

verifies the axioms of an antipode, so that $T(V)$ is indeed a Hopf algebra. For $x \in V$ we have $S(x) = -x$, hence $S * I(x) = I * S(x) = 0$. As V generates $T(V)$ as an algebra this is easy to check.

3.2.3. *Enveloping algebras* Let $\mathfrak{g}$ a Lie algebra. The universal enveloping algebra is the quotient of the tensor algebra $T(\mathfrak{g})$ by the ideal J generated by $x \otimes y - y \otimes x - [x, y]$, $x, y \in \mathfrak{g}$.

LEMMA 6. *J is a Hopf ideal, i.e. $\Delta(J) \subset \mathcal{H} \otimes J + J \otimes \mathcal{H}$ and $S(J) \subset J$.*

PROOF. The ideal J is generated by primitive elements (according to Proposition 16 below), and any ideal generated by primitive elements is a Hopf ideal (very easy and left to the reader). □

The quotient of a Hopf algebra by a Hopf ideal is a Hopf algebra. Hence the universal enveloping algebra $\mathcal{U}(\mathfrak{g})$ is a cocommutative Hopf algebra.

3.3. *Some properties of Hopf algebras*

We summarise in the proposition below the main properties of the antipode in a Hopf algebra.

PROPOSITION 15 *(cf. [Sw, Proposition 4.0.1]). Let $\mathcal{H}$ be a Hopf algebra with multiplication m, comultiplication Δ, unit $u : 1 \mapsto \mathbf{1}$, co-unit ε and antipode S. Then:*

(i) *$S \circ u = u$ and $\varepsilon \circ S = \varepsilon$.*

(ii) *S is an algebra antimorphism and a coalgebra antimorphism, i.e. if τ denotes the flip we have:*

$$m \circ (S \otimes S) \circ \tau = S \circ m, \qquad \tau \circ (S \otimes S) \circ \Delta = \Delta \circ S.$$

(iii) *If $\mathcal{H}$ is commutative or cocommutative, then $S^2 = I$.*

For a detailed proof, see [K], or [M2, Section I.7].

PROPOSITION 16.

(i) *If x is a primitive element then $S(x) = -x$.*

(ii) *The linear subspace* Prim $\mathcal{H}$ *of primitive elements in $\mathcal{H}$ is a Lie algebra.*

PROOF. If x is primitive, then $(\varepsilon \otimes \varepsilon) \circ \Delta(x) = 2\varepsilon(x)$. On the other hand, $(\varepsilon \otimes \varepsilon) \circ \Delta(x) = \varepsilon(x)$, so $\varepsilon(x) = 0$. Then:

$$0 = (u \circ \varepsilon)(x) = m(S \otimes I)\Delta(x) = S(x) - x.$$

Now let x and y be primitive elements of $\mathcal{H}$. Then we can easily compute:

$$\begin{aligned}
&\Delta(xy - yx)\\
&\quad = (x \otimes \mathbf{1} + \mathbf{1} \otimes x)(y \otimes \mathbf{1} + \mathbf{1} \otimes y) - (y \otimes \mathbf{1} + \mathbf{1} \otimes y)(x \otimes \mathbf{1} + \mathbf{1} \otimes x)\\
&\quad = (xy - yx) \otimes \mathbf{1} + \mathbf{1} \otimes (xy + yx) + x \otimes y + y \otimes x - y \otimes x - x \otimes y\\
&\quad = (xy - yx) \otimes \mathbf{1} + \mathbf{1} \otimes (xy - yx).
\end{aligned}$$

□

4. Connected Hopf algebras

We introduce the crucial property of connectedness for bialgebras. The main interest resides in the possibility to implement recursive procedures in connected bialgebras, the induction taking place with respect to a filtration (e.g. the coradical filtration) or a grading. An important example of these techniques is the recursive construction of the antipode, which then "comes for free", showing that any connected bialgebra is in fact a connected Hopf algebra.

4.1. *Connected graded bialgebras*

Let k be a field with characteristic zero. We shall denote by $k[[t]]$ the ring of formal series on k, and by $k[t^{-1}, t]]$ the field of Laurent series on k. A *graded Hopf algebra* on k is a graded k-vector space:

$$\mathcal{H} = \bigoplus_{n \geqslant 0} \mathcal{H}_n$$

endowed with a product $m : \mathcal{H} \otimes \mathcal{H} \to \mathcal{H}$, a coproduct $\Delta : \mathcal{H} \to \mathcal{H} \otimes \mathcal{H}$, a unit $u : k \to \mathcal{H}$, a co-unit $\varepsilon : \mathcal{H} \to k$ and an antipode $S : \mathcal{H} \to \mathcal{H}$ fulfilling the usual axioms of a Hopf algebra, and such that:

$$\mathcal{H}_p.\mathcal{H}_q \subset \mathcal{H}_{p+q}, \tag{4}$$

$$\Delta(\mathcal{H}_n) \subset \bigoplus_{p+q=n} \mathcal{H}_p \otimes \mathcal{H}_q, \tag{5}$$

$$S(\mathcal{H}_n) \subset \mathcal{H}_n. \tag{6}$$

If we do not ask for the existence of an antipode $\mathcal{H}$ we get the definition of a *graded bialgebra*. In a graded bialgebra $\mathcal{H}$ we shall consider the increasing filtration:

$$\mathcal{H}^n = \bigoplus_{p=0}^{n} \mathcal{H}_p.$$

Suppose moreover that $\mathcal{H}$ is *connected*, i.e. $\mathcal{H}_0$ is one-dimensional. Then we have:

$$\operatorname{Ker}\varepsilon = \bigoplus_{n\geqslant 1}\mathcal{H}_n.$$

PROPOSITION 17. *For any $x \in \mathcal{H}^n$, $n \geqslant 1$, we can write*:

$$\Delta x = x \otimes \mathbf{1} + \mathbf{1} \otimes x + \widetilde{\Delta} x, \qquad \widetilde{\Delta} x \in \bigoplus_{p+q=n,\, p\neq 0, q\neq 0} \mathcal{H}_p \otimes \mathcal{H}_q.$$

The map $\widetilde{\Delta}$ is coassociative on $\operatorname{Ker}\varepsilon$ *and $\widetilde{\Delta}_k = (I^{\otimes k-1} \otimes \widetilde{\Delta})(I^{\otimes k-2} \otimes \widetilde{\Delta})\dots\widetilde{\Delta}$ sends $\mathcal{H}^n$ into $(\mathcal{H}^{n-k})^{\otimes k+1}$.*

PROOF. Thanks to connectedness we clearly can write:

$$\Delta x = a(x \otimes 1) + b(1 \otimes x) + \widetilde{\Delta} x$$

with $a, b \in k$ and $\widetilde{\Delta} x \in \operatorname{Ker}\varepsilon \otimes \operatorname{Ker}\varepsilon$. The co-unit property then tells us that, with $k \otimes \mathcal{H}$ and $\mathcal{H} \otimes k$ canonically identified with $\mathcal{H}$:

$$x = (\varepsilon \otimes I)(\Delta x) = bx, \qquad x = (I \otimes \varepsilon)(\Delta x) = ax, \tag{7}$$

hence $a = b = 1$. We shall use the following two variants of Sweedler's notation:

$$\Delta x = \sum_{(x)} x_1 \otimes x_2, \tag{8}$$

$$\widetilde{\Delta} x = \sum_{(x)} x' \otimes x'', \tag{9}$$

the second being relevant only for $x \in \operatorname{Ker}\varepsilon$. if x is homogeneous of degree n we can suppose that the components x_1, x_2, x', x'' in the expressions above are homogeneous as well, and we have then $|x_1| + |x_2| = n$ and $|x'| + |x''| = n$. We easily compute:

$$\begin{aligned}(\Delta \otimes I)\Delta(x) = {} & x \otimes 1 \otimes 1 + 1 \otimes x \otimes 1 + 1 \otimes 1 \otimes x \\ & + \sum_{(x)} x' \otimes x'' \otimes 1 + x' \otimes 1 \otimes x'' + 1 \otimes x' \otimes x'' \\ & + (\widetilde{\Delta} \otimes I)\widetilde{\Delta}(x)\end{aligned}$$

and

$$\begin{aligned}(I \otimes \Delta)\Delta(x) = {} & x \otimes 1 \otimes 1 + 1 \otimes x \otimes 1 + 1 \otimes 1 \otimes x \\ & + \sum_{(x)} x' \otimes x'' \otimes 1 + x' \otimes 1 \otimes x'' + 1 \otimes x' \otimes x'' \\ & + (I \otimes \widetilde{\Delta})\widetilde{\Delta}(x),\end{aligned}$$

hence the co-associativity of $\widetilde{\Delta}$ comes from the one of Δ. Finally it is easily seen by induction on k that for any $x \in \mathcal{H}^n$ we can write:

$$\widetilde{\Delta}_k(x) = \sum_{x} x^{(1)} \otimes \cdots \otimes x^{(k+1)}, \tag{10}$$

with $|x^{(j)}| \geqslant 1$. The grading imposes:

$$\sum_{j=1}^{k+1} |x^{(j)}| = n,$$

so the maximum possible for any degree $|x^{(j)}|$ is $n - k$. □

4.2. *Connected filtered bialgebras*

A *filtered Hopf algebra* over k is a k-vector space together with an increasing $\mathbb{Z}_+$-indexed filtration:

$$\mathcal{H}^0 \subset \mathcal{H}^1 \subset \cdots \subset \mathcal{H}^n \subset \cdots, \qquad \bigcup_n \mathcal{H}^n = \mathcal{H}$$

endowed with a product $m : \mathcal{H} \otimes \mathcal{H} \to \mathcal{H}$, a coproduct $\Delta : \mathcal{H} \to \mathcal{H} \otimes \mathcal{H}$, a unit (morphism) $u : k \to \mathcal{H}$, a co-unit $\varepsilon : \mathcal{H} \to k$ and an antipode $S : \mathcal{H} \to \mathcal{H}$ fulfilling the usual axioms of a Hopf algebra, and such that:

$$\mathcal{H}^p . \mathcal{H}^q \subset \mathcal{H}^{p+q}, \tag{11}$$

$$\Delta(\mathcal{H}^n) \subset \sum_{p+q=n} \mathcal{H}^p \otimes \mathcal{H}^q, \tag{12}$$

$$S(\mathcal{H}^n) \subset \mathcal{H}^n. \tag{13}$$

If we do not ask for the existence of an antipode $\mathcal{H}$ we get the definition of a *filtered bialgebra*. For any $x \in \mathcal{H}$ we set:

$$|x| := \min\{n \in \mathbb{N},\ x \in \mathcal{H}^n\}. \tag{14}$$

Any graded bialgebra or Hopf algebra is obviously filtered by the canonical filtration associated to the grading:

$$\mathcal{H}^n := \bigoplus_{i=0}^{n} \mathcal{H}_i, \tag{15}$$

and in that case, if x is an homogeneous element, x is of degree n if and only if $|x| = n$. We say that the filtered bialgebra $\mathcal{H}$ is connected if $\mathcal{H}^0$ is one-dimensional. There is an analogue of Proposition 17 in the connected filtered case, the proof of which is very similar:

PROPOSITION 18. *For any* $x \in \mathcal{H}^n$, $n \geqslant 1$, *we can write*:

$$\Delta x = x \otimes \mathbf{1} + \mathbf{1} \otimes x + \widetilde{\Delta} x, \quad \widetilde{\Delta} x \in \sum_{p+q=n,\, p \neq 0,\, q \neq 0} \mathcal{H}^p \otimes \mathcal{H}^q. \tag{16}$$

The map $\widetilde{\Delta}$ *is coassociative on* $\operatorname{Ker} \varepsilon$ *and* $\widetilde{\Delta}_k = (I^{\otimes k-1} \otimes \widetilde{\Delta})(I^{\otimes k-2} \otimes \widetilde{\Delta}) \ldots \widetilde{\Delta}$ *sends* $\mathcal{H}^n$ *into* $(\mathcal{H}^{n-k})^{\otimes k+1}$.

The following theorem is due to S. Montgomery [Mo, Lemma 1.1].

THEOREM 4. *Let $\mathcal{H}$ be any pointed Hopf algebra. Then the coradical filtration endows $\mathcal{H}$ with a structure of a filtered Hopf algebra.*

PROOF. It only remains to show that for any $n \in \mathbb{N}$ the inclusion $S(H^n) \subset H^n$ holds, and that for any $p, q \in \mathbb{N}$:

$$\mathcal{H}^p\mathcal{H}^q \subset \mathcal{H}^{p+q},$$

which, together with Proposition 13, will prove the result. Recall that a pointed coalgebra is a coalgebra in which any simple subcoalgebra is one-dimensional. In this case any simple subcoalgebra is linearly generated by a unique grouplike element. Any grouplike element g in a Hopf algebra admits an inverse Sg, where S is the antipode. It follows that the coradical $\mathcal{H}^0$ of a pointed Hopf algebra $\mathcal{H}$ is a Hopf subalgebra of $\mathcal{H}$, precisely the Hopf algebra of the group of the grouplike elements of $\mathcal{H}$ (cf. Example 3.2.1).

The proof proceeds by induction: the inclusion $S\mathcal{H}_0 \subset \mathcal{H}_0$ obviously holds. Suppose that $S\mathcal{H}^k \subset \mathcal{H}^k$ for all $k \leqslant n-1$. Using the definition of $\mathcal{H}^n$:

$$\mathcal{H}^n = \mathcal{H}^0 \wedge \mathcal{H}^{n-1} = \mathcal{H}^{n-1} \wedge H^0$$

and the formula

$$Sx = \sum_{(x)} Sx_2 \otimes Sx_1,$$

(cf. Proposition 15) we deduce the inclusion $S\mathcal{H}^n \subset \mathcal{H}^n$. Now, suppose that the inclusion $\mathcal{H}^k\mathcal{H}^0 \subset \mathcal{H}^k$ holds for $k \leqslant n-1$ (this is obviously the case for $k = 0$). Then we have:

$$\begin{aligned}\Delta(\mathcal{H}^n\mathcal{H}^0) &\subset \big(\mathcal{H}^0 \otimes \mathcal{H} + \mathcal{H} \otimes \mathcal{H}^{n-1}\big)\big(\mathcal{H}^0 \otimes \mathcal{H}^0\big)\\ &\subset \mathcal{H}^0 \otimes \mathcal{H} + \mathcal{H} \otimes \mathcal{H}^{n-1}\mathcal{H}^0\\ &\subset \mathcal{H}^0 \otimes \mathcal{H} + \mathcal{H} \otimes \mathcal{H}^{n-1}.\end{aligned}$$

So $\mathcal{H}^n\mathcal{H}^0 \subset \Delta^{-1}(\mathcal{H}^0 \otimes \mathcal{H} + \mathcal{H} \otimes \mathcal{H}^{n-1}) = \mathcal{H}^n$. Similarly on the other side we have $\mathcal{H}^0\mathcal{H}^n \subset \mathcal{H}^n$ for any n. Suppose now that the inclusion:

$$\mathcal{H}^p\mathcal{H}^q \subset \mathcal{H}^{p+q}$$

holds for any p, q such that $p + q \leqslant n - 1$. Choose p, q with $p + q = n$ and compute:

$$\begin{aligned}\Delta(\mathcal{H}^p\mathcal{H}^q) &\subset \big(\mathcal{H}^0 \otimes \mathcal{H}^p + \mathcal{H}^p \otimes \mathcal{H}^{p-1}\big)\big(\mathcal{H}^0 \otimes \mathcal{H}^q + \mathcal{H}^q \otimes \mathcal{H}^{q-1}\big)\\ &\subset \mathcal{H}^0\mathcal{H}^0 \otimes \mathcal{H}^p\mathcal{H}^q + \mathcal{H}^p\mathcal{H}^0 \otimes \mathcal{H}^{p-1}\mathcal{H}^q + \mathcal{H}^0\mathcal{H}^q \otimes \mathcal{H}^p\mathcal{H}^{q-1}\\ &\quad + \mathcal{H}^p\mathcal{H}^q \otimes \mathcal{H}^{p-1}\mathcal{H}^{q-1}\\ &\subset \mathcal{H}^0 \otimes \mathcal{H} + \mathcal{H} \otimes \mathcal{H}^{p+q-1}\end{aligned}$$

thanks to the induction hypothesis and the property already proved when one of the indices is equal to zero. Thus $\mathcal{H}^p\mathcal{H}^q \subset \mathcal{H}^{p+q}$, which finishes the proof of the theorem. □

REMARK 3. The proof used only the property that the coradical is a Hopf subalgebra of $\mathcal{H}$. The pointedness of $\mathcal{H}$ [TW] implies this property but is not strictly necessary.

REMARK 4. The image of k under the unit map u is a one-dimensional simple subcoagebra of $\mathcal{H}$. If $\mathcal{H}$ is an irreducible coalgebra, by Proposition 7 it is the unique one, and then the coradical is $\mathcal{H}^0 = k.\mathbf{1}$. Any irreducible Hopf algebra is then pointed, and connected with respect to the coradical filtration.

4.3. *The convolution product*

An important result is that any connected filtered bialgebra is indeed a filtered Hopf algebra, in the sense that the antipode comes for free. We give a proof of this fact as well as a recursive formula for the antipode with the help of the *convolution product*: let $\mathcal{H}$ be a (connected filtered) bialgebra, and let $\mathcal{A}$ be any k-algebra (which will be called the *target algebra*): the convolution product on $\mathcal{L}(\mathcal{H}, \mathcal{A})$ is given by

$$\begin{aligned}\varphi * \psi(x) &= m_{\mathcal{A}}(\varphi \otimes \psi)\Delta(x)\\ &= \sum_{(x)} \varphi(x_1)\psi(x_2).\end{aligned}$$

PROPOSITION 19. *The map $e = u_{\mathcal{A}} \circ \varepsilon$, given by $e(\mathbf{1}) = \mathbf{1}_{\mathcal{A}}$ and $e(x) = 0$ for any $x \in \operatorname{Ker}\varepsilon$, is a unit for the convolution product. Moreover the set*

$$G(\mathcal{A}) := \{\varphi \in \mathcal{L}(\mathcal{H}, \mathcal{A}),\ \varphi(\mathbf{1}) = \mathbf{1}_{\mathcal{A}}\}$$

endowed with the convolution product is a group.

PROOF. The first statement is straightforward. To prove the second let us consider the formal series:

$$\begin{aligned}\varphi^{*-1}(x) &= \big(e - (e - \varphi)\big)^{*-1}(x)\\ &= \sum_{k\geqslant 0} (e - \varphi)^{*k}(x).\end{aligned}$$

Using $(e - \varphi)(\mathbf{1}) = 0$ we have immediately $(e - \varphi)^{*k}(\mathbf{1}) = 0$, and for any $x \in \operatorname{Ker}\varepsilon$:

$$(e - \varphi)^{*k}(x) = m_{\mathcal{A},k-1}(\underbrace{\varphi \otimes \cdots \otimes \varphi}_{k \text{ times}})\widetilde{\Delta}_{k-1}(x). \tag{17}$$

When $x \in \mathcal{H}^n$ this expression vanishes for $k \geqslant n + 1$. The formal series thus consists of only finite number of terms for any x, which proves the result. □

COROLLARY 5. *Any connected filtered bialgebra $\mathcal{H}$ is a filtered Hopf algebra. The antipode is defined by*

$$S(x) = \sum_{k\geqslant 0} (u\varepsilon - I)^{*k}(x). \tag{18}$$

It is given by $S(\mathbf{1}) = \mathbf{1}$ *and recursively by any of the two formulas for* $x \in \operatorname{Ker}\varepsilon$:

$$S(x) = -x - \sum_{(x)} S(x')x'', \tag{19}$$

$$S(x) = -x - \sum_{(x)} x'S(x''). \tag{20}$$

PROOF. The antipode, when it exists, is the inverse of the identity for the convolution product on $\mathcal{L}(\mathcal{H}, \mathcal{H})$. One just needs then to apply Proposition 19 with $\mathcal{A} = \mathcal{H}$. The two recursive formulas come directly from the two equalities:

$$m(S \otimes I)\Delta(x) = m(I \otimes S)\Delta(x) = 0$$

fulfilled by any $x \in \operatorname{Ker}\varepsilon$. □

Let $\mathfrak{g}(\mathcal{A})$ be the subspace of $\mathcal{L}(\mathcal{H}, \mathcal{A})$ formed by the elements α such that $\alpha(\mathbf{1}) = 0$. It is clearly a subalgebra of $\mathcal{L}(\mathcal{H}, \mathcal{A})$ for the convolution product. We have:

$$G(\mathcal{A}) = e + \mathfrak{g}(\mathcal{A}).$$

From now on we shall suppose that the ground field k is of characteristic zero. For any $x \in \mathcal{H}^n$ the exponential

$$e^{*\alpha}(x) = \sum_{k \geqslant 0} \frac{\alpha^{*k}(x)}{k!} \tag{21}$$

is a finite sum (ending at $k = n$). It is a bijection from $\mathfrak{g}(\mathcal{A})$ onto $G(\mathcal{A})$. Its inverse is given by

$$\operatorname{Log}(1 + \alpha)(x) = \sum_{k \geqslant 1} \frac{(-1)^{k-1}}{k} \alpha^{*k}(x). \tag{22}$$

This sum again ends at $k = n$ for any $x \in \mathcal{H}^n$. Let us introduce a decreasing filtration on $\mathcal{L} = \mathcal{L}(\mathcal{H}, \mathcal{A})$:

$$\mathcal{L}^n := \{\alpha \in \mathcal{L},\ \alpha_{|\mathcal{H}^{n-1}} = 0\}. \tag{23}$$

Clearly $\mathcal{L}_0 = \mathcal{L}$ and $\mathcal{L}_1 = \mathfrak{g}(\mathcal{A})$. We define the valuation $\operatorname{val}\varphi$ of an element φ of $\mathcal{L}$ as the biggest integer k such that φ is in $\mathcal{L}_k$. We shall consider in the sequel the ultrametric distance on $\mathcal{L}$ induced by the filtration (and associated valuation):

$$d(\varphi, \psi) = 2^{-\operatorname{val}(\varphi - \psi)}. \tag{24}$$

For any $\alpha, \beta \in \mathfrak{g}(\mathcal{A})$ let $[\alpha, \beta] = \alpha * \beta - \beta * \alpha$.

PROPOSITION 20. *We have the inclusion*

$$\mathcal{L}^p * \mathcal{L}^q \subset \mathcal{L}^{p+q}, \tag{25}$$

and moreover the metric space $\mathcal{L}$ *endowed with the distance defined by* (24) *is complete.*

PROOF. Take any $x \in \mathcal{H}^{p+q-1}$, and any $\alpha \in \mathcal{L}_p$ and $\beta \in \mathcal{L}_q$. We have:

$$(\alpha * \beta)(x) = \sum_{(x)} \alpha(x_1)\beta(x_2).$$

Recall that we denote by $|x|$ the minimal n such that $x \in \mathcal{H}^n$. Since $|x_1| + |x_2| = |x| \leqslant p + q - 1$, either $|x_1| \leqslant p - 1$ or $|x_2| \leqslant q - 1$, so the expression vanishes. Now if (ψ_n) is a Cauchy sequence in $\mathcal{L}$ it is immediate to see that this sequence is *locally stationary*, i.e. for any $x \in \mathcal{H}$ there exists $N(x) \in \mathbb{N}$ such that $\psi_n(x) = \psi_{N(x)}(x)$ for any $n \geqslant N(x)$. Then the limit of (ψ_n) exists and is clearly defined by

$$\psi(x) = \psi_{N(x)}(x).$$

□

As a corollary the Lie algebra $\mathcal{L}_1 = \mathfrak{g}(\mathcal{A})$ is *pro-nilpotent*, in the sense that it is the projective limit of the Lie algebras $\mathfrak{g}(\mathcal{A})/\mathcal{L}^n$, which are nilpotent.

4.4. *Algebra morphisms and cocycles*

Let $\mathcal{H}$ be a connected filtered Hopf algebra over k, and let $\mathcal{A}$ be a k-algebra. A *cocycle from $\mathcal{H}$ to $\mathcal{A}$* is a linear morphism $\tau : \mathcal{H} \to \mathcal{A}$ such that $\tau(xy) = \tau(yx)$ for any $x, y \in \mathcal{H}$. It is indeed a 1-cocycle in the cohomology of the Lie algebra $\mathcal{H}$ with values in $\mathcal{A}$ considered as a trivial $\mathcal{H}$-module. In the case where $\mathcal{A}$ is the ground field k cocycles are just traces.

We shall also consider unital algebra morphisms from $\mathcal{H}$ to the target algebra $\mathcal{A}$. When the algebra $\mathcal{A}$ is commutative we shall call them, by slight abuse of language, *characters*. It is clear that any character in our sense is a cocycle. We recover of course the usual notion of character when the algebra $\mathcal{A}$ is the ground field k.

The notions of character and cocycle involve only the algebra structure of $\mathcal{H}$. On the other hand the convolution product on $\mathcal{L}(\mathcal{H}, \mathcal{A})$ involves only the *coalgebra* structure on $\mathcal{H}$. Let us consider now the full Hopf algebra structure on $\mathcal{H}$ and see what happens to algebra morphisms and cocycles under the convolution product:

PROPOSITION 21. *Let $\mathcal{H}$ be a connected filtered Hopf algebra over k, and let $\mathcal{A}$ be a k-algebra. Then,*

(i) *The convolution of two cocycles in $\mathcal{L}(\mathcal{H}, \mathcal{A})$ is a cocycle.*

(ii) *If τ is a cocycle such that $\tau(\mathbf{1}) = \mathbf{1}_{\mathcal{A}}$, then the inverse τ^{*-1} is a cocycle as well.*

(iii) *In the case of a commutative algebra $\mathcal{A}$ the characters from $\mathcal{H}$ to $\mathcal{A}$ form a group $G_1(\mathcal{A})$ under the convolution product, and for any $\varphi \in G_1(\mathcal{A})$ the inverse is given by:*

$$\varphi^{*-1} = \varphi \circ S. \tag{26}$$

PROOF. Using the fact that Δ is an algebra morphism we have for any $x, y \in \mathcal{H}$:

$$f * g(xy) = \sum_{(x)(y)} f(x_1 y_1) g(x_2 y_2).$$

If f and g are cocycles we get:

$$\begin{aligned} f * g(xy) &= \sum_{(x)(y)} f(y_1 x_1) g(y_2 x_2) \\ &= f * g(yx). \end{aligned}$$

If $\mathcal{A}$ is commutative and if f and g are characters we get:

$$\begin{aligned} f * g(xy) &= \sum_{(x)(y)} f(x_1) f(y_1) g(x_2) g(y_2) \\ &= \sum_{(x)(y)} f(x_1) g(x_2) f(y_1) g(y_2) \\ &= (f * g)(x)(f * g)(y). \end{aligned}$$

The unit $e = u_{\mathcal{A}} \circ \varepsilon$ is both a cocycle and an algebra morphism. The formula for the inverse of a character comes easily from the commutativity of the following diagram:

$$\begin{tikzcd}
 & \mathcal{H}\otimes\mathcal{H} \arrow[rr, "S\otimes I"] \arrow[rrrr, dashed, bend left, "(f\circ S)*f"] & & \mathcal{H}\otimes\mathcal{H} \arrow[rr, "f\otimes f"] \arrow[dr, "m"] & & \mathcal{A}\otimes\mathcal{A} \arrow[dr, "m_{\mathcal{A}}"] & \\
\mathcal{H} \arrow[ur, "\Delta"] \arrow[dr, "\Delta"'] \arrow[rr, "\varepsilon"'] \arrow[rrrrrr, dashed, bend left=10, "e"] & & k \arrow[rr, "u"'] & & \mathcal{H} \arrow[rr, "f"'] & & \mathcal{A} \\
 & \mathcal{H}\otimes\mathcal{H} \arrow[rr, "I\otimes S"] \arrow[rrrr, dashed, bend right, "f*(f\circ S)"'] & & \mathcal{H}\otimes\mathcal{H} \arrow[rr, "f\otimes f"] \arrow[ur, "m"] & & \mathcal{A}\otimes\mathcal{A} \arrow[ur, "m_{\mathcal{A}}"'] &
\end{tikzcd}$$

Finally the fact that the inverse of a cocycle τ such that $\tau(\mathbf{1}) = \mathbf{1}_{\mathcal{A}}$ is a cocycle comes from (i) and from the formula:

$$\tau^{-1}(x) = \sum_{k \geqslant 0} (e - \tau)^{*k}(x). \qquad \square$$

We call *infinitesimal characters with values in the algebra* $\mathcal{A}$ those elements α of $\mathcal{L}(\mathcal{H}, \mathcal{A})$ such that:

$$\alpha(xy) = e(x)\alpha(y) + \alpha(x)e(y).$$

PROPOSITION 22.

(i) *Suppose that $\mathcal{A}$ is a commutative algebra. Let $G_1(\mathcal{A})$ (resp. $\mathfrak{g}_1(\mathcal{A})$) be the set of characters of $\mathcal{H}$ with values in $\mathcal{A}$ (resp. the set of infinitesimal characters of $\mathcal{H}$ with values in $\mathcal{A}$). Then $G_1(\mathcal{A})$ is a subgroup of G, the exponential restricts to a bijection from $\mathfrak{g}_1(\mathcal{A})$ onto $G_1(\mathcal{A})$, and $\mathfrak{g}_1(\mathcal{A})$ is a Lie subalgebra of $\mathfrak{g}(\mathcal{A})$.*

(ii) *Suppose that $\mathcal{A}$ is an algebra (not necessarily commutative). Let $G_2(\mathcal{A})$ (resp. $\mathfrak{g}_2(\mathcal{A})$) be the set of cocycles φ from $\mathcal{H}$ to $\mathcal{A}$ such that $\varphi(\mathbf{1}) = \mathbf{1}_{\mathcal{A}}$ (resp. $\varphi(\mathbf{1}) = 0$). Then $G_2(\mathcal{A})$ is a subgroup of $G(\mathcal{A})$, the exponential restricts to a bijection from $\mathfrak{g}_2(\mathcal{A})$ onto $G_2(\mathcal{A})$, and $\mathfrak{g}_2(\mathcal{A})$ is a Lie subalgebra of $\mathfrak{g}(\mathcal{A})$.*

PROOF. Part of these results are a reformulation of Proposition 21 and some points are straightforward. The only non-trivial point concerns $\mathfrak{g}_1(\mathcal{A})$ and $G_1(\mathcal{A})$. Take two infinitesimal characters α and β with values in $\mathcal{A}$ and compute:

$$\begin{aligned}(\alpha * \beta)(xy) &= \sum_{(x)(y)} \alpha(x_1x_2)\beta(y_1y_2) \\ &= \sum_{(x)(y)} \big(\alpha(x_1)e(y_1) + e(x_1)\alpha(y_1)\big).\big(\beta(x_2)e(y_2) + e(x_2)\alpha(y_2)\big) \\ &= (\alpha * \beta)(x)e(y) + \alpha(x)\beta(y) + \beta(x)\alpha(y) + e(x)(\alpha * \beta)(y).\end{aligned}$$

Using the commutativity of $\mathcal{A}$ we immediately get:

$$[\alpha, \beta](xy) = [\alpha, \beta](x)e(y) + e(x)[\alpha, \beta](y),$$

which shows that $\mathfrak{g}_1(\mathcal{A})$ is a Lie algebra. Now for $\alpha \in \mathfrak{g}_1(\mathcal{A})$ we have:

$$\alpha^{*n}(xy) = \sum_{k=0}^{n} \binom{n}{k} \alpha^{*k}(x)\alpha^{*(n-k)}(y),$$

as is easily seen by induction on n. A straightforward computation then yields:

$$e^{*\alpha}(xy) = e^{*\alpha}(x)e^{*\alpha}(y).$$

□

REMARK 5. The groups $G(\mathcal{A})$, $G_1(\mathcal{A})$ and $G_2(\mathcal{A})$ depend functorially on the target algebra $\mathcal{A}$. In particular, when the Hopf algebra $\mathcal{H}$ itself is commutative, the correspondence $\mathcal{A} \mapsto G_1(\mathcal{A})$ is a *group scheme*. In this case it is possible to reconstruct the Hopf algebra $\mathcal{H}$ from the group scheme by means of the Cartier–Milnor–Moore theorem: in the case when the components of the filtration are finite-dimensional, we have

$$\mathcal{H} = \big(\mathcal{U}\big(\mathfrak{g}_1(k)\big)\big)^{\circ}, \tag{27}$$

where $\mathfrak{g}_1(k)$ is the Lie algebra of infinitesimal characters with values in the base field k, where $\mathcal{U}(\mathfrak{g}_1(k))$ stands for its enveloping algebra, and where $(-)^{\circ}$ stands for the restricted dual [Sw].

In the case when the Hopf algebra $\mathcal{H}$ is not commutative this is no longer possible to reconstruct it from $G_1(k)$, and moreover the inclusion $G_2(k) \subset G(k)$ may be strict.

4.5. *Hochschild cohomology*

Let $\mathcal{H}$ be any coalgebra, and let M be a bicomodule over $\mathcal{H}$, i.e. a vector space endowed with a left comodule structure $\Phi : M \to \mathcal{H} \otimes M$ and a right comodule structure $\Psi : M \to M \otimes \mathcal{H}$ which are compatible, in the sense that the following diagram commutes:

$$\begin{tikzcd}
\mathcal{H} \otimes M \otimes \mathcal{H} & \mathcal{H} \otimes M \arrow[l, "I\otimes\Psi"'] \\
M \otimes \mathcal{H} \arrow[u, "\Phi\otimes I"'] & M \arrow[l, "\Psi"'] \arrow[u, "\Phi"']
\end{tikzcd}$$

The space C^n of *n-cochains on M* consists of the linear maps $L : M \to \mathcal{H}^{\otimes n}$. The coboundary operator $\delta : C^n \to C^{n+1}$ is defined by

$$\delta L := (I \otimes L) \circ \Phi + \sum_{i=1}^{n} (-1)^i \Delta_i \circ L + (-1)^{n+1} (L \otimes I) \circ \Psi. \tag{28}$$

Here $\Delta_i : \mathcal{H}^{\otimes n} \to \mathcal{H}^{\otimes n+1}$ stands for $I^{i-1} \otimes \Delta \otimes I^{n-i}$, i.e. the coproduct applied to the i-th factor in $\mathcal{H}^{\otimes n}$. The reader is invited to check that δ^2 vanishes, thus yielding a cohomology $HH^\bullet(A, M)$.

Now suppose that $\mathcal{H}$ is a connected filtered bialgebra with unit and co-unit. We consider (as in [BK] from which we borrow this paragraph) the particular case where $M = \mathcal{H}$, $\Phi = \Delta$ and $\Psi = (I \otimes u \circ \varepsilon) \circ \Delta$ (in other words, $\Psi(x) = x \otimes \mathbf{1}$). The coboundary operator then looks like:

$$\delta L := (I \otimes L) \circ \Delta + \sum_{i=1}^{n} (-1)^i \Delta_i \circ L + (-1)^{n+1} (L \otimes \mathbf{1}), \tag{29}$$

with $(L \otimes \mathbf{1})(x) := L(x) \otimes \mathbf{1}$. The Hochschild cohomology in this case is denoted by $HH_\varepsilon^\bullet(\mathcal{H})$. A 1-cochain $L : \mathcal{H} \to \mathcal{H}$ is a cocycle (i.e. verifies $\delta L = 0$) if and only if:

$$\Delta \circ L = (I \otimes L) \circ \Delta + L \otimes \mathbf{1}. \tag{30}$$

A 1-coboundary is given (using Sweedler's notation) by $x \mapsto \sum_{(x)} L(x_2)x_1 - L(x)\mathbf{1}$ with $L \in M^* = C^0$. The Hochschild cohomology described here revealed in [BK] to be a crucial tool to prove the locality of the regularised Feynman rules (see Sections 7.5 and 8), as well as in the understanding of the *Dyson–Schwinger equations* [BK,K4,KY].

5. Renormalisation in connected filtered Hopf algebras

We describe in this section renormalisation à la Connes–Kreimer [K1,CK,CK1] in the abstract context of connected filtered Hopf algebras: the objects to be renormalised are characters with values in a commutative unital target algebra $\mathcal{A}$ endowed with a *renormalisation scheme*, i.e. a splitting $\mathcal{A} = \mathcal{A}_- \oplus \mathcal{A}_+$ into two subalgebras. An important example is given by the *minimal subtraction scheme* of the algebra $\mathcal{A}$ of meromorphic functions of one variable z, where $\mathcal{A}_+$ is the algebra of meromorphic functions which are holomorphic at $z = 0$, and where $\mathcal{A}_- = z^{-1}\mathbb{C}[z^{-1}]$ stands for the "polar parts". Any $\mathcal{A}$-valued character φ admits a unique *Birkhoff decomposition*

$$\varphi = \varphi_-^{*-1} * \varphi_+,$$

where φ_+ is an $\mathcal{A}_+$-valued character, and where $\varphi(\operatorname{Ker} \varepsilon) \subset \mathcal{A}_-$. In the MS scheme case described just above, the renormalised character is the scalar-valued character given by the evaluation of φ_+ at $z = 0$ (whereas the evaluation of φ at $z = 0$ does not necessarily make sense).

Keeping the MS scheme and supposing that the Hopf algebra $\mathcal{H}$ is *graded*, we can introduce the notion of *locality* in this purely algebraic framework, and then define the renormalisation group and the beta function of a local character, along the lines of [CK2, CM2].

5.1. *The Birkhoff decomposition*

We consider here the situation where the algebra $\mathcal{A}$ admits a *renormalisation scheme*, i.e. a splitting into two subalgebras:

$$\mathcal{A} = \mathcal{A}_- \oplus \mathcal{A}_+$$

with $\mathbf{1} \in \mathcal{A}_+$. Let π be the projection from $\mathcal{A}$ onto $\mathcal{A}_-$ parallel to $\mathcal{A}_+$.

THEOREM 5.

(i) *Let $\mathcal{H}$ be a connected filtered Hopf algebra. Let $G(\mathcal{A})$ be the group of those $\varphi \in \mathcal{L}(\mathcal{H}, \mathcal{A})$ such that $\varphi(\mathbf{1}) = \mathbf{1}_{\mathcal{A}}$ endowed with the convolution product. Any $\varphi \in G(\mathcal{A})$ admits a unique Birkhoff decomposition*:

$$\varphi = \varphi_-^{*-1} * \varphi_+, \tag{31}$$

where φ_- sends $\mathbf{1}$ to $\mathbf{1}_{\mathcal{A}}$ and $\operatorname{Ker}\varepsilon$ *into $\mathcal{A}_-$, and where φ_+ sends $\mathcal{H}$ into $\mathcal{A}_+$. The maps φ_- and φ_+ are given on* $\operatorname{Ker}\varepsilon$ *by the following recursive formulas*:

$$\varphi_-(x) = -\pi\Biggl(\varphi(x) + \sum_{(x)} \varphi_-(x')\varphi(x'')\Biggr), \tag{32}$$

$$\varphi_+(x) = (I - \pi)\Biggl(\varphi(x) + \sum_{(x)} \varphi_-(x')\varphi(x'')\Biggr). \tag{33}$$

(ii) *If $\tau \in G(\mathcal{A})$ is a cocycle, the components τ_- and τ_+ occurring in the Birkhoff decomposition of τ are cocycles as well.*

(iii) *If the algebra $\mathcal{A}$ is commutative and if φ is a character, the components φ_- and φ_+ occurring in the Birkhoff decomposition of φ are characters as well.*

PROOF. The proof goes along the same lines as the proof of [CK1, Theorem 4]: for the first assertion it is immediate from the definition of π that φ_- sends $\operatorname{Ker}\varepsilon$ into $\mathcal{A}_-$, and that φ_+ sends $\operatorname{Ker}\varepsilon$ into $\mathcal{A}_+$. It only remains to check equality $\varphi_+ = \varphi_- * \varphi$, which is an easy computation:

$$\begin{aligned}\varphi_+(x) &= (I - \pi)\Biggl(\varphi(x) + \sum_{(x)} \varphi_-(x')\varphi(x'')\Biggr)\\ &= \varphi(x) + \varphi_-(x) + \sum_{(x)} \varphi_-(x')\varphi(x'')\\ &= (\varphi_- * \varphi)(x).\end{aligned}$$

To prove the second assertion it is sufficient to prove that τ_- is a cocycle whenever τ is a cocycle. The same property for τ_+ comes then from Proposition 19. We prove the formula $\tau_-(xy) = \tau_-(yx)$ by induction on the integer $d = |x| + |y|$: it is true for $d \leqslant 1$. Suppose the formula is true up to $d - 1$ and take any $x, y \in \mathcal{H}$ with $|x| + |y| = d$. Decompose $\Delta(xy)$ with the second version of Sweedler's notation:

$$\Delta(xy) = xy \otimes \mathbf{1} + \mathbf{1} \otimes xy + x \otimes y + y \otimes x$$

$$\begin{aligned}&+\sum_{(x)}(x'y\otimes x''+x'\otimes x''y)+\sum_{(y)}(xy'\otimes y''+y'\otimes xy'')\\&+\sum_{(x)(y)}x'y'\otimes x''y''.\end{aligned}$$

We have then:

$$\begin{aligned}\tau_-(xy)=-\pi\Bigg(&\tau(xy)+\tau_-(x)\tau(y)+\tau_-(y)\tau(x)\\&+\sum_{(x)}\big(\tau_-(x'y)\tau(x'')+\tau_-(x')\tau(x''y)\big)\\&+\sum_{(y)}\big(\tau_-(xy')\tau(y'')+\tau_-(y')\tau(xy'')\big)\\&+\sum_{(x)(y)}\tau_-(x'y')\tau(x''y'')\Bigg),\end{aligned}$$

whereas:

$$\begin{aligned}\tau_-(yx)=-\pi\Bigg(&\tau(yx)+\tau_-(y)\tau(x)+\tau_-(x)\tau(y)\\&+\sum_{(y)}\big(\tau_-(y'x)\tau(y'')+\tau_-(y')\tau(y''x)\big)\\&+\sum_{(x)}\big(\tau_-(yx')\tau(x'')+\tau_-(x')\tau(yx'')\big)+\sum_{(x)(y)}\tau_-(y'x')\tau(y''x'')\Bigg).\end{aligned}$$

Using the cocycle property for τ and the induction hypothesis we see that the two expressions are the same.

The proof of assertion (iii) goes exactly as in [CK1] and relies on the following *Rota–Baxter relation* in $\mathcal{A}$:

$$\pi(a)\pi(b)=\pi\big(\pi(a)b+a\pi(b)\big)-\pi(ab),\tag{34}$$

which is easily verified by decomposing a and b into their $\mathcal{A}_\pm$-parts. Let φ be a character of $\mathcal{H}$ with values in $\mathcal{A}$. Suppose that we have $\varphi_-(xy)=\varphi_-(x)\varphi_-(y)$ for any $x,y\in\mathcal{H}$ such that $|x|+|y|\leqslant d-1$, and compute for x,y such that $|x|+|y|=d$:

$$\varphi_-(x)\varphi_-(y)=\pi(X)\pi(Y),$$

with $X=\varphi(x)-\sum_{(x)}\varphi_-(x')\varphi(x'')$ and $Y=\varphi(y)-\sum_{(y)}\varphi_-(y')\varphi(y'')$. Using the formula

$$\pi(X)=-\varphi_-(x),$$

we get

$$\varphi_-(x)\varphi_-(y)=-\pi\big(XY+\varphi_-(x)Y+X\varphi_-(y)\big),$$

hence:

$$\begin{aligned}\varphi_-(x)\varphi_-(y) = -\pi\Big(&\varphi(x)\varphi(y) + \varphi_-(x)\varphi(y) + \varphi(x)\varphi_-(y)\\ &+ \sum_{(x)} \varphi_-(x')\varphi(x'')\big(\varphi(y) + \varphi_-(y)\big)\\ &+ \sum_{(y)} \big(\varphi(x) + \varphi_-(x)\big)\varphi_-(y')\varphi(y'')\\ &+ \sum_{(x)(y)} \varphi_-(x')\varphi(x'')\varphi_-(y')\varphi(y'')\Big).\end{aligned}$$

We have to compare this expression with

$$\begin{aligned}\varphi_-(xy) = -\pi\Big(&\varphi(xy) + \varphi_-(x)\varphi(y) + \varphi_-(y)\varphi(x)\\ &+ \sum_{(x)} \big(\varphi_-(x'y)\varphi(x'') + \varphi_-(x')\varphi(x''y)\big)\\ &+ \sum_{(y)} \big(\varphi_-(xy')\varphi(y'') + \varphi_-(y')\varphi(xy'')\big)\\ &+ \sum_{(x)(y)} \varphi_-(x'y')\varphi(x''y'')\Big).\end{aligned}$$

These two expressions are easily seen to be equal using the commutativity of the algebra $\mathcal{A}$, the character property for φ and the induction hypothesis. □

Remark 6. Define the *Bogoliubov preparation map* as the map $b: G(\mathcal{A}) \to \mathcal{L}(\mathcal{H}, \mathcal{A})$ recursively given by

$$b(\varphi)(x) = \varphi(x) + \sum_{(x)} \varphi_-(x')\varphi(x''). \tag{35}$$

Then the components of φ in the Birkhoff decomposition read:

$$\varphi_- = e - \pi \circ b(\varphi), \qquad \varphi_+ = e + (I - \pi) \circ b(\varphi). \tag{36}$$

The Bogoliubov preparation map can also be written in a more concise form:

$$b(\varphi) = \varphi_- * (\varphi - e). \tag{37}$$

Plugging Eq. (37) into (36) and setting $\alpha := \varphi - e$ we get the following expression for φ_-:

$$\varphi_- = e - P(\varphi_- * \alpha) \tag{38}$$

$$\begin{aligned}&= e - P(\alpha) + P\big(P(\alpha) * \alpha\big) + \cdots\\ &\quad + (-1)^n \underbrace{P\big(P(\ldots P(}_{n \text{ times}}\alpha) * \alpha\big) \cdots * \alpha\big) + \cdots,\end{aligned} \tag{39}$$

where $P : \mathcal{L}(\mathcal{H}, \mathcal{A}) \to \mathcal{L}(\mathcal{H}, \mathcal{A})$ is the projection defined by $P(\alpha) = \pi \circ \alpha$.

5.2. *The BCH approach to Birkhoff decomposition*

Let $\mathcal{L}$ be any complete filtered Lie algebra. Thus $\mathcal{L}$ has a decreasing filtration $(\mathcal{L}_n)$ of Lie subalgebras such that $[\mathcal{L}_m, \mathcal{L}_n] \subseteq \mathcal{L}_{m+n}$ and $\mathcal{L} \cong \lim_{\leftarrow} \mathcal{L}/\mathcal{L}_n$ (i.e., $\mathcal{L}$ is complete with respect to the topology induced by the filtration). Let A be the completion of the enveloping algebra $\mathcal{U}(\mathcal{L})$ for the decreasing filtration naturally coming from that of $\mathcal{L}$. The functions

$$\exp : A_1 \to 1 + A_1, \quad \exp(a) = \sum_{n=0}^{\infty} \frac{a^n}{n!}, \tag{40}$$

$$\log : 1 + A_1 \to A_1, \quad \log(1 + a) = -\sum_{n=1}^{\infty} \frac{(-a)^n}{n} \tag{41}$$

are well defined and are the inverse of each other. The Baker–Campbell–Hausdorff formula says that for any $x, y \in \mathcal{L}_1$ [R,V]:

$$\exp(x)\exp(y) = \exp\bigl(C(x, y)\bigr) = \exp\bigl(x + y + \mathrm{BCH}(x, y)\bigr), \tag{42}$$

where $\mathrm{BCH}(x, y)$ is an element of $\mathcal{L}_2$ given by a Lie series of which the first few terms are

$$\begin{aligned}\mathrm{BCH}(x, y) = \frac{1}{2}[x, y] + \frac{1}{12}\bigl[x, [x, y]\bigr] + \frac{1}{12}\bigl[y, [y, x]\bigr] \\ - \frac{1}{24}\bigl[x, \bigl[y, [x, y]\bigr]\bigr] + \cdots .\end{aligned} \tag{43}$$

Now let $P : \mathcal{L} \to \mathcal{L}$ be any linear map preserving the filtration of $\mathcal{L}$. We define $\tilde{P}$ to be $\mathrm{Id}_{\mathcal{L}} - P$. For $a \in \mathcal{L}_1$, define $\chi(a) = \lim_{n\to\infty} \chi_{(n)}(a)$ where $\chi_{(n)}(a)$ is given by the *BCH-recursion*

$$\begin{aligned}\chi_{(0)}(a) &:= a, \\ \chi_{(n+1)}(a) &= a - \mathrm{BCH}\bigl(P\bigl(\chi_{(n)}(a)\bigr), \tilde{P}\bigl(\chi_{(n)}(a)\bigr)\bigr),\end{aligned} \tag{44}$$

and where the limit is taken with respect to the topology given by the filtration. Then the map $\chi : \mathcal{L}_1 \to \mathcal{L}_1$ satisfies

$$\chi(a) = a - \mathrm{BCH}\bigl(P\bigl(\chi(a)\bigr), \tilde{P}\bigl(\chi(a)\bigr)\bigr). \tag{45}$$

This map appeared in [EGK2,EGK3], where also more details can be found. The following proposition [EGM,M2] gives further properties of the map χ.

PROPOSITION 23. *For any linear map $P : \mathcal{L} \to \mathcal{L}$ preserving the filtration of $\mathcal{L}$ there exists a (usually non-linear) unique map $\chi : \mathcal{L}_1 \to \mathcal{L}_1$ such that $(\chi - \mathrm{Id}_{\mathcal{L}})(\mathcal{L}_i) \subset \mathcal{L}_{2i}$ for any $i \geqslant 1$, and such that, with $\tilde{P} := \mathrm{Id}_{\mathcal{L}} - P$ we have*

$$\forall a \in \mathcal{L}_1, \quad a = C\bigl(P\bigl(\chi(a)\bigr), \tilde{P}\bigl(\chi(a)\bigr)\bigr). \tag{46}$$

This map is bijective, and its inverse is given by

$$\chi^{-1}(a) = C\bigl(P(a), \tilde{P}(a)\bigr) = a + \mathrm{BCH}\bigl(P(a), \tilde{P}(a)\bigr). \tag{47}$$

PROOF. Equation (46) can be rewritten as

$$\chi(a) = F_a\bigl(\chi(a)\bigr),$$

with $F_a : \mathcal{L}_1 \to \mathcal{L}_1$ defined by

$$F_a(b) = a - \mathrm{BCH}\bigl(P(b), \tilde{P}(b)\bigr).$$

This map F_a is a contraction with respect to the metric associated with the filtration: indeed if $b, \varepsilon \in \mathcal{L}_1$ with $\varepsilon \in \mathcal{L}_n$, we have

$$F_a(b+\varepsilon) - F_a(b) = \mathrm{BCH}\bigl(P(b), \tilde{P}(b)\bigr) - \mathrm{BCH}\bigl(P(b+\varepsilon), \tilde{P}(b+\varepsilon)\bigr).$$

The right-hand side is a sum of iterated commutators in each of which ε does appear at least once. So it belongs to $\mathcal{L}_{n+1}$. So the sequence $F_a^n(b)$ converges in $\mathcal{L}_1$ to a unique fixed point $\chi(a)$ for F_a.

Let us remark that for any $a \in \mathcal{L}_i$, then, by a straightforward induction argument, $\chi_{(n)}(a) \in \mathcal{L}_i$ for any n, so $\chi(a) \in \mathcal{L}_i$ by taking the limit. Then $\chi(a) - a = \mathrm{BCH}(P(\chi(a)), \tilde{P}(\chi(a)))$ clearly belongs to $\mathcal{L}_{2i}$. Now consider the map $\psi : \mathcal{L}_1 \to \mathcal{L}_1$ defined by $\psi(a) = C(P(a), \tilde{P}(a))$. It is clear from the definition of χ that $\psi \circ \chi = \mathrm{Id}_{\mathcal{L}_1}$. Then χ is injective and ψ is surjective. The injectivity of ψ will be an immediate consequence of the following lemma:

LEMMA 7. *The map ψ increases the ultrametric distance given by the filtration.*

PROOF. For any $x, y \in \mathcal{L}_1$ the distance $d(x, y)$ is given by 2^{-n} where $n = \sup\{k \in \mathbb{N}, x - y \in \mathcal{L}_k\}$. We have then to prove that $\psi(x) - \psi(y) \notin \mathcal{L}_{n+1}$. But

$$\begin{aligned}\psi(x) - \psi(y) &= x - y + \mathrm{BCH}\bigl(P(x), \tilde{P}(x)\bigr) - \mathrm{BCH}\bigl(P(y), \tilde{P}(y)\bigr)\\ &= x - y + \bigl(\mathrm{BCH}\bigl(P(x), \tilde{P}(x)\bigr)\\ &\quad - \mathrm{BCH}\bigl(P(x) - P(x-y), \tilde{P}(x) - \tilde{P}(x-y)\bigr)\bigr).\end{aligned}$$

The rightmost term inside the large brackets clearly belongs to $\mathcal{L}_{n+1}$. As $x - y \notin \mathcal{L}_{n+1}$ by hypothesis, this proves the claim. □

The map ψ is then a bijection, so χ is also bijective, which proves Proposition 23. □

COROLLARY 6. *For any $a \in \mathcal{L}_1$ we have the following equality taking place in $1 + A_1 \subset A$:*

$$\exp(a) = \exp\bigl(P\bigl(\chi(a)\bigr)\bigr)\exp\bigl(\tilde{P}\bigl(\chi(a)\bigr)\bigr). \tag{48}$$

Suppose now that $\mathcal{L} = \mathcal{L}(\mathcal{H}, \mathcal{A})$ (with the setup and notations of Section 5.1), and that the operator P is now the projection defined by $P(a) = \pi \circ a$. It is then clear that the first factor in the right-hand side of (48) is an element of G which sends $\mathrm{Ker}\,\varepsilon$ into $\mathcal{A}_-$, and that the second factor is an element of G which sends $\mathcal{H}$ into $\mathcal{A}_+$. By uniqueness of the Birkhoff decomposition we see then that (48) is the Birkhoff decomposition of $\exp(a)$. Starting with $a \in \mathfrak{g}_1(\mathcal{A})$ (resp. $a \in \mathfrak{g}_2(\mathcal{A})$) gives the Birkhoff decomposition of $\exp(a)$

in the group $G_1(\mathcal{A})$ of $\mathcal{A}$-valued characters of $\mathcal{H}$ (resp. in the group $G_2(\mathcal{A})$ of $\mathcal{A}$-valued inversible cocycles of $\mathcal{H}$).

Putting (39) and (48) together we get for any $\alpha \in \mathcal{L}_1$ the following *non-commutative Spitzer identity*:

$$e - P(\alpha) + P\bigl(P(\alpha) * \alpha\bigr) + \cdots + (-1)^n \underbrace{P\bigl(P\bigl(\ldots P(}_{n \text{ times}}\alpha) * \alpha\bigr) \cdots * \alpha\bigr) + \cdots$$
$$= \exp\bigl[-P\bigl(\chi\bigl(\log(e + \alpha)\bigr)\bigr)\bigr]. \tag{49}$$

This identity is valid for any filtration-preserving Rota–Baxter operator P in a complete filtered Lie algebra. For a detailed treatment of these aspects, see [EGK2,EGK3,EGM, EMP1].

5.3. *An application to number theory: renormalised multiple zeta values*

Multiple zeta values [ENR,Ho2]:

$$\zeta(s_1, \ldots, s_k) = \sum_{n_1 > \cdots > n_k \geqslant 1} n_1^{-s_1} \cdots n_k^{-s_k} \tag{50}$$

are well defined for $s_1 > 1$ and $s_j \geqslant 1$, $j \geqslant 2$. They satisfy three families of relations [ENR], among them *quasi-shuffle relations*, a typical example of which is given by

$$\zeta(s_1, s_2) + \zeta(s_2, s_1) + \zeta(s_1 + s_2) = \zeta(s_1)\zeta(s_2). \tag{51}$$

The functions $\sigma_s : t \mapsto t^{-s}$ are *classical symbols*, for which regularising iterated sums makes sense; more precisely the expression

$$\zeta(s_1 - z, \ldots, s_k - z) \tag{52}$$

can be extended to be a meromorphic function of $z \in \mathbb{C}$. Now let V be the commutative algebra generated by the functions $\sigma_s : t \mapsto t^{-s}, s \in \mathbb{C}$ on $[1, +\infty[$. It gives rise to the *Hoffman quasi-shuffle Hopf algebra* [Ho1], defined by $\mathcal{H} = \mathbb{C} \oplus V \oplus V^{\otimes 2} \oplus \cdots$. The coproduct is given by the deconcatenation:

$$\varDelta(x_1 \otimes \cdots \otimes x_n) = \sum_{k=0}^{n} (x_1 \otimes \cdots \otimes x_k) \otimes (x_{k+1} \otimes \cdots \otimes x_n). \tag{53}$$

The product $*$ is recursively defined by

$$(x \otimes u) * (y \otimes v) = x \otimes \bigl(u * (y \otimes v)\bigr) + y \otimes \bigl((x \otimes u) * v\bigr)$$
$$+ (xy) \otimes (u * v). \tag{54}$$

for $x, y \in V$ and $u, v \in \mathcal{H}$. This defines a connected filtered Hopf algebra structure on $\mathcal{H}$, with $\mathcal{H}^n = \mathbb{C} \oplus V \oplus \cdots \oplus V^{\otimes n}$. One can show with some care that the multiple zeta functions give rise to a character $\varPhi$ of $\mathcal{H}$ with values into meromorphic functions such that we have:

$$\varPhi(\sigma_{s_1} \otimes \cdots \otimes \sigma_{s_k})(z)_{|z=0} = \zeta(s_1, \ldots, s_k) \tag{55}$$

whenever $\operatorname{Re} s_1 > 1$ and $\operatorname{Re} s_j \geqslant 1$, $j \geqslant 2$. We can then define the renormalised zeta value at any $s_j \in \mathbb{Z}$ by

$$\zeta^{\mathcal{R}}(s_1, \dots, s_k) := \Phi_+(\sigma_{s_1} \otimes \cdots \otimes \sigma_{s_k})(z)_{|z=0}. \tag{56}$$

It can be moreover shown that these values are rational at negative arguments. See [MP1, MP2], and also [GZ] for a somewhat different approach.

6. Connected graded Hopf algebras

6.1. *The grading biderivation*

Let $\mathcal{H} = \bigoplus_{n \geqslant 0} \mathcal{H}_n$ be a connected graded Hopf algebra, and let $\mathcal{A}$ be a commutative unital algebra. The grading induces a biderivation Y defined on homogeneous elements by

$$\begin{aligned} Y : \mathcal{H}_n &\to \mathcal{H}_n \\ x &\mapsto nx. \end{aligned}$$

Exponentiating we get a one-parameter group θ_t of automorphisms of the Hopf algebra $\mathcal{H}$, defined on $\mathcal{H}_n$ by

$$\theta_t(x) = e^{nt}x.$$

The following result is immediate.

LEMMA 8. *$\varphi \mapsto \varphi \circ Y$ is a derivation of $(\mathcal{L}(\mathcal{H}, \mathcal{A}), *)$, and $\varphi \mapsto \varphi \circ \theta_t$ is an automorphism of $(\mathcal{L}(\mathcal{H}, \mathcal{A}), *)$ for any complex t.*

Using the fact that $e \circ Y = 0$ we easily compute for any $\mathcal{A}$-valued infinitesimal character α:

$$\begin{aligned} (\alpha \circ Y)(xy) &= \alpha\big(Y(x).y + x.Y(y)\big) \\ &= (\alpha \circ Y)(x)e(y) + (e \circ Y)(x)\alpha(y) \\ &\quad + \alpha(x)(e \circ Y)(y) + e(x)(\alpha \circ Y)(y) \\ &= (\alpha \circ Y)(x)e(y) + e(x)(\alpha \circ Y)(y). \end{aligned}$$

So we have proved:

LEMMA 9. *The map $\alpha \mapsto \alpha \circ Y$ is a linear automorphism of the space of infinitesimal characters of $\mathcal{H}$ with values in $\mathcal{A}$. Its inverse is given by $\alpha \mapsto \alpha \circ Y^{-1}$, where $Y^{-1}(x) = |x|^{-1}x$ for x homogeneous of positive degree, and $Y^{-1}(\mathbf{1}) = 0$.*

REMARK 7. The notation Y^{-1} is of course slightly incorrect, as the inverse of Y does not make sense on $\mathcal{H}_0$. The convention $Y^{-1}(\mathbf{1}) = 0$ is arbitrary: any other value of $Y^{-1}(\mathbf{1})$ would give the same result, as infinitesimal characters vanish at $\mathbf{1}$.

6.2. *The Dynkin operator*

Let $\mathcal{H}$ be a *commutative* connected graded Hopf algebra. The *Dynkin operator* is the linear endomorphism D of $\mathcal{H}$ defined by

$$D = S * Y. \tag{57}$$

The grading Y is a derivation and, due to commutativity, the antipode S is an algebra morphism. The Dynkin operator D is then an infinitesimal character with values in $\mathcal{H}$ itself [EGP]. The following theorem is due to K. Ebrahimi-Fard, J. Gracia-Bondìa and F. Patras (see [EGP] for a proof).

THEOREM 6. *Let $\mathcal{A}$ be any commutative unital algebra. Right composition with D is a bijective map from the group $G_1(\mathcal{A})$ of $\mathcal{A}$-valued characters onto the Lie algebra $\mathfrak{g}_1(\mathcal{A})$ of $\mathcal{A}$-valued infinitesimal characters. The inverse map Γ is given by*

$$\Gamma(\alpha) = \sum_n \sum_{k_1,\ldots,k_n \in \mathbb{N}^*} \frac{\alpha_{k_1} * \cdots * \alpha_{k_n}}{k_1(k_1+k_2)\cdots(k_1+\cdots+k_n)}, \tag{58}$$

where α_k is defined by $\alpha_k(x) = \alpha(x)$ if $|x| = k$ and $\alpha_k(x) = 0$ if $|x| \neq k$.

The Dynkin operator is quasi-idempotent, namely:

$$D \circ D = D \circ Y. \tag{59}$$

The terminology comes from the fact that on the tensor Hopf algebra $T(V)$ (which is *cocommutative*) the operator $S * Y$ coincides with the classical Dynkin operator:

$$D(x_1 \otimes \cdots \otimes x_k) = [\ldots[x_1, x_2], \ldots, x_k] \tag{60}$$

(see [PR] for an account of the Dynkin operator in the cocommutative context).

6.3. *Examples of connected graded Hopf algebras*

6.3.1. *The Hopf algebra of positive integers* This example is a simplified version of the one given by D. Kreimer in [K2, §2.1]. Consider the algebra $\mathcal{N}$ of the multiplicative semigroup $\mathbb{N}^* = \{1, 2, 3, \ldots\}$ of positive integers. As a vector space it admits a basis $(e_n)_{n\in\mathbb{N}^*}$ with product given by $e_n.e_m = e_{nm}$ and extended by linearity. We endow $\mathcal{N}$ with a structure of commutative cocommutative connected graded Hopf algebra thanks to the decomposition of any integer into a product of prime factors; namely we set $\Delta(e_1) = e_1 \otimes e_1$, and for any prime p:

$$\Delta(e_p) = e_p \otimes e_1 + e_1 \otimes e_p,$$

and we extend Δ to an algebra isomorphism. Hence,

$$\Delta(e_{p_1\cdots p_k}) = \sum_{I \amalg J = \{1,\ldots,k\}} e_{p_I} \otimes e_{p_J},$$

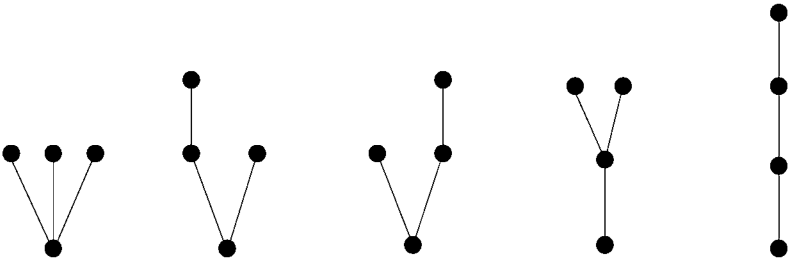


Fig. 1. The planar rooted trees with four vertices.

where p_I denotes the product of the primes p_j, $j \in I$. The grading is clearly given by the number of prime factors (including multiplicities). The antipode is given by

$$S(e_n) = (-1)^{|n|} e_n.$$

Suppose that the ground field is $k = \mathbb{C}$. The map $n \mapsto n^z$ defines a character φ of $\mathcal{N}$ with values into the holomorphic functions. Then the Riemann Zeta function is nothing but the evaluation of φ on the element

$$\omega = e_1 + e_2 + e_3 + \cdots = \prod_{p \text{ prime}} \frac{1}{1 - e_p}.$$

Here $1/(1 - e_p)$ stands for the infinite sum $e_1 + e_p + e_{p^2} + \cdots$. Of course ω is not an element of $\mathcal{N}$: it makes sense (as well as the abstract Euler product expansion on the right-hand side) only in the completion of $\mathcal{N}$ with respect to the *fine decreasing filtration* defined by the vector space grading $d(n) = n - 1$ (it is indeed an algebra filtration, as $mn - 1 \geqslant m - 1 + n - 1$). But evaluating the character φ on both sides of this equality gives the well-known Euler product expression of the zeta function.

6.3.2. *Tensor and symmetric algebras* The tensor Hopf algebra $T(V)$ of any vector space V is obviously graded. The symmetric Hopf algebra is a particular case of an enveloping Hopf algebra, with V viewed as an Abelian Lie algebra. The Hopf algebra $S(V)$ is a cocommutative commutative connected graded Hopf algebra. Note that an enveloping algebra is not graded in general, since the quotienting ideal which is generated by the $x \otimes y - y \otimes x - [x, y]$ is not homogeneous.

6.3.3. *Planar decorated rooted trees* We borrow in this section some material from [F]. A *planar rooted tree* is an oriented connected contractible graph, with a finite number of vertices, together with an embedding of it into the plane, such that only one vertex has only outgoing edges (the root). We have drawn the planar rooted trees with four vertices (see Fig. 1).

Let $\mathcal{T}$ be the set of planar rooted trees. Let V be a vector space on some field k, and let t be a planar rooted tree. The *space of decorations of t by V* is the vector space $V^{\otimes |t|}$, where $|t|$ is the number of vertices of t. A planar rooted tree t together with an indecomposable element of $V^{\otimes |t|}$ is called a *planar decorated rooted tree*. Choosing a total order on the

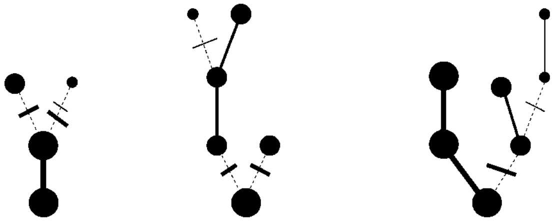


Fig. 2. An example of bi-admissible couple of cuts. From thickest to thinnest: trunk, middle and crown.

vertices amounts then to "decorate" the vertex number i with v_i. Let us consider the vector space:

$$\mathcal{T}_V = \bigoplus_{t \in \mathcal{T}} V^{\otimes |t|},$$

let $\mathcal{H}_V$ be the (noncommutative) free algebra generated by $\mathcal{T}_V$. Products of decorated trees (decorated forests) generate $\mathcal{H}_V$ as a graded vector space, the degree of a decorated forest being given by the total number of vertices. The connected graded Hopf algebra structure on $\mathcal{H}_V$ is given by the co-unit ε sending $\mathbf{1}$ to 1 and any nonempty decorated forest to 0, and by a coproduct which we describe shortly as follows (see Fig. 2).

An *elementary cut* on a tree is a cut on some edge of the given tree. An admissible cut is a cut such that any path starting from the root contains at most one elementary cut. The *empty cut* is considered as elementary, as well as the *total cut*, i.e. a cut below the root. A cut on a forest is said to be admissible if its restriction to any tree factor is admissible. Any elementary cut c sends a forest F to a couple $(P^c(F), R^c(F))$, the *crown* (or *pruning*) and the *trunk*, respectively. The trunk of a tree is a tree, but the crown of a tree is a forest. Let $\operatorname{Adm}(F)$ be the set of admissible cuts of the forest F, and let $\operatorname{Adm}^*(F)$ be the set of elementary cuts discarding the empty cut and the total cut. The coproduct:

$$\Delta(F) = \sum_{c \in \operatorname{Adm} F} P^c(F) \otimes R^c(F)$$

is graded, co-associative and compatible with the product [F]. The compatibility with the product is clear (due to the definition of an admissible cut for a forest). There is a beautiful proof of the co-associativity in [F] using induction on the degree and grafting of any forest on a decorated root. We propose here a more intuitive proof: say that a couple (c_1, c_2) of cuts is *bi-admissible* if both cuts c_1, c_2 are admissible and if c_1 never bypasses c_2, i.e. if c_2 never cuts the trunk of c_1. Any bi-admissible couple $c = (c_1, c_2)$ of cuts c defines a crown $P^c(F) = P^{c_2}(F)$, a trunk $R^c(F) = R^{c_1}(F)$, and a middle $M^c(F)$: Let $\operatorname{Adm}_2 F$ the set of bi-admissible couples of cuts of the forest F. It is quite straightforward to set down the formula for the iterated coproduct:

$$(\Delta \otimes I) \circ \Delta(F) = (I \otimes \Delta) \circ \Delta(F) = \sum_{c \in \operatorname{Adm}_2 F} P^c(F) \otimes M^c(F) \otimes R^c(F).$$

Of course the n-fold iterated coproduct admits a similar expression, involving n-admissible n-uples of admissible cuts and $n+1$ "level segments" of the forest, from the crown down to the trunk.

By Corollary 5 the connected graded bialgebra $\mathcal{H}_V$ thus obtained admits an antipode given on $\operatorname{Ker}\varepsilon$ by any of the two recursive formulas:

$$S(F) = -F - \sum_{c\in\operatorname{Adm}^*(F)} S\big(P^c(F)\big).R^c(F) \tag{61}$$

$$= -F - \sum_{c\in\operatorname{Adm}^*(F)} P^c(F).S\big(R^c(F)\big). \tag{62}$$

The square of the antipode does not in general coincide with the identity. The Hopf algebra $\mathcal{H}_V$ is *self-dual* when the vector space V is finite-dimensional. This crucial property is used in [F] for giving a complete description of the Lie algebra of the primitive elements of $\mathcal{H}_V$.

6.3.4. *Decorated rooted trees* The construction is the same except that we consider rooted trees independently from any embedding into the plane, and we consider the *free commutative* algebra generated by decorated rooted trees. We thus obtain a commutative Hopf algebra $\mathcal{H}'_V$ which is clearly a quotient of $\mathcal{H}_V$. This Hopf algebra is thoroughly investigated in [F,K1,K2].

6.3.5. *Planar binary trees [LR,F]* A planar rooted tree is *binary* if any inner vertex (i.e. any vertex different from the root and the leaves) has one ingoing edge and two outgoing edges. A *decorated binary tree* is a binary tree t together with an indecomposable element $v = v_1 \otimes \cdots \otimes v_{|t|}$ of $V^{\otimes|t|}$, where V is some vector space and $|t|$ is the grading given by the number of inner vertices of t. Choosing a total order of the inner vertices amounts then to "decorate" the inner vertex number i with v_i. Any decorated binary tree different from the vertical stick | is made of two subtrees grafted on a common root. More precisely one writes:

$$t = t^l \vee_{d_t} t^r, \tag{63}$$

where $d_t \in V$ is the decoration of the inner vertex closest to the root of t, and where t^l and t^r are the left-hand side of t and the right-hand side of t, respectively. Recursively with respect to the grading, we define the products $\prec$, $\succ$ and $*$ on the space $\mathcal{H}_V^P$ of decorated planar binary trees by:

$$\begin{aligned}
&x \prec | = | \succ x = x,\\
&x \succ | = | \prec x = 0,\\
&x \prec y = x^l \vee_{d_x} (x^r * y) \quad \text{if } x \neq |,\\
&x \succ y = (x * y^l) \vee_{d_y} y^r \quad \text{if } x \neq |,\\
&x * y = x \prec y + x \succ y.
\end{aligned}$$

Note that $| \prec |$ and $| \succ |$ are not defined. This endows $\mathcal{H}_V^P$ with a *dendriform algebra* structure augmented with a unit ([LR,F], see also [EMP2] and [EM2]). In particular the product $* = \prec + \succ$ (extended to $| * | = |$) is associative with unit $|$. The coproduct is defined (also recursively) by

$$\Delta(|) = | \otimes |,$$
$$\Delta(x) = \sum_{(x^l),(x^r)} \big((x^l)' * (x^r)'\big) \otimes \big((x^l)'' \vee_{d_x} (x^r)''\big) + x \otimes |.$$

Then $(\mathcal{H}_V^P, *, \Delta)$ is a connected graded bialgebra. The co-unit is given by $\varepsilon(|) = 1$ and $\varepsilon(t) = 0$ if $|t| \geqslant 1$. Moreover the augmentation ideal $\operatorname{Ker} \varepsilon$ is a *Hopf dendriform algebra*, in the sense that Δ is compatible with both products $\prec$ and $\succ$. This fact can be used to prove that $(\mathcal{H}_V^P, *, \Delta)$ is in fact isomorphic with the Hopf algebra $\mathcal{H}_V$ of Section 6.3.3, see [F].

6.3.6. *Some other related examples* Other combinatorial Hopf algebras which can be found in the literature [BF1,BF2,BF3,GL,M1,MR] are closely related to the Hopf algebras $\mathcal{H}_V$ of planar decorated rooted trees or to its "non-planar" quotient $\mathcal{H}'_V$: let us mention the *Grossman–Larson* Hopf algebra decorated by V [GL] which is isomorphic to the graded dual of $\mathcal{H}'_V$, the *Brouder–Frabetti* Hopf algebra [BF3] which is isomorphic to the Hopf algebra of (non-decorated) planar rooted forests of Section 6.3.3, and finally the *Malvenuto–Reutenauer* Hopf algebra $\mathcal{MR}$ [MR] which is isomorphic to a Hopf algebra $\mathcal{H}_V$ for a certain decoration space V. This last result, which uses the *bi-dendriform structure* in an essential way, implies that the Lie subalgebra of primitive elements of $\mathcal{MR}$ is free in characteristic zero. For a detailed account, see [F] and [F2].

6.3.7. *Formal diffeomorphisms and the Connes–Moscovici Hopf algebra* Let $\mathcal{A}$ be a commutative unital algebra. Let $\mathcal{D}_\mathcal{A}$ be the group (more precisely the group scheme) of "formal diffeomorphisms tangent to identity", i.e. $\mathcal{A}$-valued formal series:

$$f(x) = x + a_2x^2 + a_3x^3 + \cdots \tag{64}$$

endowed with the composition. This group scheme is given by a commutative Hopf algebra $\mathcal{H}_{\mathrm{CM}}$, i.e. $\mathcal{D}_\mathcal{A}$ is the group of $\mathcal{A}$-valued characters of the Hopf algebra $\mathcal{H}_{\mathrm{CM}}$, which is built as follows [CMo,CK2]: let $\mathcal{H}$ be the algebra with generators X, Y, δ_n, $n \geqslant 1$ and relations:

$$[Y, X] = X, \qquad [Y, \delta_n] = n\delta_n, \qquad [\delta_n, \delta_m] = 0, \qquad [X, \delta_n] = \delta_{n+1}, \tag{65}$$

endowed with the coproduct:

$$\Delta(Y) = Y \otimes 1 + 1 \otimes Y,$$
$$\Delta(X) = X \otimes 1 + 1 \otimes X + \delta_1 \otimes Y,$$
$$\Delta(\delta_1) = \delta_1 \otimes 1 + 1 \otimes \delta_1.$$

Then $\mathcal{H}_{\mathrm{CM}}$ is the (Hopf) subalgebra of $\mathcal{H}$ generated by the δ_n's. This Hopf algebra is connected, graded by $|\delta_n| = n$. The slightly different bialgebra obtained as the coordinate ring of the semigroup of formal series:

$$f(x) = a_1x + a_2x^2 + a_3x^3 + \cdots \tag{66}$$

(with a_1 not necessarily equal to $1_{\mathcal{A}}$) is known as the *Faà di Bruno bialgebra* [FdB,BFK]. In the Hopf algebra of (undecorated) planar forests, the elements

$$v_n := \sum_{\substack{t \text{ planar rooted tree} \\ \text{with } n \text{ vertices}}} t$$

freely generate (as an algebra) a Hopf subalgebra which is a noncommutative version of $\mathcal{H}_{\text{CM}}$. See [BFK] and [F]. There is Hopf algebra morphism from $\mathcal{H}_{\text{CM}}$ into the Hopf algebra of φ^3 theory described in [CK2, Section 7.6.1].

7. Hopf algebras of Feynman graphs

We treat this example (more exactly this family of examples) in a separate section for two main reasons: firstly the Hopf algebras appearing here are pointed but not connected, and secondly this is the very example where a link is established with quantum field theory. The non-connectedness is not a very serious problem: as we shall see there is a natural connected quotient. The formula for the coproduct will differ slightly from that of Connes–Kreimer in order to deal with this non-connectedness problem, but both will agree on the connected quotient. We follow [K1] quite closely, with some modifications in order to allow self-loops.

7.1. *Discarding exterior structures*

Feynman graphs are made of internal and external edges of different types, and an external edge comes with a vector attached to it (an *exterior momentum*). The sum of all exterior momenta of a given graph must be equal to zero, reflecting the global conservation of momenta in an interaction. The *Feynman rules* attach to a graph together with such an external structure an integral which can be divergent. This integral can be regularised by various procedures, among them *dimensional regularisation*: the idea is to “let the dimension of the space of momenta vary in the complex numbers”, a procedure which has been recently given a precise geometrical contents by A. Connes and M. Marcolli [CM2, §15]. The divergent integral is now replaced by a meromorphic function with poles at least at the entire dimensions where the original integral diverges [C, Chapter 4], [E].

The approach of renormalisation by A. Connes and D. Kreimer can be summarised as follows: organise Feynman graphs with their exterior structures into a graded Hopf algebra, understand the (regularised, e.g. by means of dimensional regularisation) Feynman rules as a character of this Hopf algebra with values into some algebra $\mathcal{A}$ (e.g. the meromorphic functions), choose a renormalisation scheme, i.e. a splitting $\mathcal{A} = \mathcal{A}_+ \oplus \mathcal{A}_-$ into two subalgebras, apply the method of Section 5 to extract a renormalised value, and finally recognise that this method agrees with algorithms already developed by physicists, such as the Bogoliubov–Parasiuk–Hepp–Zimmerman (BPHZ) algorithm. Our first step will consist in constructing a Hopf algebra from Feynman diagrams without exterior structure (i.e. with exterior momenta nullified).

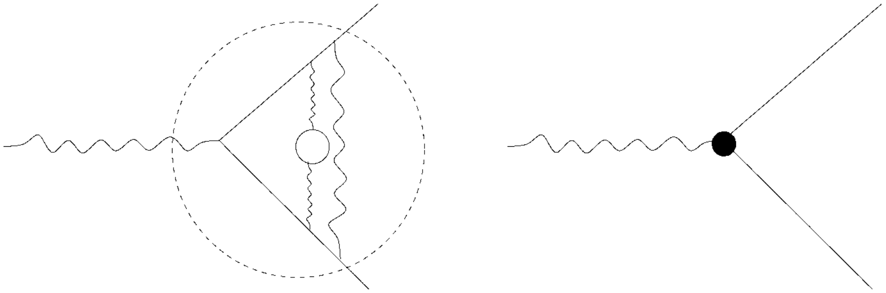


Fig. 3. A QED interaction graph and its residue.

7.2. *Operations on Feynman graphs*

A *Feynman graph* is a (non-oriented, non-planar) graph with a finite number of vertices and edges. An *internal edge* is an edge connected at both ends to a vertex (which can be the same in case of a self-loop), an *external edge* is an edge with one open end, the other end being connected to a vertex. A Feynman graph is called by physicists *vacuum graph, tadpole graph, self-energy graph*, resp. *interaction graph* if its number of external edges is 0, 1, 2, resp. > 2.

The edges (internal or external) will be of different types labelled by a positive integer $(1, 2, 3, \ldots)$, each type being represented by the way the corresponding edge is drawn (full, dashed, wavy, various colours, etc.). Let $\tau(e) \in \mathbb{N}^*$ be the type of the edge e. For any vertex v let $\mathrm{st}(v)$ be the *star* of v, i.e. the set of all edges attached to v, *with self-loops counted twice*. Hence the valence of the vertex is given by the cardinal of $\mathrm{st}(v)$. Finally to each vertex we associate its *type*, the sequence $(n_1, \ldots, n_r)$ of positive integers where n_j stands for the number of edges of type j in $\mathrm{st}(v)$. Let $T(v)$ be the type of the vertex v.

For example, in φ^n *theory* there is only one type of edge, and two types of vertices: the bivalent vertices and the n-valent vertices. In quantum electrodynamics there are two types of edges: the fermion edges (usually drawn full), and the boson edges (usually drawn wavy), and three types of vertices: bivalent boson–boson vertices, bivalent fermion–fermion vertices, and trivalent vertices with two fermion edges and one boson edge. Most of the pictures will be drawn in φ^3 or φ^4 theory, or in quantum electrodynamics.

A *one-particle irreducible graph* (in short, 1PI graph) is a connected graph which remains connected when we cut any internal edge. A disconnected graph is said to be *locally* 1PI if any of its connected components is 1PI. The *residue* of a connected 1PI graph is the graph with only one vertex obtained by shrinking all internal edges to a point. Of course any connected 1PI graph has the same type as its residue (see Fig. 3).

A *subgraph* of a Feynman graph is either the empty graph, or a *nonempty* (connected or disconnected) set of internal edges together with the vertices they encounter and the stars of those vertices, these data forming altogether a locally 1PI graph. A *proper subgraph* of Γ is a subgraph different from the empty graph or the whole graph Γ itself. If γ is a subgraph inside a graph Γ, the *contracted graph* Γ/γ is the graph obtained by replacing all connected components of γ by their residues inside Γ. As an example the residue of a graph Γ is equal to Γ/Γ (see Fig. 4).

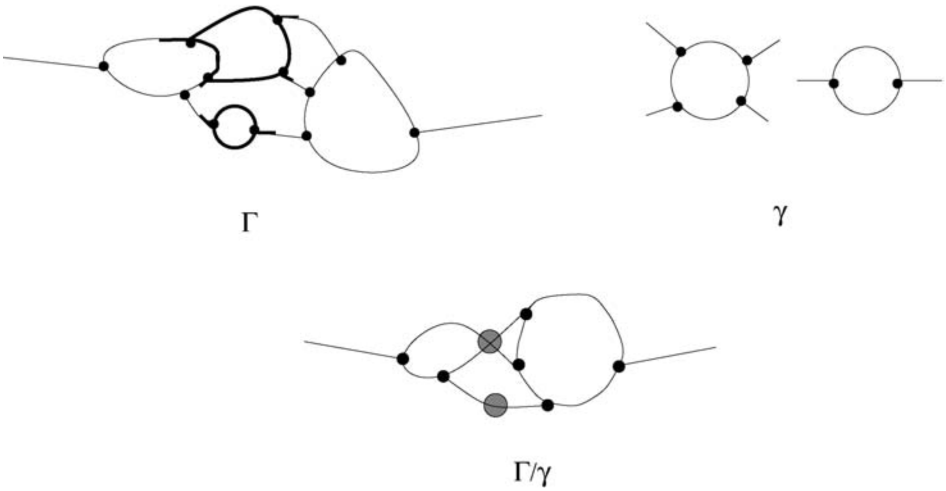


Fig. 4. A subgraph γ inside a graph Γ in φ^3 theory. The contracted graph Γ/γ does not belong to φ^3.

7.3. *The graded Hopf algebra structure*

Fix a set $\mathcal{T} = \{T_1, \dots, T_k\}$ of finite sequences of positive integers, which will be the possible vertex types we want to deal with. Let $V_{\mathcal{T}}$ be the vector space generated by all connected 1PI Feynman graphs with vertex types in $\mathcal{T}$, and all residues of those. Let $\mathcal{B}_{\mathcal{T}} = S(V_{\mathcal{T}})$ be the free commutative algebra generated by V. We shall identify the unit $\mathbf{1}$ with the empty graph and any element of $\mathcal{B}_{\mathcal{T}}$ with a linear combination of disconnected locally 1PI graphs. The algebra structure is obvious, the co-unit is given by $\varepsilon(\mathbf{1}) = 1$ and $\varepsilon(\Gamma) = 0$ for any nonempty graph Γ. The grading (at least, one possible grading) is given on connected graphs by the *loop number*

$$L := I - V + 1,$$

where I is the number of internal edges and V is the number of vertices of a given graph. This grading is extended to non-connected graphs in such a way that it is compatible with the algebra structure. It is important to notice that any nonempty subgraph has a non-vanishing loop number. The coproduct is given on connected 1PI graphs by the following formula:

$$\begin{aligned}\Delta(\Gamma) &= \sum_{\substack{\gamma \text{ subgraph of } \Gamma \\ \Gamma/\gamma \in V_{\mathcal{T}}}} \gamma \otimes \Gamma/\gamma \\ &= \Gamma \otimes \operatorname{res} \Gamma + \mathbf{1} \otimes \Gamma + \sum_{\substack{\gamma \text{ proper subgraph of } \Gamma \\ \Gamma/\gamma \in V_{\mathcal{T}}}} \gamma \otimes \Gamma/\gamma \quad \text{if } L(\Gamma) \geqslant 1, \\ \Delta(\Gamma) &= \Gamma \otimes \Gamma \quad \text{if } L(\Gamma) = 0,\end{aligned}$$

and extended to non-connected graphs by multiplicativity. We leave it to the reader as an easy exercise to show that the coproduct respects the loop number as well. Figure 5

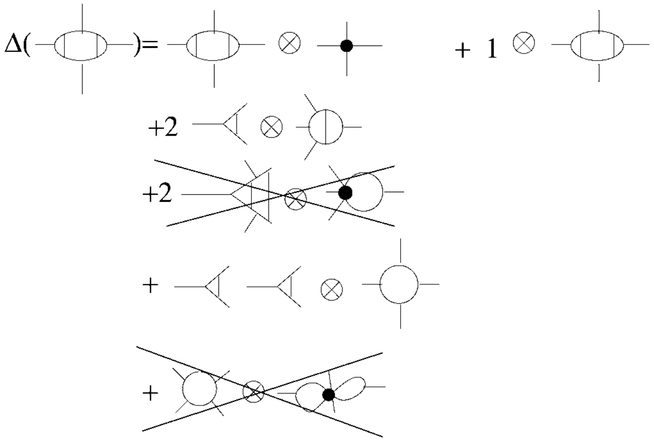


Fig. 5. An example of coproduct in φ^3 theory.

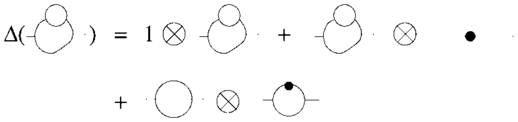


Fig. 6. Another example of coproduct in φ^3 theory.

illustrates a coproduct computation in φ^3 theory. Two terms of the sum have been removed because the corresponding contracted graphs have a vertex the type of which is outside $\mathcal{T}$ (here a pentavalent and an hexavalent vertex, respectively), and then does not belong to $V_{\mathcal{T}}$. On the other hand residues with any number of external edges are allowed.

Figure 6 illustrates another coproduct computation in φ^3 theory, with a bivalent vertex arising in the contracted graph.

Proposition 24. *$\mathcal{B}_{\mathcal{T}}$ is a pointed graded bialgebra.*

Proof. All axioms of a pointed graded bialgebra have been already given by the construction, except coassociativity of the coproduct. But we have for any 1PI graph of positive degree:

$$(\varDelta \otimes I)\varDelta(\varGamma) = \sum_{\substack{\delta \subset \gamma \subset \varGamma \\ \gamma/\delta \in \mathcal{B}_{\mathcal{T}},\ \varGamma/\gamma \in \mathcal{B}_{\mathcal{T}}}} \delta \otimes \gamma/\delta \otimes \varGamma/\gamma,$$

whereas

$$(I \otimes \varDelta)\varDelta(\varGamma) = \sum_{\substack{\delta \subset \varGamma, \widetilde{\gamma} \subset \varGamma/\delta \\ \varGamma/\delta \in \mathcal{B}_{\mathcal{T}},\ (\varGamma/\delta)/\widetilde{\gamma} \in \mathcal{B}_{\mathcal{T}}}} \delta \otimes \widetilde{\gamma} \otimes (\varGamma/\delta)/\widetilde{\gamma}.$$

There is an obvious bijection $\gamma \mapsto \widetilde{\gamma} = \gamma/\delta$ from subgraphs of Γ containing δ onto subgraphs of Γ/δ, given by shrinking δ. As we have the obvious "transitive shrinking property"

$$\Gamma/\gamma = (\Gamma/\delta)/\widetilde{\gamma},$$

the two expressions coincide. □

In order to build up a graded Hopf algebra from $\mathcal{B}_\mathcal{T}$, two choices are possible: first we can add formally the inverses of the grouplike elements, i.e. the degree zero graphs: let Σ be the set of degree zero connected 1PI graphs, let Σ^{-1} be another copy of the same set, with elements labelled γ^{-1}, $\gamma \in \Sigma$. Let $\widetilde{V}_\mathcal{T}$ be the vector space generated by $V_\mathcal{T}$ and Σ^{-1}, and consider

$$\widetilde{\mathcal{H}}_\mathcal{T} = S(\widetilde{V}_\mathcal{T})/J, \tag{67}$$

where J is the ideal generated by $\gamma\gamma^{-1} - \mathbf{1}$, $\gamma \in \Sigma$. The coproduct on $S(V_\mathcal{T})$ is extended to $S(\widetilde{V}_\mathcal{T})$ by saying that the elements of Σ^{-1} are grouplike. The ideal J is a bi-ideal, and so $\widetilde{\mathcal{H}}_\mathcal{T}$ is a pointed graded bialgebra. An antipode is easily given inductively with respect to the degree, as any degree zero element has an antipode given by $S(\gamma) = \gamma^{-1}$, $S(\gamma^{-1}) = \gamma$ for any $\gamma \in \Sigma$. The second option consists in killing the degree zero graphs (except the empty graph). We set:

$$\mathcal{H}_\mathcal{T} = \mathcal{B}_\mathcal{T}/K,$$

where K is the ideal generated by $\gamma - \mathbf{1}, \gamma \in \Sigma$. It is easily seen to be a bi-ideal. The quotient is then a connected graded bialgebra, hence a Hopf algebra thanks to Corollary 5. We can identify the quotient with $S(V'_\mathcal{T})$, where $V'_\mathcal{T}$ stands for the vector space generated by connected 1PI graphs with loop number $\geqslant 1$. The coproduct is then given by Kreimer's formula

$$\Delta(\Gamma) = \Gamma \otimes \mathbf{1} + \mathbf{1} \otimes \Gamma + \sum_{\substack{\gamma \text{ proper subgraph of } \Gamma \\ \Gamma/\gamma \in V_\mathcal{T}}} \gamma \otimes \Gamma/\gamma.$$

7.4. *External structures*

We shall be very sketchy here. Let W be a finite-dimensional vector space (the *momentum space*). Keeping the notations of the preceding subsection and following [CK2] and [K1], a *specified graph* will be a couple (Γ, σ) where Γ is a connected graph in $V_\mathcal{T}$ with E external lines, and σ is a distribution on the vector subspace $M_\Gamma = M_E \subset W^E$ defined by

$$M_E = \left\{ (p_1, \ldots, p_E) \sum_{k=1}^{E} p_k = 0 \right\}.$$

In order to get a Hopf algebra structure for specified graphs we must further discriminate the type of a vertex: once the number of edges of each type is fixed for a vertex, we add an extra nonzero natural number, so that there are "several kinds of vertices of the same type". This comes from the Lagrangian of the given quantum field theory we are dealing with: each monomial of degree n_i with respect to the field ϕ_i ($i \in \{1, \dots, k\}$) gives rise to vertices of type $T = (n_1, \dots, n_k)$, and there are as many kinds of vertices of type T as monomials of "field degree" T inside the Lagrangian. For example, in φ^3 theory with mass, the terms $(m^2/2)\varphi^2$ and $(\partial\varphi)^2/2$ give rise to two different kinds of bivalent vertices. When taking residues we must specify the kind for the unique remaining vertex: then when considering a contracted graph Γ/γ we must consider the kind of every contracted vertex (corresponding to a connected component of the subgraph γ). This gives rise to contracted graphs $\Gamma/\gamma(i)$ where i is a multi-index.

To any vertex of kind (T, i) corresponds a specific distribution $\sigma_{T,i}$ on M_Γ, where Γ is any graph whose residue gives a vertex of type T. This extends to non-connected graphs by considering multi-indices i. Now $\mathcal{T}$ stands for the set of all *kinds* (T, i) of vertices we can encounter, $V_\mathcal{T}$ stands for the space generated by all connected 1PI graphs with vertex kinds in $\mathcal{T}$, and all residues of those. Let $V'_\mathcal{T}$ be the space generated by all connected 1PI Feynman graphs with vertex kinds in $\mathcal{T}$ and nonzero loop number, let $(V'_\mathcal{T})_E$ be the subspace of $V'_\mathcal{T}$ of graphs with E external edges, and finally let $W'_\mathcal{T}$ be the corresponding space of specified graphs:

$$W'_\mathcal{T} = \sum_{E=0}^{\infty} (V'_\mathcal{T})_E \otimes \mathcal{D}'(M_E).$$

We directly give the connected version of the Hopf algebra: it is given by $\mathcal{H}_\mathcal{T} = S(W'_\mathcal{T})$, and the coproduct is given on connected specified graphs by

$$\begin{aligned}\Delta(\Gamma, \sigma) = {} & (\Gamma, \sigma) \otimes \mathbf{1} + \mathbf{1} \otimes (\Gamma, \sigma) \\ & + \sum_{\gamma \text{ proper subgraph of } \Gamma} \sum_{i, \Gamma/\gamma(i) \in V_\mathcal{T}} (\gamma, \sigma_{T,i}) \otimes (\Gamma/\gamma(i), \sigma).\end{aligned}$$

7.5. *Feynman rules*

Let $\mathcal{T}$ be a set of "kinds of vertices" defining the set of Feynman graphs of a given quantum field theory, as explained in Section 7.4 above. Each (internal or external) edge of a graph comes with its *propagator* Δ_e, which is a distribution on the *coordinate space*, i.e. on some given finite-dimensional vector space V of dimension D, which will be endowed with a *Euclidean* metric here. The propagator Δ_e is determined by the type $\tau(e)$ of the edge e. We consider the Fourier transform $\sigma_e = \mathcal{F}(\Delta_e)$, which is a function on the dual $W = V^*$, the *momentum space*. The Feynman rules associate to each 1PI graph Γ (with E external edges, I internal edges and loop number L) together with external momenta $(p_1, \dots, p_E) \in W^E$ the following integral:

$$\begin{aligned}J_\Gamma(p_1, \dots, p_E) = (2\pi)^{-LD} \frac{1}{S(\Gamma)} \int_{W_\Gamma} & I_\Gamma(p_1, \dots, p_E; k_1, \dots, k_I) \\ & \times \delta_{W_\Gamma}\, dk_1 \dots dk_I. \qquad (68)\end{aligned}$$

Here W_Γ is the affine subspace of W^I (of dimension LD) defined by the vanishing of the sum of momenta at each vertex of the graph, $\delta_{W_\Gamma} dk_1 \ldots dk_I$ is the canonical volume form on W_Γ coming from the Lebesgue measure $dk_1 \ldots dk_I$ on W^I, $S(\Gamma)$ is the symmetry factor (i.e. the order of the automorphism group of Γ), and finally the integrand I_Γ is defined by

$$\begin{aligned} &I_\Gamma(p_1, \ldots, p_E; k_1, \ldots, k_I) \\ &\quad = \prod_{v \text{ vertex}} \lambda_v \prod_{e \text{ external edge}} \sigma_e(p_e) \prod_{e \text{ internal edge}} \sigma_e(k_e). \end{aligned} \tag{69}$$

Here λ_v is an interaction term which is determined by the *kind* of the vertex v (see Section 7.4 above) and the momenta flowing into it. The λ_v's contain the coupling constants of the theory.

The integral (68) is usually divergent. It can be regularised by various techniques: let us mention *dimensional regularisation*, which replaces the ill-defined $J_\Gamma(p_1, \ldots, p_E)$ by a meromorphic function $J_\Gamma(p_1, \ldots, p_E)(z)$ of one variable z, with a possible pole at $z = D = \dim W$. See [C,HV,E], and see [CM2] for a conceptual approach to "spaces of complex dimension z". The regularised integral extends naturally by multiplicativity to a character J of the Hopf algebra $\mathcal{H}_\mathcal{T}$ with values into meromorphic functions. The *renormalised Feynman rule in the minimal subtraction scheme* is then defined by the scalar-valued character $J_+(z)_{|z=D}$, where J_+ is the second component in the Birkhoff decomposition (see Section 5.1).

7.6. *Two examples*

7.6.1. *φ^3 theory* φ^3 theory in D dimensions is given by a classical action functional:

$$S(\varphi) = \int_V \mathcal{L}(\varphi)\, d^D\varphi, \tag{70}$$

where φ is a classical field (i.e. some function on the space V), and the Lagrangian $\mathcal{L}$ is given by

$$\mathcal{L}(\varphi) = -\frac{m^2\varphi^2}{2} + \frac{(\partial\varphi)^2}{2} + \lambda\frac{\varphi^3}{3!}, \tag{71}$$

where λ is the *coupling constant* of the theory. There is only one type of edge with corresponding propagator, which in momentum space[4] is written:

$$\sigma(k) = \frac{1}{\|k\|^2 + m^2}. \tag{72}$$

The three terms in the Lagrangian give rise to three kinds of vertices, respectively two kinds of bivalent vertices $\underset{0}{-\!\!\times\!\!-}$, $\underset{1}{-\!\!\times\!\!-}$ and a trivalent vertex $-\!\!<$. The interaction terms are

[4] We write the propagator in the Euclidean setting after "Wick rotation". In Minkowski space it would be written $\sigma(k) = \frac{1}{\|k\|^2 - m^2 + i\varepsilon}$ where $\|k\|^2$ now stands for the Minkowski scalar product $-k_1^2 + k_2^2 + \cdots + k_D^2$. For $\varepsilon \to 0$ there is a pole on the "mass shell" $\|k\|^2 = m^2$.

given by

$$\begin{aligned}
\lambda(\underset{0}{-\!\!\times\!\!-}) &= m,\\
\lambda(\underset{1}{-\!\!\times\!\!-}) &= \|p\|^2,\\
\lambda(-\!\!<) &= \mu^{3-D/2}\lambda,
\end{aligned}$$

where p is the momentum flowing into the corresponding vertex of kind $\underset{1}{-\!\!\times\!\!-}$. Here μ stands for an arbitrary mass (the *'t Hooft mass*) so that the coupling constant λ remains dimensionless for any D [CK1]. This gives concretely the integral J_Γ defined in (68). The theory is super-renormalisable for $D < 6$ (i.e. the integrals J_Γ are convergent except for a finite number of graphs), renormalisable for $D = 6$ and non-renormalisable for $D > 6$ [C,CK1].

7.6.2. *Quantum electrodynamics* We sketchily follow [VS], to which we refer for the details (see also [BF1,BF2,BF3]). The dimension D is equal to 4, the coordinates of $k \in W$ are denoted by k_ν, $\nu = 1, 2, 3, 4$. The integrand I_Γ takes values into the space $M_4(\mathbb{C})$ of 4×4 matrices. The 4×4 Dirac matrices are denoted by γ^ν, $\nu = 1, 2, 3, 4$ [Hu, Chapter 6]. There are two types of lines: the electron lines — and the photon lines $\sim\!\sim$, coming with their propagators σ_{electron} and σ_{photon}. There are two kinds of vertices: a bivalent electron–electron vertex $-\!\!\times\!\!-$ and a trivalent vertex $\sim\!\sim\!<$. The symmetry factor $S(\Gamma)$ of a graph is always equal to 1 except for vacuum graphs. The interaction terms are given by

$$\begin{aligned}
\lambda(-\!\!\times\!\!-) &= m,\\
\lambda(\sim\!\sim\!<) &= e\gamma^\nu,
\end{aligned}$$

where m and e are the (non-renormalised) mass and electric charge of the electron, respectively. The ν-dependence interaction term $e\gamma^\nu$ of any trivalent vertex is removed by summing over $\nu = 1, 2, 3, 4$ in combination with the attached photon edge, as the expression of the photon propagator contains an index ν as well [VS, §3].

The Hopf algebra $\mathcal{H}_\mathcal{T}$ is constructed as in Section 7.4, except that the external structure σ of a specified graph (Γ, σ) is now an element of $\mathcal{D}'(M_\Gamma) \otimes M_4(\mathbb{C})^*$. The regularised Feynman rules then yield a character of $\mathcal{H}_\mathcal{T}$ with values into the meromorphic functions (and not into the noncommutative algebra of 4×4-matrices of meromorphic functions). W. Van Suijlekom proved in [VS] that the *Ward–Takahashi identities* generate a Hopf ideal of $\mathcal{H}_\mathcal{T}$ on which the regularised Feynman rule character vanishes. This seems to be a general pattern for gauge theories [K3,VS2].

8. The renormalisation group and the beta function

Let $\mathcal{H}$ be a connected graded Hopf algebra over the complex numbers. Let $\mathcal{A}$ be the algebra of germs of meromorphic functions at some $z_0 \in \mathbb{C}$. The algebra $\mathcal{A}$ admits a splitting into two subalgebras:

$$\mathcal{A} = \mathcal{A}_+ \oplus \mathcal{A}_-,$$

where $\mathcal{A}_+$ is the algebra of germs of holomorphic functions at $z = 0$, and $\mathcal{A}_-$ is equal to $z^{-1}\mathbb{C}[z^{-1}]$. We denote by Y (resp. θ_t) the biderivation (resp. the one-parameter group of automorphisms) of the Hopf algebra $\mathcal{H}$ induced by the graduation. We denote as before by $G(\mathcal{A})$ the group of the elements $\varphi \in \mathcal{L}(\mathcal{H}, \mathcal{A})$ such that $\varphi(\mathbf{1}) = \mathbf{1}_{\mathcal{A}}$ (with the convolution products), and by $\mathfrak{g}(\mathcal{A})$ the subalgebra of $\mathcal{L}(\mathcal{H}, \mathcal{A})$ of the elements $\varphi \in \mathcal{L}(\mathcal{H}, \mathcal{A})$ such that $\varphi(\mathbf{1}) = 0$. We will sometimes write G for $G(\mathcal{A})$ and $\mathfrak{g}$ for $\mathfrak{g}(\mathcal{A})$, dropping the target algebra $\mathcal{A}$ when no confusion is possible.

Recall that $G = \exp \mathfrak{g}$. Dropping the mention of the target algebra we shall consider as before the subgroups G_1 (resp. G_2) of G formed by the characters of $\mathcal{H}$ with values in $\mathcal{A}$ (resp. by the elements of G which enjoy the cocycle property), as well as the Lie subalgebras $\mathfrak{g}_1$ (resp. $\mathfrak{g}_2$) of derivations of $\mathcal{H}$ with values in $\mathcal{A}$ (resp. of $\mathfrak{g}$ which enjoy the cocycle property). Recall that $G_1 = \exp \mathfrak{g}_1$ and $G_2 = \exp \mathfrak{g}_2$.

8.1. *The renormalisation map*

We construct here a bijection $R : \mathfrak{g} \to \mathfrak{g}$ thanks to the biderivation Y:

PROPOSITION 25. *The equation*

$$\varphi \circ Y = \varphi * \gamma \tag{73}$$

defines a bijective correspondence

$$\begin{aligned} \widetilde{R} : G &\to \mathfrak{g} \\ \varphi &\mapsto \gamma. \end{aligned}$$

Equivalently, the equation

$$e^{*\alpha} \circ Y = e^{*\alpha} * \gamma \tag{74}$$

defines a (non-linear) bijective correspondence

$$\begin{aligned} R : \mathfrak{g} &\to \mathfrak{g} \\ \alpha &\mapsto \gamma, \end{aligned}$$

and $R = \widetilde{R} \circ \exp$.

PROOF. Equation (73) yields for any homogeneous $x \in \mathcal{H}$:

$$|x|\varphi(x) = \gamma(x) + \sum_{(x)} \varphi(x')\gamma(x''),$$

which determines γ (recursively with respect to $|x|$) from φ and vice-versa, starting from $\varphi(\mathbf{1}) = \mathbf{1}_{\mathcal{A}}$ and $\gamma(\mathbf{1}) = 0$. In other words Eq. (73) determines a bijection $\widetilde{R}$ from G to $\mathfrak{g}$ such that $\gamma = \widetilde{R}(\varphi)$. The rest of the proposition follows then immediately. □

Equation (74) yields the following explicit expression for R:

$$R(\alpha) = e^{*-\alpha} * \left(e^{*\alpha} \circ Y\right). \tag{75}$$

PROPOSITION 26. *If moreover the Hopf algebra $\mathcal{H}$ is commutative, the correspondence $\widetilde{R}$ reduces to right composition with the Dynkin operator*:

$$\widetilde{R}(\varphi) = \varphi \circ D. \tag{76}$$

PROOF. We have:

$$\varphi \circ D = \varphi \circ (S * Y) = (\varphi \circ S) * (\varphi \circ Y) = \varphi^{*-1} * (\varphi \circ Y), \tag{77}$$

hence:

$$\varphi \circ Y = \varphi * (\varphi \circ D). \tag{78}$$

So $\varphi \circ D = \widetilde{R}(\varphi)$ according to Proposition 25 (see [EGP]). □

There is another explicit formula:

PROPOSITION 27.

$$R(\alpha) = \int_0^1 e^{*-s\alpha} * (\alpha \circ Y) * e^{*s\alpha}\, ds = \frac{1 - e^{-\operatorname{ad}\alpha}}{\operatorname{ad}\alpha}.(\alpha \circ Y). \tag{79}$$

PROOF. For any $u \in \mathbb{C}$ we have

$$e^{*u\alpha} \circ Y = e^{*u\alpha} * R(u\alpha).$$

Setting $u = t + s$ and using the group property $e^{*(t+s)}\alpha = e^{*t\alpha} * e^{*s\alpha}$ as well as the derivation property

$$(e^{*t\alpha} * e^{*s\alpha}) \circ Y = (e^{*t\alpha} \circ Y) * e^{*s\alpha} + e^{*t\alpha} * (e^{*s\alpha} \circ Y),$$

we get:

$$e^{*(t+s)\alpha} \circ Y = e^{*(t+s)\alpha} * (R(s\alpha) + e^{*-s\alpha} * R(t\alpha) * e^{*s\alpha}). \tag{80}$$

Setting $\gamma(t) = R(t\alpha)$ the above equation reads:

$$\gamma(t+s) = \gamma(s) + e^{*-s\alpha} * \gamma(t) * e^{*s\alpha}. \tag{81}$$

We have $\gamma(0) = 0$, and differentiating this equation with respect to s at $s = 0$ yields:

$$\dot{\gamma}(t) = \dot{\gamma}(0) + \big[\gamma(t), \alpha\big]. \tag{82}$$

Differentiating once again with respect to t gives then

$$\ddot{\gamma}(t) = \big[\dot{\gamma}(t), \alpha\big]. \tag{83}$$

The solution of this first order differential equation is given by

$$\dot{\gamma}(t) = e^{*-t\alpha} * \dot{\gamma}(0) * e^{*t\alpha}. \tag{84}$$

Expanding the equation $e^{*t\alpha} \circ Y = e^{*t\alpha} * \gamma(t)$ up to order 1 in $t = 0$ yields immediately

$$\dot{\gamma}(0) = \alpha \circ Y. \tag{85}$$

Integrating and setting $t = 1$ then proves the proposition. □

COROLLARY 7. *The correspondence R sends infinitesimal characters to infinitesimal characters and cocycles to cocycles.*

PROOF. The first assertion follows immediately from Propositions 27, 22 and Lemma 9. The second assertion follows directly from Proposition 27. □

REMARK 8. If the Hopf algebra $\mathcal{H}$ is cocommutative, then, thanks to the commutativity of $\mathcal{A}$, the convolution product is commutative. The correspondence R becomes then linear and we simply have

$$R(\alpha) = \alpha \circ Y. \tag{86}$$

8.2. *Inverting $\widetilde{R}$: the scattering-type formula*

We shall give an explicit expression of the map $\widetilde{R}^{-1} : \mathfrak{g} \to G$. It takes the form

$$\widetilde{R}^{-1}(\gamma) = \lim_{t \to +\infty} \exp(-tA) \exp tB,$$

(cf. Theorem 7 below), where A and B live in a semi-direct product Lie algebra $\widetilde{\mathfrak{g}} = \mathfrak{g} \rtimes \mathbb{C}$. We have to describe this semi-direct product and the corresponding semi-direct product group $\widetilde{G} = G \rtimes \mathbb{C}$, and then we must endow $\widetilde{G}$ with a topology so that the above limit makes sense. We adapt here the proof of Theorem 2 in [CK2]. To be precise, we define the Lie algebra

$$\widetilde{g} := \mathfrak{g} \rtimes \mathbb{C}.Z_0, \tag{87}$$

where the action of Z_0 on $\mathfrak{g}$ is given by the derivation

$$Z_0(\gamma) = \gamma \circ Y. \tag{88}$$

The corresponding group is $\widetilde{G} = G \rtimes \mathbb{C}$, where the right action of $\mathbb{C}$ on G is given by

$$\varphi.t = \varphi \circ \theta_t, \tag{89}$$

so that the product is given by $(\varphi, t)(\psi, s) = (\varphi * (\psi \circ \theta_t), t + s)$. We shall not delve out the Lie group structure for $\widetilde{G}$ here, but we shall define the exponential map $\exp : \widetilde{\mathfrak{g}} \to \widetilde{G}$. It should of course coincide with the exponential already defined on G, and should verify

$$\exp tZ_0 = (e, t),$$

so that $\exp tZ_0$ indeed acts on G by composition with $\theta_t = \exp tY$ on the right. We should be able in principle to define $\exp(tZ_0 + \gamma)$ by means of the Baker–Campbell–Hausdorff formula as long as convergence problems can be handled here. We prefer, like in [CK2], cf. also [CM1], to give an alternative definition based on Araki's expansion formula [Ar]:

$$\begin{aligned}\exp(tZ_0 + \gamma) = \sum_{n=0}^{\infty} \int_{\sum_{j=0}^{n} u_j = 1,\, u_j \geqslant 0} & \exp(u_0 tZ_0)\gamma \exp(u_1 tZ_0)\gamma \cdots \gamma \\ & \times \exp(u_n tZ_0)\, du_1 \ldots du_n.\end{aligned} \tag{90}$$

Let us check that the sum above makes sense in our particular context: setting $v_j = u_j + u_{j+1} + \cdots + u_n$ we get

$$
\begin{aligned}
&\exp(-tZ_0)\exp(tZ_0+\gamma) = \exp(-tZ_0).\\
&\exp(tZ_0).\sum_{n=0}^{\infty}\int_{0\leqslant v_n\leqslant\cdots\leqslant v_1\leqslant 1}\exp(-tv_1Z_0)\gamma\exp(tv_1Z_0)\cdots\\
&\qquad\times\exp(-tv_nZ_0)\gamma\exp(tv_nZ_0)\,dv_1\dots dv_n\\
&\quad=\sum_{n=0}^{\infty}\int_{0\leqslant v_n\leqslant\cdots\leqslant v_1\leqslant 1}(\gamma\circ\theta_{-tv_1})*\cdots*(\gamma\circ\theta_{-tv_n})\,dv_1\dots dv_n.
\end{aligned}
$$

The sum here is well defined as a locally finite sum, as it ends at $n = n_0$ when evaluated at any $x = \mathcal{H}^{n_0}$. It remains to check that the exponential thus defined enjoys the one-parameter group property. Indeed, for any s, t real we have:

$$
\begin{aligned}
&\exp t(Z_0+\gamma)\exp s(Z_0+\gamma)\\
&\quad= e^{tZ_0}\left(\sum_{p=0}^{\infty}t^p\int_{0\leqslant v_p\leqslant\cdots\leqslant v_1\leqslant 1}(\gamma\circ\theta_{-tv_1})*\cdots*(\gamma\circ\theta_{-tv_p})\,dv_1\dots dv_p\right)\\
&\qquad\times e^{sZ_0}\left(\sum_{p=0}^{\infty}s^q\int_{0\leqslant w_q\leqslant\cdots\leqslant w_1\leqslant 1}(\gamma\circ\theta_{-sw_1})*\cdots\right.\\
&\qquad\left.*(\gamma\circ\theta_{-sw_q})\,dw_1\dots dw_q\right)\\
&\quad= e^{(t+s)Z_0}\sum_{p,q=0}^{\infty}t^ps^q\int\int_{0\leqslant v_p\leqslant\cdots\leqslant v_1\leqslant 1,\,0\leqslant w_q\leqslant\cdots\leqslant w_1\leqslant 1}(\gamma\circ\theta_{-s-tv_1})*\cdots\\
&\qquad*(\gamma\circ\theta_{-s-tv_p})*(\gamma\circ\theta_{-sw_1})*\cdots\\
&\qquad*(\gamma\circ\theta_{-sw_q})\,dv_1\dots dv_p\,dw_1\dots dw_q\\
&\quad= e^{(t+s)Z_0}\sum_{p,q=0}^{\infty}\int\int_{0\leqslant v_p\leqslant\cdots\leqslant v_1\leqslant t,\,0\leqslant w_q\leqslant\cdots\leqslant w_1\leqslant s}(\gamma\circ\theta_{-s-v_1})*\cdots\\
&\qquad*(\gamma\circ\theta_{-s-v_p})*(\gamma\circ\theta_{-w_1})*\cdots\\
&\qquad*(\gamma\circ\theta_{-w_q})\,dv_1\dots dv_p\,dw_1\dots dw_q\\
&\quad= e^{(t+s)Z_0}\sum_{n=0}^{\infty}\sum_{p+q=n}\int\int_{s\leqslant v_p\leqslant\cdots\leqslant v_1\leqslant t+s,\,0\leqslant w_q\leqslant\cdots\leqslant w_1\leqslant s}(\gamma\circ\theta_{-v_1})*\cdots\\
&\qquad*(\gamma\circ\theta_{-v_p})*(\gamma\circ\theta_{-w_1})*\cdots\\
&\qquad*(\gamma\circ\theta_{-w_q})\,dv_1\dots dv_p\,dw_1\dots dw_q\\
&\quad= e^{(t+s)Z_0}\sum_{n=0}^{\infty}\int_{s\leqslant u_p\leqslant\cdots\leqslant u_1\leqslant t+s}(\gamma\circ\theta_{-u_1})*\cdots*(\gamma\circ\theta_{-u_n})\,du_1\dots du_n\\
&\quad= \exp(t+s)(Z_0+\gamma).
\end{aligned}
$$

We can now state the main theorem of this section.

THEOREM 7. *Let $\gamma \in \mathfrak{g}$. Then*:

(i) *For any real t the product* $\exp -tZ_0 \exp t(Z_0 + \gamma)$ *belongs to G.*

(ii) *The product above admits a limit when $t \to +\infty$ for the topology on G induced by the simple convergence topology on $\mathcal{L}(\mathcal{H}, \mathcal{A})$.*

(iii) *The inverse of the renormalisation map is given by*

$$\widetilde{R}^{-1}(\gamma) = \lim_{t\to+\infty} \exp -tZ_0 \exp t(Z_0 + \gamma). \tag{91}$$

(iv) $\widetilde{R}^{-1}$ *sends $\mathfrak{g}_1$ into G_1 and $\mathfrak{g}_2$ into G_2.*

PROOF. The first assertion comes directly from the expression:

$$\begin{aligned}&\exp(-tZ_0)\exp(tZ_0 + t\gamma)\\&\quad= \sum_{n=0}^{\infty} \int_{0\leqslant v_n\leqslant\cdots\leqslant v_1\leqslant 1} (t\gamma \circ \theta_{-tv_1}) * \cdots * (t\gamma \circ \theta_{-tv_n})\, dv_1 \ldots dv_n.\end{aligned}$$

The right-hand side belongs manifestly to G. The change of variables $v_j \to tv_j$ yields:

$$\begin{aligned}\exp(-tZ_0)\exp(tZ_0 + t\gamma) = \sum_{n=0}^{\infty} \int_{0\leqslant v_n\leqslant\cdots\leqslant v_1\leqslant t} &(\gamma \circ \theta_{-v_1}) * \cdots\\ &* (\gamma \circ \theta_{-v_n})\, dv_1 \ldots dv_n.\end{aligned}$$

To prove the second assertion it suffices to prove that the integrals

$$I_n := \int_{0\leqslant v_n\leqslant\cdots\leqslant v_1\leqslant +\infty} (\gamma \circ \theta_{-v_1}) * \cdots * (\gamma \circ \theta_{-v_n})\, dv_1 \ldots dv_n \tag{92}$$

converge, as the sum $I_0 + I_1 + I_2 + \cdots$ is locally finite. The convergence is easily seen by induction on n: indeed we have $I_0 = e$ and the crucial equality valid for any $x \in \operatorname{Ker}\varepsilon$:

$$Y^{-1}(x) = \int_0^{\infty} \theta_{-t}(x)\, dt. \tag{93}$$

It follows that we have for any $a \in \mathfrak{g}$

$$\int_0^{\infty} a \circ \theta_{-t}\, dt = a \circ Y^{-1}. \tag{94}$$

A simple computation then gives

$$I_n = \int_0^{\infty} (I_{n-1} * \gamma) \circ \theta_{-v_n}\, dv_n = (I_{n-1} * \gamma) \circ Y^{-1},$$

which inductively shows the convergence of the integrals I_n. Now equation (E) can be rewritten as

$$\begin{aligned}&\varphi(x) = (\varphi * \gamma) \circ Y^{-1}(x) \quad \forall x \in \operatorname{Ker}\varepsilon,\\&\varphi(\mathbf{1}) = \mathbf{1}_{\mathcal{A}}.\end{aligned}$$

As $\gamma = \widetilde{R}(\varphi)$ this means that

$$\widetilde{R}^{-1}(\gamma) = e + T\big(\widetilde{R}^{-1}(\gamma)\big), \tag{95}$$

where T is the transformation of $\mathcal{L} = \mathcal{L}(\mathcal{H}, \mathcal{A})$ defined by

$$T(\psi) = (\psi * \gamma) \circ Y^{-1} = \int_0^\infty (\psi * \gamma) \circ \theta_{-t}\, dt.$$

The transformation T is a contraction on $\mathcal{L}$ for the distance associated with the filtration. $\widetilde{R}^{-1}(\gamma)$ is then the limit of the sequence (φ_n) defined by $\varphi_0 = e$ and $\varphi_{n+1} = e + T(\varphi_n)$. A straightforward computation yields

$$\varphi_n = \sum_{k=0}^{n} I_k. \tag{96}$$

Hence we have

$$\widetilde{R}^{-1}(\gamma) = \sum_{k=0}^{\infty} I_k, \tag{97}$$

which proves assertion (iii). Finally assertion (iv) comes from the fact that the derivation Z_0 acts on $\mathfrak{g}_1$ and $\mathfrak{g}_2$. We can then consider semi-direct products:

$$\begin{aligned} \widetilde{\mathfrak{g}}_1 &= \mathfrak{g}_1 \rtimes \mathbb{C}.Z_0, &\qquad \widetilde{G}_1 &= G_1 \rtimes \mathbb{C},\\ \widetilde{\mathfrak{g}}_2 &= \mathfrak{g}_2 \rtimes \mathbb{C}.Z_0, &\qquad \widetilde{G}_2 &= G_2 \rtimes \mathbb{C}, \end{aligned}$$

and thus replace the group G by any of the two groups G_1, G_2 in assertions (i)–(iii), which proves assertion (iv) and ends the proof of the theorem. □

COROLLARY 8. *The inverse of* $R : \mathfrak{g} \to \mathfrak{g}$ *is given by*

$$R^{-1}(\gamma) = \lim_{t \to +\infty} \operatorname{Log}\big(\exp -tZ_0 \exp t(Z_0 + \gamma)\big), \tag{98}$$

and R^{-1} *sends* $\mathfrak{g}_1$ (*resp.* $\mathfrak{g}_2$) *into* $\mathfrak{g}_1$ (*resp.* $\mathfrak{g}_2$).

8.3. *The beta-function*

Exponentiating the grading derivation Y we get a one-parameter group θ_t of automorphisms of the Hopf algebra $\mathcal{H}$, defined on $\mathcal{H}_n$ by

$$\theta_t(x) = e^{nt} x. \tag{99}$$

The map $\varphi \mapsto \varphi \circ Y$ is a derivation of $(\mathcal{L}(\mathcal{H}, \mathcal{A}), *)$, and $\varphi \mapsto \varphi \circ \theta_t$ is an automorphism of $(\mathcal{L}(\mathcal{H}, \mathcal{A}), *)$ for any complex t. We will rather consider the one-parameter group $\varphi \mapsto \varphi \circ \theta_{tz}$ of automorphisms of the algebra $(\mathcal{L}(\mathcal{H}, \mathcal{A}), *)$, i.e.:

$$\varphi^t(x)(z) := e^{tz|x|} \varphi(x)(z). \tag{100}$$

Differentiating at $t = 0$ we get:

$$\frac{d}{dt}_{|t=0}\varphi^t = z(\varphi \circ Y). \tag{101}$$

Let $G_{\mathcal{A}}$ be any of the three groups $G(\mathcal{A})$, $G_1(\mathcal{A})$ and $G_2(\mathcal{A})$. We denote by $G_{\mathcal{A}}^{\mathrm{loc}}$ the set of *local* elements of $G_{\mathcal{A}}$, i.e. those $\varphi \in G_{\mathcal{A}}$ such that the negative part of the Birkhoff decomposition of φ^t does not depend on t, namely:

$$G_{\mathcal{A}}^{\mathrm{loc}} = \left\{ \varphi \in G_{\mathcal{A}} \,\middle|\, \frac{d}{dt}(\varphi^t)_- = 0 \right\}. \tag{102}$$

In particular the dimensional-regularised Feynman rules verify this property: in physical terms, the counter terms do not depend on the choice of the arbitrary mass-parameter μ ('t Hooft's mass) and one must introduce in dimensional regularisation in order to get dimensionless expressions, which is indeed a manifestation of locality (see [CK2]). We also denote by $G_{\mathcal{A}-}^{\mathrm{loc}}$ the elements φ of $G_{\mathcal{A}}^{\mathrm{loc}}$ such that $\varphi = \varphi_-^{*-1}$. Since composition on the right with Y is a derivation for the convolution product, the map $\widetilde{R}$ of Section 8.1 verifies a cocycle property:

$$\widetilde{R}(\varphi * \psi) = \widetilde{R}(\psi) + \psi^{*-1} * \widetilde{R}(\varphi) * \psi. \tag{103}$$

We summarise some key results of [CK3] in the following proposition.

PROPOSITION 28.

(i) *For any $\varphi \in G_{\mathcal{A}}$ there is a one-parameter family h_t in $G_{\mathcal{A}}$ such that $\varphi^t = \varphi * h_t$, and we have*

$$\dot{h}_t := \frac{d}{dt}h_t = h_t * z\widetilde{R}(h_t) + z\widetilde{R}(\varphi) * h_t. \tag{104}$$

(ii) *$z\widetilde{R}$ restricts to a bijection from $G_{\mathcal{A}}^{\mathrm{loc}}$ onto $\mathfrak{g}_{\mathcal{A}} \cap \mathcal{L}(\mathcal{H}, \mathcal{A}_+)$. Moreover it is a bijection from $G_{\mathcal{A}}^{\mathrm{loc}}{}_-$ onto those elements of $\mathfrak{g}_{\mathcal{A}}$ with values in the constants, i.e.:*

$$\mathfrak{g}_{\mathcal{A}}^c = \mathfrak{g}_{\mathcal{A}} \cap \mathcal{L}(\mathcal{H}, \mathbb{C}).$$

(iii) *For $\varphi \in G_{\mathcal{A}}^{\mathrm{loc}}$, the constant term of h_t, defined by*

$$F_t(x) = \lim_{z \to 0} h_t(x)(z) \tag{105}$$

is a one-parameter subgroup of $G_{\mathcal{A}} \cap \mathcal{L}(\mathcal{H}, \mathbb{C})$, the scalar-valued characters of $\mathcal{H}$.

PROOF. For any $\varphi \in G_{\mathcal{A}}$ one can write

$$\varphi^t = \varphi * h_t \tag{106}$$

with $h_t \in G_{\mathcal{A}}$. From (106), (101) and (73) we immediately get:

$$\varphi * \dot{h}_t = \varphi * h_t * z\widetilde{R}(\varphi * h_t).$$

Equation (104) then follows from the cocycle property (103). This proves the first assertion. Now take any character $\varphi \in G_{\mathcal{A}}^{\mathrm{loc}}$ with Birkhoff decomposition $\varphi = \varphi_-^{*-1} * \varphi_+$ and write the Birkhoff decomposition of φ^t:

$$\begin{aligned}\varphi^t &= \left(\varphi^t\right)_-^{*-1} * \left(\varphi^t\right)_+ \\ &= (\varphi_-)^{*-1} * \left(\varphi^t\right)_+ \\ &= \left(\varphi * \varphi_+^{*-1}\right) * \left(\varphi^t\right)_+ \\ &= \varphi * h_t,\end{aligned}$$

with h_t taking values in $\mathcal{A}_+$. Then $z\widetilde{R}(\varphi)$ also takes values in $\mathcal{A}_+$, as a consequence of Eq. (104) at $t = 0$. Conversely, suppose that $z\widetilde{R}(\varphi)$ takes values in $\mathcal{A}_+$. We show that h_t also takes values in $\mathcal{A}_+$ for any t, which immediately implies that φ belongs to $G_{\mathcal{A}}^{\mathrm{loc}}$.

For any $\gamma \in \mathfrak{g}_{\mathcal{A}}$, let us introduce the linear transformation U_γ of $\mathfrak{g}_{\mathcal{A}}$ defined by

$$U_\gamma(\delta) := \gamma * \delta + z\delta \circ Y.$$

If γ belongs to $\mathfrak{g}_{\mathcal{A}} \cap \mathcal{L}(\mathcal{H}, \mathcal{A}_+)$ then U_γ restricts to a linear transformation of $\mathfrak{g}_{\mathcal{A}} \cap \mathcal{L}(\mathcal{H}, \mathcal{A}_+)$.

LEMMA 10. *For any $\varphi \in G_{\mathcal{A}}$, $n \in \mathbb{N}$ we have*

$$z^n \varphi \circ Y^n = \varphi * U^n_{z\widetilde{R}(\varphi)}(e).$$

PROOF. Case $n = 0$ is obvious, $n = 1$ is just the definition of $\widetilde{R}$. We check thus by induction, using again the fact that composition on the right with Y is a derivation for the convolution product:

$$\begin{aligned}z^{n+1}\varphi \circ Y^{n+1} &= z\left(z^n\varphi \circ Y^n\right) \circ Y \\ &= z\left(\varphi * U^n_{z\widetilde{R}(\varphi)}(e)\right) \circ Y \\ &= z(\varphi \circ Y) * U^n_{z\widetilde{R}(\varphi)}(e) + z\varphi * \left(U^n_{z\widetilde{R}(\varphi)}(e) \circ Y\right) \\ &= \varphi * \left(z\widetilde{R}(\varphi) * U^n_{z\widetilde{R}(\varphi)}(e) + zU^n_{z\widetilde{R}(\varphi)}(e) \circ Y\right) \\ &= \varphi * U^{n+1}_{z\widetilde{R}(\varphi)}(e).\end{aligned}$$ □

Let us go back to the proof of Proposition 28. According to Lemma 10 we have for any t, at least formally:

$$\varphi^t = \varphi * \exp\left(tU_{z\widetilde{R}(\varphi)}\right)(e). \tag{107}$$

We still have to fix up the convergence of the exponential just above in the case when $z\widetilde{R}(\varphi)$ belongs to $\mathcal{L}(\mathcal{H}, \mathcal{A}_+)$. Let us consider the following decreasing bifiltration of $\mathcal{L}(\mathcal{H}, \mathcal{A}_+)$:

$$\mathcal{L}_+^{p,q} = \left(z^q\mathcal{L}(\mathcal{H}, \mathcal{A}_+)\right) \cap \mathcal{L}^p,$$

where $\mathcal{L}^p$ is the set of those $\alpha \in \mathcal{L}(\mathcal{H}, \mathcal{A})$ such that $\alpha(x) = 0$ for any $x \in \mathcal{H}$ of degree $\leqslant p - 1$. In particular $\mathcal{L}^1 = \mathfrak{g}_0$. Considering the associated filtration

$$\mathcal{L}_+^n = \sum_{p+q=n} \mathcal{L}_+^{p,q},$$

we see that for any $\gamma \in \mathfrak{g}_0 \cap \mathcal{L}(\mathcal{H}, \mathcal{A}_+)$ the transformation U_γ increases the filtration by 1, i.e.:

$$U_\gamma\big(\mathcal{L}_+^n\big) \subset \mathcal{L}_+^{n+1}.$$

The algebra $\mathcal{L}(\mathcal{H}, \mathcal{A}_+)$ is not complete with respect to the topology induced by this filtration, but the completion is $\mathcal{L}(\mathcal{H}, \widehat{\mathcal{A}_+})$, where $\widehat{\mathcal{A}_+} = \mathbb{C}[\![z]\!]$ stands for algebra of formal series. Hence the right-hand side of (107) is convergent in $\mathcal{L}(\mathcal{H}, \widehat{\mathcal{A}_+})$ with respect to this topology. Hence for any $\gamma \in \mathcal{L}(\mathcal{H}, \mathcal{A}_+)$ and for φ such that $z\widetilde{R}(\varphi) = \gamma$ we have $\varphi^t = \varphi * h_t$ with $h_t \in \mathcal{L}(\mathcal{H}, \widehat{\mathcal{A}_+})$ for any t. On the other hand we already know that h_t takes values in meromorphic functions for each t. So h_t belongs to $\mathcal{L}(\mathcal{H}, \mathcal{A}_+)$, which proves the first part of the second assertion. Equation (104) at $t = 0$ reads:

$$z\widetilde{R}(\varphi) = \dot{h}(0) = \frac{d}{dt}_{|t=0}\big(\varphi^t\big)_+. \tag{108}$$

For $\varphi \in G^{\mathrm{loc}}_{\mathcal{A}_-}$ we have, thanks to the property $\varphi(\operatorname{Ker}\varepsilon) \subset \mathcal{A}_-$:

$$\begin{aligned} h_t(x) = (\varphi^t)_+(x) = (I - \pi)\left(\varphi^t(x) + \sum_{(x)} \varphi^{*-1}(x')\varphi^t(x'')\right) \\ = t(I - \pi)\left(z|x|\varphi(x) + z\sum_{(x)} \varphi^{*-1}(x')\varphi(x'')|x''|\right) + O\big(t^2\big) \\ = t\operatorname{Res}(\varphi \circ Y) + O\big(t^2\big), \end{aligned}$$

hence

$$\dot{h}(0) = \operatorname{Res}(\varphi \circ Y). \tag{109}$$

From Eqs. (101), (73) and (109) we get

$$z\widetilde{R}(\varphi) = \operatorname{Res}(\varphi \circ Y) \tag{110}$$

for any $\varphi \in G^{\mathrm{loc}}_{\mathcal{A}_-}$, hence $z\widetilde{R}(\varphi) \in \mathfrak{g}^c$. Conversely let β in $\mathfrak{g}^c$. Consider $\psi = \widetilde{R}^{-1}(z^{-1}\beta)$. This element of $G_{\mathcal{A}}$ verifies by definition, thanks to Eq. (73):

$$z\psi \circ Y = \psi * \beta.$$

Hence for any $x \in \operatorname{Ker}\varepsilon$ we have

$$z\psi(x) = \frac{1}{|x|}\left(\beta(x) + \sum_{(x)} \psi(x')\beta(x'')\right).$$

As $\beta(x)$ is a constant (as a function of the complex variable z) it is easily seen by induction on $|x|$ that the right-hand side evaluated at z has a limit when z tends to zero. Thus $\psi(x) \in \mathcal{A}_-$, and then

$$\psi = \widetilde{R}^{-1}\left(\frac{1}{z}\beta\right) \in G^{\mathrm{loc}}_{\mathcal{A}_-},$$

which proves assertion (ii).

Let us prove assertion (iii). Equation $\varphi^t = \varphi * h_t$ together with $(\varphi^t)^s = \varphi^{t+s}$ yields:

$$h_{s+t} = h_s * (h_t)^s. \tag{111}$$

Taking values at $z = 0$ immediately yields the one-parameter group property

$$F_{s+t} = F_s * F_t \tag{112}$$

thanks to the fact that evaluation at $z = 0$ is an algebra morphism. □

We can now give a definition of the beta-function [CK2,EM1,S]: for any $\varphi \in G_{\mathcal{A}}^{\mathrm{loc}}$, the beta-function of φ is the generator of the one-parameter group F_t defined by Eq. (105) in Proposition 28. It is the element of the dual $\mathcal{H}^*$ defined by

$$\beta(\varphi) := \frac{d}{dt}\bigg|_{t=0} F_t(x) \tag{113}$$

for any $x \in \mathcal{H}$.

PROPOSITION 29. *For any $\varphi \in G_{\mathcal{A}}^{\mathrm{loc}}$ the beta-function of φ coincides with the one of the negative part φ_-^{*-1} in the Birkhoff decomposition. It is given by any of the three expressions*:

$$\begin{aligned}\beta(\varphi) &= \operatorname{Res} \widetilde{R}(\varphi)\\ &= \operatorname{Res}\bigl(\varphi_-^{*-1} \circ Y\bigr)\\ &= -\operatorname{Res}(\varphi_- \circ Y).\end{aligned}$$

PROOF. The third equality will be derived from the second by taking residues on both sides of the equation:

$$0 = \widetilde{R}(e) = \widetilde{R}(\varphi_-) + \varphi_-^{*-1} * \widetilde{R}\bigl(\varphi_-^{*-1}\bigr) * \varphi_-,$$

which is a special instance of the cocycle formula (103). Suppose first $\varphi \in G_{\mathcal{A}_-}^{\mathrm{loc}}$, hence $\varphi_-^{*-1} = \varphi$. Then $z\widetilde{R}(\varphi)$ is a constant according to assertion (ii) of Proposition 28. The proposition then follows from Eq. (109) evaluated at $z = 0$, and Eq. (110). Suppose now $\varphi \in G_{\mathcal{A}}^{\mathrm{loc}}$, and consider its Birkhoff decomposition. As both components belong to $G_{\mathcal{A}}^{\mathrm{loc}}$ we can apply Proposition 28 to them. In particular we have:

$$\begin{aligned}\varphi^t &= \varphi * h_t,\\ (\varphi_-^{*-1})^t &= \varphi_-^{*-1} * v_t,\\ (\varphi_+)^t &= \varphi_+ * w_t,\end{aligned}$$

and equality $\varphi^t = (\varphi_-^{*-1})^t * (\varphi_+)^t$ yields:

$$h_t = (\varphi_+)^{*-1} * v_t * \varphi_+ * w_t. \tag{114}$$

We denote by F_t, V_t, W_t the one-parameter groups obtained from h_t, v_t, w_t, respectively, by letting the complex variable z go to zero. It is clear that $\varphi^+_{|z=0} = e$, and similarly that W_t is the constant one-parameter group reduced to the co-unit ε. Hence Eq. (114) at $z = 0$

reduces to

$$F_t = V_t, \tag{115}$$

hence the first assertion. The cocycle equation (103) applied to the Birkhoff decomposition reads:

$$\widetilde{R}(\varphi) = \widetilde{R}(\varphi_+) + (\varphi_+)^{*-1} * \widetilde{R}(\varphi_-^{*-1}) * \varphi_+.$$

Taking residues of both sides yields:

$$\operatorname{Res} \widetilde{R}(\varphi) = \operatorname{Res} \widetilde{R}(\varphi_-^{*-1}),$$

which ends the proof. □

The one-parameter group $F_t = V_t$ above is the *renormalisation group* of φ [CK3].

REMARK 9. As it is possible to reconstruct φ_- from $\beta(\varphi)$ using the scattering-type formula of Theorem 7, the term φ_- (i.e. the divergence structure of φ) is uniquely determined by its residue.

References

[Ab] E. Abe, Hopf Algebras, Cambridge Univ. Press, Cambridge, 1980.

[ABS] M. Aguiar, N. Bergeron, F. Sottile, Combinatorial Hopf algebras and generalized Dehn–Sommerville relations, Compos. Math. 142 (2006) 1–30.

[Ar] H. Araki, Expansional in Banach algebras, Ann. Sci. École Norm. Sup. (4) 6 (1973) 67–84.

[B] N. Bourbaki, Algèbre, Chapitre 8, Hermann, Paris, 2007.

[Br] Ch. Brouder, Quantum field theory meets Hopf algebra, hep-th/0611153, 2006.

[BF1] Ch. Brouder, A. Frabetti, Noncommutative renormalization for massless QED, hep-th/0011161, 2000.

[BF2] Ch. Brouder, A. Frabetti, Renormalization of QED with planar binary trees, European Phys. J. C 19 (2001) 715–741.

[BF3] Ch. Brouder, A. Frabetti, QED Hopf algebras on planar binary trees, J. Algebra 267 (2003) 298–322.

[BFK] Ch. Brouder, A. Frabetti, Ch. Krattenthaler, Noncommutative Hopf algebra of formal diffeomorphisms, Adv. Math. 200 (2) (2006) 479–524.

[BK] Ch. Bergbauer, D. Kreimer, Hopf algebras in renormalization theory: Locality and Dyson–Schwinger equations from Hochschild cohomology, in: IRMA Lect. Math. Theor. Phys., vol. 10, Eur. Math. Soc., Zürich, 2006, pp. 133–164.

[BP] N.N. Bogoliubov, O.S. Parasiuk, On the multiplication of causal functions in the quantum theory of fields, Acta Math. 97 (1957) 227–266.

[BS] Ch. Brouder, W. Schmitt, Renormalization as a functor on bialgebras, J. Pure Appl. Algebra 209 (2) (2007) 477–495.

[C] J. Collins, Renormalization, Cambridge Monogr. Math. Phys., Cambridge Univ. Press, Cambridge, 1984.

[CK] A. Connes, D. Kreimer, Hopf algebras, renormalisation and noncommutative geometry, Comm. Math. Phys. 199 (1998) 203–242.

[CK1] A. Connes, D. Kreimer, Renormalization in Quantum Field Theory and the Riemann–Hilbert problem I: The Hopf algebra structure of graphs and the main theorem, Comm. Math. Phys. 210 (2000) 249–273.

[CK2] A. Connes, D. Kreimer, Renormalization in Quantum Field Theory and the Riemann–Hilbert problem II: The β-function, diffeomorphisms and the renormalization group, Comm. Math. Phys. 216 (2001) 215–241.

[CK3] A. Connes, D. Kreimer, Insertion and elimination: the doubly infinite Lie algebra of Feynman graphs, Ann. Henri Poincaré 3 (3) (2002) 411–433.

[CM1] A. Connes, M. Marcolli, Renormalization, the Riemann–Hilbert correspondence and motivic Galois theory, math.QA/0411114, 2004.

[CM2] A. Connes, M. Marcolli, A walk in the noncommutative garden, math.QA/0601054, 2006.

[CMo] A. Connes, H. Moscovici, Hopf algebras, cyclic cohomology and the transverse index theorem, Comm. Math. Phys. 198 (1998) 198–246.

[Di] J. Dixmier, Algèbres enveloppantes, Gautier-Villars, Paris, 1974.

[DF] R.K. Dennis, B. Farb, Noncommutative Algebra, Springer-Verlag, Berlin, 1993.

[DK] Yu.A. Drozd, V.V. Kirichenko, Finite Dimensional Algebras, Springer-Verlag, Berlin, 1994 (English edition).

[DNR] S. Dăscălescu, C. Năstăsescu, S. Raianu, Hopf Algebras, an Introduction, Pure Appl. Math., vol. 235, Dekker, 2001.

[E] P. Etingof, Note on dimensional regularization, in: Quantum Fields and Strings: A Course for Mathematicians, vol. 1, Amer. Math. Soc./IAS, 1999, pp. 597–607.

[EG1] K. Ebrahimi-Fard, L. Guo, Matrix representation of renormalization in perturbative quantum field theory, hep-th/0508155, 2005.

[EG2] K. Ebrahimi-Fard, L. Guo, Quasi-shuffles, mixable shuffles and Hopf algebras, J. Algebraic Combin. 24 (1) (2006) 83–101.

[EGK1] K. Ebrahimi-Fard, L. Guo, D. Kreimer, Integrable renormalization I: The ladder case, J. Math. Phys. 45 (10) (2004) 3758–3769.

[EGK2] K. Ebrahimi-Fard, L. Guo, D. Kreimer, Integrable renormalization II: The general case, Ann. H. Poincaré 6 (2005) 369–395.

[EGK3] K. Ebrahimi-Fard, L. Guo, D. Kreimer, Spitzer's identity and the algebraic Birkhoff decomposition in pQFT, J. Phys. A 37 (2004) 11036–11052.

[EGM] K. Ebrahimi-Fard, L. Guo, D. Manchon, Birkhoff type decompositions and the Baker–Campbell–Hausdorff recursion, Comm. Math. Phys. 267 (2006) 821–845.

[EGGV] K. Ebrahimi-Fard, J.M. Gracia-Bondia, L. Guo, J.C. Varilly, Combinatorics of renormalization as matrix calculus, Phys. Lett. B 632 (4) (2006) 552–558.

[EGP] K. Ebrahimi-Fard, J. Gracia-Bondia, F. Patras, A Lie theoretic approach to renormalization, hep-th/0609035, Comm. Math. Phys. 276 (2007) 519–549.

[EM1] K. Ebrahimi-Fard, D. Manchon, On matrix differential equations in the Hopf algebra of renormalization, Adv. Theor. Math. Phys. 10 (2006) 879–913.

[EM2] K. Ebrahimi-Fard, D. Manchon, A Magnus- and Fer- type formula in dendriform algebras, arXiv:0707.0607[math.CO], 2007.

[EMP1] K. Ebrahimi-Fard, D. Manchon, F. Patras, A noncommutative Bohnenblust–Spitzer identity for Rota–Baxter algebras solves Bogoliubov's recursion, arXiv: 0705.1265, 2007.

[EMP2] K. Ebrahimi-Fard, D. Manchon, F. Patras, New identities in dendriform algebras, arXiv: 0705.2636 [math.CO], 2007.

[ENR] M. Espie, J.-Ch. Novelli, G. Racinet, Formal Computations about Multiple Zeta Values, IRMA Lect. Math. Theor. Phys., vol. 3, Eur. Math. Soc., Berlin, 2003.

[FdB] F. Faà di Bruno, Sullo sviluppo delle funzioni, Ann. Sci. Mat. Fis. Roma 6 (1855) 479–480.

[F] L. Foissy, Les algèbres de Hopf des arbres enracinés décorés I, II, Bull. Sci. Math. 126 (2002) 193–239, 249–288.

[F2] L. Foissy, Bidendriform bialgebras, trees and free quasi-symmetric functions, math/0505207, 2005.

[FG] H. Figueroa, J.M. Gracia-Bondía, Combinatorial Hopf algebras in Quantum Field Theory I, Rev. Math. Phys. 17 (2005) 881–976.

[GL] R. Grossman, R.G. Larson, Hopf-algebraic structure of families of trees, J. Algebra 126 (1) (1989) 184–210.

[GZ] L. Guo, B. Zhang, Renormalization of multiple zeta values, math.NT/0606076, 2006.

[H] K. Hepp, Proof of the Bogoliubov–Parasiuk theorem on renormalization, Comm. Math. Phys. 2 (1966) 301–326.

[Hu] K. Huang, Quantum Field Theory: From Operators to Path Integrals, Wiley, New York, 1998.

[Ho1] M.E. Hoffman, Quasi-shuffle products, J. Algebraic Combin. 11 (2000) 49–68.

[Ho2] M.E. Hoffman, The Hopf algebra structure of multiple harmonic sums, Nuclear Phys. B Proc. Suppl. 135 (2004) 215–219, math.QA/0406589.

[HV] G. 't Hooft, M. Veltman, Regularization and renormalization of gauge fields, Nucl. Phys. B 44 (1972) 189–213.

[J] N. Jacobson, Basic Algebra II, second ed., Freeman, New York, 1989.

[JR] S.A. Joni, G.-C. Rota, Coalgebras and bialgebras in combinatorics, Stud. Appl. Math. 61 (1979) 93–139.

[K] C. Kassel, Quantum Groups, Springer-Verlag, Berlin, 1995.

[K1] D. Kreimer, On the Hopf algebra structure of perturbative quantum field theories, Adv. Theor. Math. Phys. 2 (1998).

[K2] D. Kreimer, Structures in Feynman graphs: Hopf algebras and symmetries, Proc. Sympos. Pure Math., vol. 73, Amer. Math. Soc., Providence, RI, 2005, hep-th/0202110.

[K3] D. Kreimer, Anatomy of a gauge theory, Ann. Phys. 3212 (2006) 2757.

[K4] D. Kreimer, Dyson–Schwinger Equations: From Hopf Algebras to Number Theory, in Universality and Renormalization, in: Fields Inst. Commun., vol. 50, Amer. Math. Soc., Providence, RI, 2007, 225 p.

[KY] D. Kreimer, K. Yeats, An étude in non-linear Dyson–Schwinger equations, Nucl. Phys. Proc. Suppl. 160 (2006) 116.

[LR] J.-L. Loday, M. Ronco, Hopf algebra of the planar binary trees, Adv. Math. 139 (1998) 293–309.

[M1] D. Manchon, L'algèbre de Hopf bitensorielle, Comm. Algebra 25 (5) (1997) 1537–1551.

[M2] D. Manchon, Hopf algebras, from basics to applications to renormalization, Rencontr. Math. Glanon 2001 (published in 2003); math.QA/0408405.

[MP1] D. Manchon, S. Paycha, Shuffle relations for regularised integrals of symbols, Comm. Math. Phys. 270 (2007) 13–51.

[MP2] D. Manchon, S. Paycha, Chen sums of symbols and renormalised multiple zeta values, math.NT/0702135, 2007.

[MR] C. Malvenuto, Ch. Reutenauer, Duality between quasi-symmetric functions and the Solomon descent algebra, J. Algebra 177 (3) (1995) 967–982.

[Mo] S. Montgomery, Some remarks on filtrations of Hopf algebras, Comm. Algebra 21 (3) (1993) 999–1007.

[PR] F. Patras, Ch. Reutenauer, On Dynkin and Klyachko idempotents in graded bialgebras, Adv. Appl. Math. 28 (2002) 560–579.

[R] Ch. Reutenauer, Free Lie Algebras, Oxford Univ. Press, Oxford, 1993.

[Sp] E.R. Speer, Renormalization and Ward identities using complex space–time dimension, J. Math. Phys. 15 (1) (1974) 1–6.

[S] M. Sakakibara, On the differential equations of the characters for the renormalization group, Modern Phys. Lett. A 19 (2004) 1453–1456.

[Sw] M.E. Sweedler, Hopf Algebras, Benjamin, New York, 1969.

[TW] E.J. Taft, R.L. Wilson, On antipodes in pointed Hopf algebras, J. Algebra 28 (1974) 27–32.

[V] V.S. Varadarajan, Lie Groups, Lie Algebras and Their Representations, Springer-Verlag, Berlin, 1984.

[VS] W. Van Suijlekom, The Hopf algebra of Feynman graphs in quantum electrodynamics, Lett. Math. Phys. 77 (2006) 265–281.

[VS2] W. Van Suijlekom, Renormalization of gauge fields: A Hopf algebra approach, hep-th/0610137, 2006.

[Z] W. Zimmermann, Convergence of Bogoliubov's method of renormalization in momentum space, Comm. Math. Phys. 15 (1969) 208–234.

This page intentionally left blank

Classification of Semisimple Hopf Algebras

Akira Masuoka

Institute of Mathematics, *University of Tsukuba*, *Ibaraki* 305-8571, *Japan*
E-mail: *akira@math.tsukuba.ac.jp*

Contents

HANDBOOK OF ALGEBRA, VOL. 5
Edited by M. Hazewinkel

DOI: 10.1016/S1570-7954(07)05008-5

This page intentionally left blank

Introduction

The coordinate ring of an algebraic group is a commutative Hopf algebra. G. Hochschild was the first to study in detail commutative Hopf algebras for applications to algebraic groups. As a successor of Hochschild, M. Sweedler founded in the end of the 1960s the study of (possibly non-commutative) Hopf algebras; he apparently regarded Hopf algebras as a non-commutative analogue of algebraic groups. Almost at the same time, G. Kac, an operator-algebraist, reached the notion of a *ring group*, nowadays called a *Kac algebra*, as a non-commutative analogue of locally compact groups; a Kac algebra, in particular one of finite dimension is a C^*-algebra, hence especially a semisimple algebra over $\mathbb{C}$, which is at the same time a Hopf algebra (Definition 7.2).

In 1975, I. Kaplansky published his lecture notes "Bialgebras" [Kap], proposing 10 conjectures in the appendix. More than half of the conjectures ask whether finite-dimensional Hopf algebras, especially semisimple ones, have the same properties as finite groups. Unfortunately, Kac's pioneering work, though referred to in [Kap], did not draw attention for a long time until Y. Zhu [Z] refined some of the results by Kac [K2] in an Hopf-algebra context, in order to answer Kaplansky's conjecture 8 which states that a Hopf algebra of prime dimension is commutative and cocommutative. This work by Zhu served to draw attention to Kac's work, and motivated many people including the author, Natale and Kashina to prove classification results for semisimple (and later, even arbitrary) Hopf algebras of various special finite dimensions. On the other hand, A. Ocneanu formulated around 1986 (without proof) a correspondence between finite-dimensional Kac algebras and depth 2 subfactors of type II_1 (see [Oc]), which was proved by Szymański [Sz], David [Da], Popa [P1,P2] and others. Then a close relation of this result with Kac's work was pointed out by Izumi and Kosaki [IK], and the author [M8].

In the middle of the 1980s, the notion of a quantum group was discovered independently by V. Drinfeld, M. Jimbo and S.L. Woronowicz. Since then quantum group theory has had a strong impact on Hopf algebra theory, and especially convinced the mathematical community of the importance of braidings on tensor categories. An important aspect of a Hopf algebra, say H, is that the H-modules form a tensor category. The braidings (respectively, symmetries, i.e., involutory braidings) on such a tensor category are in 1–1 correspondence with the quasitriangular (respectively, triangular) structures on the Hopf algebra H; see [D1]. P. Etingof and S. Gelaki have produced a plethora of remarkable results especially on quasitriangular Hopf algebras, including the complete classification of finite-dimensional triangular Hopf algebras in characteristic zero [EG7].

This paper surveys recent achievements on finite-dimensional semisimple (mainly) Hopf algebras, giving explanatory examples. We will start in Section 1 by surveying the present status of some of the Kaplansky conjectures; especially, pioneering work by Larson and Radford (see [L,LR,R1,R2]) should be respectfully mentioned. Section 2 contains the classification results in dimension pq, where p, q are primes, and the lifting theorem due to Etingof and Gelaki [EG3], which makes it possible to derive results in positive characteristic from those in characteristic zero. In Section 3 we discuss semisolvability as formulated by Montgomery and Witherspoon [MW], and also describe classification results in dimensions p^3, pq^2. Section 4 contains the above-cited classification result, due to Etingof and Gelaki [EG5], of finite-dimensional triangular Hopf algebras, but restricted to semisimple (and cosemisimple) ones. Sections 5 and 6 are devoted to those results on Hopf-algebra extensions associated to a matched pair of finite groups, which includes an exact cohomology sequence due to Kac [K1] and its consequences. Sections 7 and 8 are devoted to topics on finite-dimensional Kac algebras. In Section 7, we describe the famous example of an 8-dimensional Kac algebra due to Kac and Paljutkin [KP], and its generalizations. Finally in Section 8, we review the correspondence mentioned above between finite-dimensional Kac algebras and depth 2 subfactors of type II_1, and its relation with Kac's work [K1] on cohomology.

Because of my restricted knowledge, some important relevant topics are not discussed, including those on the Frobenius–Schur indicators [LM], and on the exponent [Ks1], [EG4], as well as various generalized results for weak Hopf algebras [ENO].

As survey articles which treat of relevant topics, let me recommend [A], [Mo2,Mo3] and [SS].

We work over a fixed ground field k. We will state explicitly when there are assumptions on k, especially on its characteristic $\operatorname{ch} k$. In Sections 7, 8 we suppose $k = \mathbb{C}$, the complex number field. We let H denote a Hopf algebra (over k) with coalgebra structure

$$\Delta : H \to H \otimes H, \quad \Delta(a) = \sum a_1 \otimes a_2, \qquad \varepsilon : H \to k$$

and antipode S. Let $H^+ = \operatorname{Ker} \varepsilon$ denote the augmentation ideal. In most sections below, H will be supposed to be finite-dimensional, in which case H^* denotes the dual Hopf algebra of H.

1. Kaplansky's conjectures

Let us survey the present status of some of Kaplansky's conjectures; see the introduction. First of all Kaplansky stated the following one, as an analogue of the Lagrange theorem on groups.

CONJECTURE 1. *A Hopf algebra H is free, as a left or right module, over every Hopf subalgebra K.*

Immediately, Oberst and Schneider [OS] gave a counter-example with $\dim H = \infty$.

THEOREM 1.1. *(See [NZ,S1].) Suppose* $\dim H < \infty$. *Then Conjecture 1 is true. Moreover, H has a normal K-basis in the sense that there is a left K-linear and right H/K^+H-colinear isomorphism $H \simeq K \otimes H/K^+H$.*

Note here that H/K^+H is a quotient (right H-module) coalgebra of H. Recently, Skryabin [Sk] proved the same result in the generalized situation that H is weakly finite, and K is a finite-dimensional left (or right) coideal subalgebra.

Next, let us pick on:

CONJECTURE 10. *Suppose that k is algebraically closed, and let $d > 0$ be an integer such that* $\operatorname{ch} k \nmid d$. *Then the number of isomorphism classes of Hopf algebras of dimension d is finite.*

THEOREM 1.2.

(1) *(See [St].) Let k be as above. Then the number of isomorphism classes of semisimple and cosemisimple Hopf algebras of a given dimension is finite.*

(2) *(See [AS,BDG,G,Mu].) Conjecture* 10 *is not true in general.*

Each of [AS,BDG,G,Mu] constructed a family of Hopf algebras of a fixed dimension that consists of infinitely many isomorphism classes. In [M6], it was proved that for all Hopf algebras in each family, their comodule categories are tensor-equivalent to each other, or in other words, the duals of those Hopf algebras are twists (Definition 4.2) of each other. Etingof and Gelaki [EG6] showed that there exist infinitely many Hopf algebras of a fixed dimension that are non-isomorphic even up to a twist.

Stefan's idea to prove part (1) above was elaborated by Etingof and Gelaki to a generalized cohomology-vanishing theorem [EG3, Theorem 1.2], which has the following as a corollary; see also [R2].

THEOREM 1.3. *(See [EG3].) The set* $\mathrm{Hom}_{\mathrm{Hopf}}(H_1, H_2)$ *of all Hopf algebra maps from a finite-dimensional semisimple Hopf algebra* H_1 *to a finite-dimensional cosemisimple Hopf algebra* H_2 *is finite.*

Suppose that H is a finite-dimensional Hopf algebra. By Radford [R1], the antipode S of H has finite order; more precisely the composite $S \circ \cdots \circ S = S^{4d}$ with $d = \dim H$ equals the identity id_H. Let us consider the conditions:

(a) H is semisimple as an algebra,

(b) H is cosemisimple as a coalgebra,

(c) $\mathrm{ch}\, k \nmid \dim H$,

(d) S is involutory, i.e., $S \circ S = \mathrm{id}_H$.

Conjectures 5, 7 are expressed by:

CONJECTURE 5. (a) *or* (b) $\Rightarrow$ (d).

CONJECTURE 7. (a) *and* (b) $\Rightarrow$ (c).

By summarizing results by Larson [L], Larson and Radford [LR], and Etingof and Gelaki [EG3], we have:

THEOREM 1.4.

(1) (a) *and* (b) $\Leftrightarrow$ (c) *and* (d).

(2) *If* $\mathrm{ch}\, k = 0$ *or* $> d^{\varphi(d)/2}$, *where* $d = \dim H$, *then* (a) $\Leftrightarrow$ (b) $\Leftrightarrow$ (d). *Here* φ *denotes the Euler function.*

Thus, Conjecture 7 is proved to be true. Conjecture 5 is true if $\mathrm{ch}\, k = 0$ or $> d^{\varphi(d)/2}$.

Suppose that k is algebraically closed. Let H be a finite-dimensional Hopf algebra which is semisimple. Then H is, as an algebra, of the form

$$H \simeq M_{n_1}(k) \times M_{n_2}(k) \times \cdots \times M_{n_r}(k), \tag{1.5}$$

where $M_n(k)$ denotes the algebra of all $n \times n$ matrices. We may suppose that $H^+ = M_{n_2}(k) \times \cdots \times M_{n_r}(k)$, and so $n_1 = 1$. In this situation Kaplansky posed:

CONJECTURE 6. *$n_i \mid \dim H$ for each $1 \leqslant i \leqslant r$.*

If $H = kG$, a finite group algebra, then this is a well-known result by Frobenius, which is generalized by part (1) of the next theorem due to Etingof and Gelaki.

THEOREM 1.6. *Let the notation be as above. We suppose in addition that H is cosemisimple, if $\operatorname{ch} k > 0$.*

(1) *(See [EG1]; see also [S2,T2].) Conjecture 6 is true if H is quasitriangular* [D1] *(or see [Mo1, p. 180]).*

(2) *(See [KSZ1,KSZ2].) Suppose that $\dim H$ is odd.*
   (i) *Every n_i $(1 \leqslant i \leqslant r)$ must be odd.*
   (ii) *Any simple component $M_{n_i}(k)$ $(2 \leqslant i \leqslant r)$ included in H^+ cannot be stable under the antipode S.*

Keep the notation be as above. The traces tr_i of the simple components $M_{n_i}(k)$ given in (1.5) span a subalgebra in H^*, which we denote by $R_k(H) = \sum_{i=1}^r k \operatorname{tr}_i$. This is precisely the base extension $\otimes_{\mathbb{Z}} k$ of the character ring of H. By [LR, Proposition 1], the character $\Lambda := \sum_{i=1}^r n_i \operatorname{tr}_i$ of the regular representation is an integral. Note that if $\operatorname{ch} k \nmid \dim H$, then $e_\Lambda := (\dim H)^{-1}\Lambda$ is a (central) primitive idempotent in $R_k(H)$ such that $\dim e_\Lambda H^* = 1$.

THEOREM 1.7. *(See [K2,Z].) Suppose $\operatorname{ch} k = 0$, and that k is algebraically closed. Then $R_k(H)$ is a semisimple algebra, and for each primitive idempotent e in $R_k(H)$,*

$$\dim eH^* \mid \dim H.$$

Finally, let us pick:

CONJECTURE 8. *A finite-dimensional Hopf algebra H of a prime dimension p is commutative and cocommutative.*

THEOREM 1.8. *Suppose $\operatorname{ch} k = 0$ or $> p$. Then Conjecture 8 is true. Therefore, if k is algebraically closed, such an H as above is isomorphic to the group algebra kC_p of the cyclic group C_p of order p.*

This theorem was first proved by Kac [K2] for Kac algebras (Definition 7.2), then by Zhu [Z] in characteristic zero (as a direct consequence of his Theorem 1.7), and finally by Etingof and Gelaki [EG3] in positive characteristic; see Remark 2.5.

2. Some classification results

ASSUMPTION 2.1. *Throughout this section, k is supposed to be algebraically closed, and H denotes a finite-dimensional Hopf algebra over k.*

PROPOSITION 2.2. *Suppose* $\operatorname{ch} k \nmid \dim H$ (*i.e., condition* (c) *in Section* 1). *Then, H is cocommutative if and only if it is isomorphic to some group algebra kG.*

PROOF. This follows since the assumption implies that if H is cocommutative, the irreducible component of H containing 1 is trivial. □

In a finite group algebra kG, the group elements g ($\in G$) are regarded as grouplikes, so that $\Delta(g) = g \otimes g$, $\varepsilon(g) = 1$, $S(g) = g^{-1}$. This defines the Hopf algebra structure on kG. We let $k^G = (kG)^*$ denote the dual Hopf algebra of kG. Let $(e_g)_{g\in G}$ denote the basis of k^G which is dual to the canonical basis $(g)_{g\in G}$ of kG. Thus,

$$\Delta(e_g) = \sum_{h\in G} e_h \otimes e_{h^{-1}g}, \qquad \varepsilon(e_g) = \delta_{1,g}, \qquad S(e_g) = e_{g^{-1}}.$$

The proposition above can be dualized so that under the same assumption, H is commutative if and only if it is isomorphic to some k^G.

DEFINITION 2.3. H is said to be *trivial*, if it is isomorphic to some kG or k^G.

THEOREM 2.4. *Suppose* $\dim H = pq$, *where p and q are primes with $p \leqslant q$.*

(1) *(See [EG2,M1,M3].) If H is semisimple and cosemisimple* (*then necessarily,* $\operatorname{ch} k \neq p, q$), *it is trivial.*
(2) *(See [Ng2,EG8].) Suppose* $\operatorname{ch} k = 0$. *If either $p = 2 < q$ or $2 < p < q \leqslant 2p + 1$, then H is necessarily trivial.*
(3) *(See [Ng1].) Suppose* $\operatorname{ch} k = 0$. *If $p = q$, then H is either trivial or isomorphic to one of the pointed Hopf algebras defined by Taft* [Tf].

It seems to be widely believed that if $p \neq q$, the conclusion of (2) should be true without any additional restriction on p, q; see [BD] for a positive result in dimension 65 which is not covered by (2).

First, in characteristic zero, part (1) was proved by the author [M1,M2,M3] in the case when $p = 2$ or $p = q$, and by Etingof and Gelaki [EG2] (see also Natale [N1]) in the remaining cases. The results were then applied to prove things in positive characteristic, by using the following lifting theorem due to Etingof and Gelaki.

REMARK 2.5. The situation is the same for the proofs of Theorems 1.6(1), 1.8, 3.10 and 3.12(2).

Suppose $\operatorname{ch} k > 0$. Let $\mathcal{O} = W(k)$ denote the ring of Witt vectors over k; it is a complete discrete variation ring of characteristic zero, with residue field k. Let K denote the quotient field of $\mathcal{O}$; it is not algebraically closed, in general.

THEOREM 2.6. *(See [EG3].)*

(1) *Given a semisimple and cosemisimple Hopf algebra H over k, there is a unique* (*up to isomorphism*) *Hopf algebra $\bar{H}$ over $\mathcal{O}$ which is free as an $\mathcal{O}$-module, and such that $\bar{H}/\mathfrak{m}\bar{H} \simeq H$, where $\mathfrak{m}$ is the maximal ideal of $\mathcal{O}$.*

(2) *The Hopf algebra* $H_0 := \bar{H} \otimes_{\mathcal{O}} K$ *over* K *is split semisimple and split cosemisimple. Moreover, if* H *(respectively,* H^**)* $\simeq M_{n_1}(k) \times \cdots \times M_{n_r}(k)$*, then* H_0 *(respectively,* H_0^**)* $\simeq M_{n_1}(K) \times \cdots \times M_{n_r}(K)$.

(3) $H \mapsto H_0$ *gives rise to a functor from the category of* [(*quasi*)*triangular*] *semisimple and cosemisimple Hopf algebras over* k *to the category of* [(*quasi*)*triangular*] *semisimple Hopf algebras over* K.

3. Semisolvable Hopf algebras

Let H be a Hopf algebra, and let R be an algebra. Suppose that

$$\rightharpoonup : H \otimes R \to R, \qquad \sigma : H \otimes H \to R$$

are linear maps that satisfy:

$$\text{(a) } \sigma \text{ is convolution-invertible,} \tag{3.1a}$$

$$\text{(b) } a \rightharpoonup xy = \sum (a_1 \rightharpoonup x)(a_2 \rightharpoonup y), \ \ a \rightharpoonup 1 = \varepsilon(a)1, \tag{3.1b}$$

$$\text{(c) } \sum \bigl(a_1 \rightharpoonup \sigma(b_1, c_1)\bigr)\sigma(a_2, b_2c_2) = \sum \sigma(a_1, b_1)\sigma(a_2b_2, c), \tag{3.1c}$$

$$\text{(d) } \sigma(a, 1) = \varepsilon(a)1 = \sigma(1, a), \tag{3.1d}$$

where $a, b, c \in H$, $x, y \in R$. Then the vector space $R \otimes H$ is made into an algebra, which we denote by $R \rtimes_\sigma H$, with unit $1 \otimes 1$ and product

$$(x \otimes a)(y \otimes b) = \sum x(a_1 \rightharpoonup y)\sigma(a_2, b_1) \otimes a_3b_2.$$

DEFINITION 3.2. (See [DT,BCM].) $R \rtimes_\sigma H$ is called the *crossed product* of H with R. It is a right H-comodule algebra with respect to the obvious H-comodule structure $\mathrm{id}_R \otimes \Delta : R \rtimes_\sigma H \to (R \rtimes_\sigma H) \otimes H$.

In what follows we impose:

ASSUMPTION 3.3. *k is algebraically closed.*

Recall from Section 1, Kaplansky's Conjecture 6. We say that a finite-dimensional semisimple algebra A is *of Frobenius type* if it posseses the property described by the conjecture, namely if $n_i \mid \dim A$ for each $1 \leqslant i \leqslant r$, where $A \simeq M_{n_1}(k) \times \cdots \times M_{n_r}(k)$. Conjecture 6 is generalized to:

CONJECTURE 6′. *If R is a finite-dimensional semisimple algebra of Frobenius type, then any crossed product $R \rtimes_\sigma H$ with a semisimple Hopf algebra H, that is necessarily semisimple, is of Frobenius type.*

This is open, as is Conjecture 6. Montgomery and Witherspoon proved:

PROPOSITION 3.4. *(See [MW].) Conjecture* $6'$ *holds true, if* H *is trivial* (*Definition* 2.3) *and if* $\operatorname{ch} k \nmid \dim H$ (*or equivalently if* H *is cosemisimple as well*).

DEFINITION 3.5. (See [MW,A].) A finite-dimensional Hopf algebra H is said to be *lower-semisolvable*, if there is a sequence

$$H = H_s \supset H_{s-1} \supset \cdots \supset H_1 \supset H_0 = k$$

of Hopf subalgebras such that each H_i is normal[1] in H_{i+1} $(0 \leqslant i < s)$, and the factor $H_{i+1}/H_i^+ H_{i+1}$ is trivial. H is said to be *upper-semisolvable* if the dual H^* is lower-semisolvable. In other words, H is *lower-semisolvable* (respectively, *upper-semisolvable*), if it is obtained as a result H_s of successive extensions (see Section 4)

$$H_i \rightarrowtail H_{i+1} \twoheadrightarrow J_i \tag{3.6a}$$

$$(\text{respectively,}\ \ J_i \rightarrowtail H_{i+1} \twoheadrightarrow H_i) \tag{3.6b}$$

$(0 \leqslant i < s)$ starting from $H_0 = k$, where the J_i are trivial Hopf algebras. We say that H is *semisolvable* if it satisfies the relaxed condition that the i-th extension is of one of the forms (3.6a) or (3.6b).

Using Proposition 3.4, Montgomery and Witherspoon (essentially) proved:

THEOREM 3.7. *(See [MW].) Conjecture* $6'$ *is true if* H *is semisolvable and if* $\operatorname{ch} k \nmid n$. *In particular, a semisolvable, semisimple and cosemisimple Hopf algebra is of Frobenius type.*

A Hopf algebra H $(\neq k)$ is said to be *simple* if it includes no normal Hopf subalgebra other than k, H.

THEOREM 3.8. *(See [N4].) Suppose* $\operatorname{ch} k = 0$. *Among all semisimple Hopf algebras of dimension* < 60, *there is only one* (*up to isomorphism*) *that is neither upper- nor lower-semisolvable. This unique Hopf algebra is a selfdual, simple Hopf algebra of dimension* 36 *which is a twist* (*Definition* 4.4) *of the group algebra* $k(D_6 \times D_6)$, *where* D_6 *denotes the dihedral group of order* 6; *see* [GN].

REMARK 3.9. According to Hoffman [H], for every finite simple group G, there is a non-trivial simple Hopf algebra as a twist of kG. But, apart from twists or pseudo-twists (Remark 4.6) of group algebras and their duals, there seems to be no example of a simple, semisimple and cosemisimple Hopf algebra (over an algebraically closed field), or even of a semisimple and cosemisimple Hopf algebra which is not semisolvable.

Let us consider semisimple and cosemisimple Hopf algebras of dimension p^e, where p is a prime. These have a similar property as p-groups, as shown below.

[1] A Hopf subalgebra K of a Hopf algebra H is *normal* if $K^+H = HK^+$.

THEOREM 3.10. *(See [M3,EG2].) Every semisimple and cosemisimple Hopf algebra H of dimension p^e, where p is a prime and $e > 0$, contains a non-trivial central grouplike element.*

By applying the result to H^*, we have a sequence of Hopf subalgebras $H = H_e \supset H_{e-1} \supset \cdots \supset H_0 = k$ which gives rise to extensions

$$H_i \rightarrowtail H_{i+1} \twoheadrightarrow kC_p \quad (0 \leqslant i < e) \tag{3.11}$$

that are *cocentral* (i.e., $k^{C_p} \subset H_{i+1}^*$ is central). In particular, H is lower- and upper-semisolvable. Here C_n denotes the cyclic group of order n.

Classifying all those H of dimension p^e would be possible, in principle, by classifying the extensions (3.11) inductively. For $e = 2$, this was done in [M2]; see also Theorem 2.4(1). For $e = 3$, we need to classify the extensions $k^G \rightarrowtail H \twoheadrightarrow kC_p$, where $G = C_{p^2}$ or $C_p \times C_p$. If $G = C_{p^2}$, H is necessarily a group algebra. If $G = C_p \times C_p$, there do exist non-trivial extensions, which can be classified up to equivalence by computing the Opext group explained in Section 5. By classifying the obtained H further into isomorphism classes, we obtain the following.

THEOREM 3.12.

(1) *(See [K1].) There exist only one* (*up to isomorphism*) *semisimple and cosemisimple Hopf algebra of dimension* 8; *it is now famous as the Kac and Paljutkin Hopf algebra* [KP] (*see Section* 7).

(2) *(See [M2].) Let p be an odd prime. There exist exactly $p + 8$ non-trivial semisimple and cosemisimple Hopf algebras of dimension p^3.*

The Hopf algebras obtained above are all selfdual. Those in (2) are not quasitriangular, although *almost cocommutative* [D2] in the sense that their character rings are commutative; see [M9].

As for the case $e = 4$, Kashina [Ks2] classified the semisimple Hopf algebras of dimension 16 in characteristic zero.

Let us summarize Natale's results in dimension pq^2.

THEOREM 3.13. *(See [N1,N2,N3].) Suppose* $\operatorname{ch} k = 0$. *Let p, q be distinct primes. Let H be a semisimple Hopf algebra of dimension pq^2.*

(1) *The following are equivalent*:
  (i) *H is not simple,*
  (ii) *H or H^* contains a non-trivial central grouplike.*

(2) *H necessarily satisfies these equivalent conditions in the following cases*:
  (a) $p = 2$ *or* 3,
  (b) $p^2 < q$,
  (c) $p > q^4$ *and* $p \not\equiv 1 \bmod q$,
  (d) *H and H^* are both of Frobenius type,*
  (e) *H or H^* is of Frobenius type, and $p < q$.*

(3) *If* $\dim H < 100$, (d) *holds.*

Moreover, Natale [N1] classified those semisimple Hopf algebras H of dimension pq^2 which satisfy the equivalent conditions (i), (ii) above, and especially those of dimension $pq^2 < 100$. It was recently shown by Galindo and Natale [GN] that a direct analogue of the Burnside $p^a q^b$ theorem for finite groups does not hold for semisimple Hopf algebras.

4. Triangular semisimple Hopf algebras

Recall from Section 3 the construction of a crossed product $R \rtimes_\sigma H$ of a Hopf algebra H with an algebra R, but in the special case when $R = k$. Then $\rightharpoonup$ must equal $\varepsilon : H \to k$, and (3.1c) turns into

$$\sum \sigma(b_1, c_1)\sigma(a, b_2 c_2) = \sum \sigma(a_1, b_1)\sigma(a_2 b_2, c). \tag{4.1}$$

DEFINITION 4.2. A linear map $\sigma : H \otimes H \to k$ which satisfies (3.1a), (3.1d) and (4.1) is called a *cocycle* on $H \otimes H$.

An equivalence relation $\sim$ among cocycles on $H \otimes H$ is defined so that $\sigma \sim \sigma'$, if and only if

$$\sum \sigma(a_1, b_1)\gamma(a_2 b_2) = \sum \gamma(a_1)\gamma(b_1)\sigma'(a_2, b_2) \quad (a, b \in H)$$

for some invertible $\gamma : H \to k$; see [Doi]. We will write simply ${}_\sigma H$ for $k \rtimes_\sigma H$.

A *(right) H-Galois object* is a right H-comodule algebra $A \neq 0$ such that $A \otimes A \to A \otimes H$, $a \otimes b \mapsto a\rho(b)$ is bijective, where $\rho : A \to A \otimes H$ denotes the H-comodule structure. Such an ${}_\sigma H$ as above is an H-Galois object. Moreover, $\sigma \mapsto {}_\sigma H$ induces an injection

$$\{\text{cocycles on } H \otimes H\}/\!\sim\ \rightarrowtail \{H\text{-Galois objects}\}/\simeq.$$

This is bijective if $\dim H < \infty$; see [DT] (or [Mo1, p. 129]).

Suppose $\dim H < \infty$, and let $J := H^*$.

DEFINITION 4.3. An element σ in $J \otimes J$ is called a *dual cocycle* in $J \otimes J$, if it is, regarded as a linear map $H \otimes H \to k$, a cocycle on $H \otimes H$.

For such a σ as above, let J^σ denote the algebra J which is endowed with the twisted coalgebra structure Δ^σ, ε^σ defined by

$$\Delta^\sigma(x) = \sigma\Delta(x)\sigma^{-1}, \qquad \varepsilon^\sigma(x) = \varepsilon(x) \quad (x \in J),$$

where Δ, ε on the right-hand side denote the coalgebra structure of J. Then, J^σ is indeed a Hopf algebra.

DEFINITION 4.4. J^σ is called the *twist* of J by σ.

Regard a (left) J-module V as a J^σ-module in the obvious way, and denote it by $\bar{V}$. Then, $V \mapsto \bar{V}$ gives a tensor-equivalence from the category J-Mod of J-modules to the category J^σ-Mod of J^σ-modules, with respect to the tensor-structure

$$\bar{V} \otimes \bar{W} \xrightarrow{\simeq} \overline{V \otimes W}, \quad v \otimes w \mapsto \sigma^{-1}(v \otimes w). \tag{4.5}$$

Every Hopf algebra K such that K-Mod is tensor-equivalent to J-Mod is isomorphic to some twist J^σ, and every tensor-equivalence is given as above. See [D1,Sb1].

REMARK 4.6. (See [Nk].) To make $J^\sigma = (J^\sigma, \Delta^\sigma, \varepsilon^\sigma)$ into a bialgebra, we have only to require that the linear map $\sigma : H \otimes H \to k$ satisfies the following, instead of (4.1): there exists an invertible $\phi : H \otimes H \otimes H \to k$ such that

$$\sum \sigma(b_1, c_1)\sigma(a, b_2c_2) = \sum \sigma(a_1, b_1)\sigma(a_2b_2, c)\phi(a_3, b_3, c_3),$$
$$\sum \phi(a_1, b_1, c_1)a_2b_2c_2 = \sum a_1b_1c_1\phi(a_2, b_2, c_2),$$

for all $a, b, c \in H$. We call such a σ ($\in J \otimes J$) a *pseudo-dual cocycle*, and J^σ the *pseudo-twist* of J by σ. Two finite-dimensional semisimple bialgebras J, J' over an algebraically closed field are pseudo-twists of each other if and only if there exists an algebra isomorphism $J \xrightarrow{\simeq} J'$ that induces an isomorphism of character rings.

To survey deeper results in the special case when $H = k^G$, we impose:

ASSUMPTION 4.7. *k is algebraically closed, and G denotes a finite group such that*

$$\operatorname{ch} k \nmid |G|.$$

We will reformulate things concerning k^G-comodule algebras (respectively, k^G-Galois objects) in terms of *G-algebras* (respectively, *Galois G-algebras*). A *G-algebra* is an algebra A on which G acts via algebra automorphisms. Then there arises the semi-direct product algebra $A \rtimes G$. A finite-dimensional G-algebra $A \neq 0$ is said to be *Galois*, if the canonical algebra map $A \rtimes G \to \operatorname{End} A$ is bijective, or equivalently, under the assumption above, if A is semisimple, $|G| = \dim A$ and the subalgebra A^G of G-invariants in A equals k.

Let A be a Galois G-algebra. Since G acts transitively on the simple components of A, their stabilizers are mutually conjugate subgroups. Fix a simple component B, say, in A, and let Γ denote its stabilizer subgroup. Let $B = M_n(k)$.

PROPOSITION 4.8. *(See [Mv].)*

(1) *B is a Galois Γ-algebra, whence $|\Gamma| = n^2$. The Γ-action $\rightharpoonup$ on $B = M_n(k)$ arises uniquely from an irreducible projective representation $\pi : \Gamma \to PGL_n(k)$ so that*

$$x \rightharpoonup b = u_x b u_x^{-1} \quad (x \in \Gamma,\ b \in B),$$

where u_x is such an element in $GL_n(k)$ that is mapped to $\pi(x)$ in $PGL_n(k)$.

(2) *A is isomorphic to the Γ-algebra* $\operatorname{Map}_\Gamma(G, B)$ *consisting of all left Γ-maps $G \to B$, where the Γ-action is induced from the right Γ multiplication G.*

Note that the map $\Gamma \to GL_n(k)$, $x \mapsto u_x$ induces an algebra isomorphism $k_\alpha \Gamma \xrightarrow{\simeq} M_n(k) = B$, where $k_\alpha \Gamma$ denotes the twisted product group algebra given by an appropriate 2-cocycle $\alpha : \Gamma \times \Gamma \to k^\times$. The induced Γ-action on $k_\alpha \Gamma$ is the Miyashita action [Mi].

PROPOSITION 4.9. *(See [Mv].) Conversely, a pair (Γ, π) of a subgroup $\Gamma \subset G$ and an irreducible projective representation $\pi : \Gamma \to PGL_n(k)$ of degree $n = |\Gamma|^{1/2}$ gives rise to a Galois G-algebra* $\mathrm{Map}_\Gamma(G, M_n(k))$. *Every Galois G-algebra comes from such a pair (Γ, π) as above*; *the pair (Γ, π) is unique up to conjugacy of subgroups and equivalence of projective representations.*

EXAMPLE 4.10.

(1) If $|G|$ is square-free, there exists no Galois G-algebra other than the trivial one $k^G = \mathrm{Map}(G, k)$.

(2) [M5] Suppose

$$G = D_{2n} = \langle a, x \mid a^2 = 1 = x^n,\ ax = x^{-1}a \rangle, \tag{4.11}$$

the dihedral group of order $2n$ ($\geqslant 4$). Non-trivial G-Galois algebras possibly arise only from the following Abelian subgroups of type $(2, 2)$, which are not conjugate to each other:

$$\Gamma_1 = \langle a, x^{n/2} \rangle \quad \text{in case } 2 \mid n,$$
$$\Gamma_2 = \langle ax, x^{n/4} \rangle \quad \text{in case } 4 \mid n,$$

Note that the Abelian group of type $(2, 2)$ has a unique (up to equivalence) irreducible projective representation of degree 2, say ϖ. Therefore there exist no (respectively, only one; respectively, exactly two) non-trivial D_{2n}-Galois algebras, if n is odd (respectively, n is even, but $4 \nmid n$; respectively, $4 \mid n$).

Let A be a Galois G-algebra, and suppose that it arises from a pair (Γ, π). The Galois Γ-algebra (or the k^Γ-Galois object) associated to the pair gives rise to a unique (up to equivalence) dual cocycle σ in $k\Gamma \otimes k\Gamma$, whence in $kG \otimes kG$. Suppose $J = kG$ in (4.5). Let z be a central element in G of order $\leqslant 2$. Define

$$R_z := \frac{1}{2}(1 \otimes 1 + 1 \otimes z + z \otimes 1 - z \otimes z) \quad (\in kG \otimes kG),$$

where we understand R_z to be $1 \otimes 1$ if $z = 1$. Then, $v \otimes w \mapsto R_z(w \otimes v)$ gives a symmetry (i.e., an involutory braiding) in kG-Mod, which coincides with the symmetry of super-vector spaces if $z \neq 1$. It transforms through (4.5) to $v \otimes w \mapsto R_z \sigma \sigma_{21}^{-1}(w \otimes v)$ in $(kG)^\sigma$-Mod, where σ_{21} denotes the image of σ under the flip $a \otimes b \mapsto b \otimes a$. This means that $R_z \sigma_{21} \sigma^{-1}$ is a triangular structure on $(kG)^\sigma$.

THEOREM 4.12. *(See [EG5].) The triangular Hopf algebra $((kG)^\sigma, R_z \sigma_{21} \sigma^{-1})$ is semisimple and cosemisimple. Conversely, every triangular, semisimple and cosemisimple Hopf algebra (H, R) of finite dimension arises from a quadruple (G, Γ, π, z) as above; the quadruple is unique up to isomorphism.* (*An isomorphism between such quadruples is defined in the obvious way.*)

We leave it as an exercise to describe the triangular Hopf algebras which arise from the quadruples $(D_{2n}, \Gamma_i, \varpi, 1)$ given in Example 4.10(2); see Example 5.7.

For the further extensive theory of (co)triangular Hopf algebras due to Etingof and Gelaki, see first of all [EG7].

5. Hopf algebra extensions of kF by k^G

Let F, G denote groups, which will be supposed to be finite in most parts below.

DEFINITION 5.1. (See [T1].) A pair of groups (F, G) together with two maps

$$G \xleftarrow{\triangleleft} G \times F \xrightarrow{\triangleright} F$$

is called a *matched pair*, if

(a) $1 \triangleright a = a, xy \triangleright a = x \triangleright (y \triangleright a)$,
(b) $x \triangleleft 1 = x, x \triangleleft ab = (x \triangleleft a) \triangleleft b$,
(c) $x \triangleright ab = (x \triangleright a)((x \triangleleft a) \triangleright b)$, and
(d) $xy \triangleleft a = (x \triangleleft (y \triangleright a))(y \triangleleft a)$,
where $a, b \in F, x, y \in G$.[2]

PROPOSITION 5.2. *(See [T1].) If $(F, G, \triangleleft, \triangleright)$ is matched pair, the Cartesian product $F \times G$ endowed with the binary operation*

$$(a, x)(b, y) = \big(a(x \triangleright b), (x \triangleleft b)y\big)$$

forms a group with unit $(1, 1)$.

This group is denoted by $F \bowtie G$. If $\triangleleft =$ triv, the trivial action (i.e., $x \triangleleft a = x$ for all $x \in G, a \in F$), then $\triangleright$ must be an action by group automorphisms, and the associated semi-direct product $F \rtimes G$ coincides with $F \bowtie G$.

PROPOSITION 5.3. *If F and G are embedded in a group Σ so that Σ factorizes in F and G in the sense that the canonical map $\mu : F \times G \to \Sigma$, $(a, x) \mapsto ax$ is bijective. Then the pair (F, G) together with the actions $\triangleleft, \triangleright$ determined by*

$$xa = (x \triangleright a)(x \triangleleft a) \quad \text{in } \Sigma,$$

where $a \in F, x \in G$, forms a matched pair, and $\mu : F \bowtie G \xrightarrow{\simeq} \Sigma$ turns into a group isomorphism.

Thus every matched-pair structure on (F, G) arises uniquely (up to isomorphism) from a factorization $F \times G \simeq \Sigma$ of some group Σ.

[2] Here $x \triangleleft a$ is the image in G of $(x, a) \in G \times F$ under $\triangleleft$ and $y \triangleright b$ is the image in F of $(y, b) \in G \times F$ under $\triangleright$.

EXAMPLE 5.4. Let $n \geqslant 3$ be an integer. The symmetric group S_n on n letters $1, 2, \ldots, n$ factorizes into

$$C_n = \{c^i \mid 0 \leqslant i < n\}, \qquad S_{n-1} = \{x \in S_n \mid x(n) = n\},$$

where $c = (1\ 2\ \ldots\ n)$, a cyclic permutation, so that we have a matched pair $(C_n, S_{n-1}, \triangleleft, \triangleright)$. Since $c^{-x(i)}xc^i$ fixes n, we see

$$x \triangleright c^i = c^{x(i)} \quad (0 \leqslant i < n),$$

while the other $\triangleleft$, non-trivial if $n > 3$, is difficult to describe explicitly.

PROPOSITION 5.5. *(See [T1].) Given a matched pair $(F, G, \triangleleft, \triangleright)$ of finite groups, the vector space $k^G \otimes kF$ is made into a Hopf algebra, denoted $k^G \blacktriangleright\!\!\triangleleft kF$, which is defined by the obvious unit and counit and the additional structures*

$$(e_x a)(e_y b) = \delta_{x \triangleleft a, y} e_x ab,$$
$$\Delta(e_x a) = \sum_{y \in G} e_{xy^{-1}}(y \triangleright a) \otimes e_y a,$$
$$S(e_x a) = e_{(x \triangleleft a)^{-1}}(x \triangleright a)^{-1},$$

where $a \in F$, $x \in G$, and $e_x a$ stands for $e_x \otimes a$.

This is the bismash product, i.e., the smash-product algebra $k^G \rtimes kF$ and cosmash-product coalgebra $k^G \blacktriangleright\!\!< kF$, which is constructed from the action $\rightharpoonup$ and the coaction ρ given by the formulae (5.9) below.

Here is an interesting results on twists (Definition 4.4) of $k^G \blacktriangleright\!\!\triangleleft kF$.

PROPOSITION 5.6. *(See [LYZ].) Suppose that a finite group Σ factorizes into subgroups in two ways as $F \times G_1 \simeq \Sigma$, $F \times G_2 \simeq \Sigma$, so that there are two matched pairs, (F, G_1), (F, G_2). Then the associated Hopf algebras $k^{G_1} \blacktriangleright\!\!\triangleleft kF$, $k^{G_2} \blacktriangleright\!\!\triangleleft kF$ are twists of each other.*

PROOF. This follows from the following fact: the category $k^G \blacktriangleright\!\!\triangleleft kF$-Mod of $k^G \blacktriangleright\!\!\triangleleft kF$-modules is canonically tensor-equivalent to the tensor category $({}^{\Sigma}_{F}\mathcal{M}_F, \otimes_{kF}, kF)$, where ${}^{\Sigma}_{F}\mathcal{M}_F$ denotes the category of those $\Sigma = F \bowtie G$-graded vector spaces $M = \bigoplus_{s \in \Sigma} M_s$ which are at the same time two-sided kF-modules so that $aM_s b \subset M_{asb}$ $(a, b \in F$, $s \in \Sigma)$. See [T3, pp. 329–330] for more details, and [Sb2,Sb3] as well. □

A natural dual cocycle σ in $(k^{G_1} \blacktriangleright\!\!\triangleleft kF)^{\otimes 2}$ such that $(k^{G_1} \blacktriangleright\!\!\triangleleft kF)^{\sigma} = k^{G_2} \blacktriangleright\!\!\triangleleft kF$ is explicitly given in [T3, (6.23)].

EXAMPLE 5.7. (See [M5].) Suppose that $n = 2m$ $(\geqslant 4)$ is an even integer. The dihedral group D_{2n} as presented by (4.11) factorizes in two ways into

(1) $F = \langle a \rangle \simeq C_2$, $G_1 = \langle x \rangle \simeq C_n$,
(2) $F = \langle a \rangle \simeq C_2$, $G_2 = \langle ax, x^m \rangle \simeq D_{2m}$,

so that $k^{G_1} {\blacktriangleright\!\!\triangleleft} kF$ and $k^{G_2} {\blacktriangleright\!\!\triangleleft} kF$ are twists of each other. We remark that the dual of $k^{G_2} {\blacktriangleright\!\!\triangleleft} kF$ coincides with the Hopf algebra $\mathcal{A}_{4m}$ defined by Definition 7.3(1). If k contains a primitive n-th root of 1, then one sees that $k^{G_1} {\blacktriangleright\!\!\triangleleft} kF \simeq kD_{2n}$. In this case the natural dual cocycle σ as given in [T3, (6.23)] corresponds to the Galois D_{2n}-algebra which arises from the pair (Γ_1, ϖ) given in Example 4.10(2).

Let us review quickly some basic facts on Hopf algebra extensions. Let J, K be finite-dimensional Hopf algebras. By an *extension* of J by K, we mean a short exact sequence $(H) = K \overset{i}{\rightarrowtail} H \overset{p}{\twoheadrightarrow} J$ of finite-dimensional Hopf algebras. This means that

(a) i is injective and p induces an isomorphism $H/i(K^+)H \simeq J$, or equivalently

(b) p is surjective and i induces an isomorphism from K onto $H^{\mathrm{co}J} := \{a \in H \mid (\mathrm{id} \otimes p) \circ \Delta(a) = a \otimes 1\}$.

We remark that H is then semisimple if J and K are. An *equivalence* between extensions (H), (H') is an isomorphism $H \xrightarrow{\simeq} H'$ that induces the identity maps on J, K. By Theorem 1.1, there exists a (possibly unit and counit-preserving) left K-linear and right J-colinear isomorphism

$$H \simeq K \otimes J. \tag{5.8}$$

The Hopf algebra structure transferred via (5.8) onto $K \otimes J$ is the bicrossed product $K^{\tau} {\blacktriangleright\!\!\triangleleft}_{\sigma} J$ constructed from certain data

$$\begin{aligned} &\rightharpoonup : J \otimes K \to K, && \sigma : J \otimes J \to K; \\ &\rho : J \to J \otimes K, && r : J \to K \otimes K. \end{aligned}$$

As an algebra, it the crossed-product algebra $K \rtimes_{\sigma} J$ constructed from $\rightharpoonup, \sigma$ (Definition 3.2), and as a coalgebra, it is the co-crossed-product coalgebra $K^{\tau} {\blacktriangleright\!<} J$ constructed from ρ, τ; see [AD, Theorem 2.20] for the compatibility conditions among the data. Thus, the exact sequence (H) is equivalent to

$$(K^{\tau} {\blacktriangleright\!\!\triangleleft}_{\sigma} J) = K \overset{\otimes 1}{\rightarrowtail} K^{\tau} {\blacktriangleright\!\!\triangleleft}_{\sigma} J \overset{\varepsilon \otimes \mathrm{id}}{\twoheadrightarrow} J.$$

Suppose in particular that $J = kF$, $K = k^G$, where F, G are finite groups. Then for each extension (H), the data $\rightharpoonup$, ρ are independent of the choice of the isomorphism (5.8), and arise uniquely from a matched-pair structure $G \xleftarrow{\triangleleft} G \times F \xrightarrow{\triangleright} F$ on (F, G) so that

$$a \rightharpoonup e_x = e_{x \triangleleft a^{-1}}, \quad \rho(a) = \sum_{y \in G} (y \triangleright a) \otimes e_y, \tag{5.9}$$

where $a \in F$, $x \in G$. Hence we can say that (H) *is associated to* the matched pair $(F, G, \triangleleft, \triangleright)$ thus obtained. Let

$$\mathrm{Opext}\big(kF, k^G, \rightharpoonup, \rho\big) \quad \text{or} \quad \mathrm{Opext}\big(kF, k^G\big)$$

denote the set of equivalence classes of those extensions which are associated to a fixed matched pair $(F, G, \triangleleft, \triangleright)$, where $\rightharpoonup, \rho$ are understood to be what arises from $\triangleleft, \triangleright$ by (5.9). In fact, this set forms an Abelian group with respect to some product that generalizes the Baer product; the unit is the extension $(k^G {\blacktriangleright\!\!\triangleleft} kF)$ given by the bismash product.

REMARK 5.10. Let $(F, G, \triangleleft, \triangleright)$ be a matched pair of finite groups, and suppose that it arises from a factorization $F \times G \simeq \Sigma$ of a group Σ. By applying inverses we have another factorization $G \times F \simeq \Sigma$, which gives rise to another matched pair $(G, F, \triangleleft', \triangleright')$; this matched pair will be said to be *opposite* to the first one. One sees that $(H) \mapsto (H^*)$ gives an isomorphism

$$\operatorname{Opext}\bigl(kF, k^G, \rightharpoonup, \rho\bigr) \simeq \operatorname{Opext}\bigl(kG, k^F, \rightharpoonup', \rho'\bigr)$$

between the associated Opext groups.

6. The Kac exact sequence

Fix a matched pair $(F, G, \triangleleft, \triangleright)$ of finite groups (Definition 5.1), and recall that it gives rise to a group, $F \bowtie G$, which encompasses F, G. We will reproduce from [M8] a double complex $E^{\cdot\cdot}$ for computing the Opext group associated to the fixed matched pair; the original construction, slightly different from ours, is due to Kac [K1]. For a group Γ, a *Γ-module* means a left module over the integral group ring $\mathbb{Z}\Gamma$. First of all, let us have a double complex of $F \bowtie G$-modules,

$$\begin{tikzcd}
 & \vdots \arrow[d] & \vdots \arrow[d] & \\
C_{\cdot\cdot} = & C_{01} \arrow[d, "\partial'"] & C_{11} \arrow[l, "\partial"'] \arrow[d, "\partial'"] & \cdots \arrow[l] \\
 & C_{00} & C_{10} \arrow[l, "\partial"'] & \cdots \arrow[l]
\end{tikzcd}$$

in which C_{pq} is the free $F \bowtie G$-module on the set $G^q \times F^p$, and ∂, ∂' are defined by

$$\begin{aligned}
&(-1)^q \partial(x_1, \dots, x_q; a_1, \dots, a_p) \\
&\quad = (x_1 \cdots x_q \triangleright a_1)\bigl(x_1 \triangleleft (x_2 \cdots x_q \triangleright a_1), \dots, \\
&\qquad x_{q-1} \triangleleft (x_q \triangleright a_1), x_q \triangleleft a_1; a_2, \dots, a_p\bigr) \\
&\qquad + \sum_{i=1}^{p-1} (-1)^i (x_1, \dots, x_q; a_1, \dots, a_i a_{i+1}, \dots, a_p) \\
&\qquad + (-1)^p (x_1, \dots, x_q; a_1, \dots, a_{p-1}), \\
&\partial'(x_1, \dots, x_q; a_1, \dots, a_p) \\
&\quad = x_1(x_2, \dots, x_q; a_1, \dots, a_p) \\
&\qquad + \sum_{i=1}^{q-1} (-1)^i (x_1, \dots, x_i x_{i+1}, \dots, x_q; a_1, \dots, a_p) \\
&\qquad + (-1)^q \bigl(x_1, \dots, x_{q-1}; x_q \triangleright a_1, (x_q \triangleleft a_1) \triangleright a_2, \dots, \\
&\qquad (x_q \triangleleft a_1 \cdots a_{p-1}) \triangleright a_p\bigr),
\end{aligned}$$

where $a_i \in F$, $x_i \in G$. The lowest horizontal complex $C_{\cdot 0}$, and the leftmost vertical complex $C_{\cdot 0}$ are the standard resolutions of $\mathbb{Z}$ over F and G, respectively. Each row or column in $C_{\cdot\cdot}$ is exact, whence the total complex $\mathrm{Tot}\, C_{\cdot\cdot}$ gives a non-standard free $F \bowtie G$-resolution of $\mathbb{Z}$. Regard $k^\times = k \setminus 0$ as a trivial $F \bowtie G$-module, and apply $\mathrm{Hom}_{F\bowtie G}(\;, k^\times)$ to $C_{\cdot\cdot}$. We then obtain a double cochain complex, $\mathrm{Hom}_{F\bowtie G}(C_{\cdot\cdot}, k^\times)$. Normalize this just for convenience, and let $D^{\cdot\cdot}$ denote the thus obtained, normalized complex; see below. Thus, D^{pq} equals, under a natural identification, the multiplicative group $\mathrm{Map}_+(G^q \times F^p, k^\times)$ consisting of the maps $f : G^q \times F^p \to k^\times$ such that $f(x_1 \dots, x_q; a_1, \dots, a_p) = 1$ whenever any a_i or $x_i = 1$. The lowest horizontal (respectively, leftmost vertical) complex in $D^{\cdot\cdot}$ is the standard complex for computing the group cohomology $H^\cdot(F, k^\times)$ (respectively, $H^\cdot(G, k^\times)$). Remove these two complexes from $D^{\cdot\cdot}$, and let $E^{\cdot\cdot}$ denote the obtained one; see below. But, we count the total dimensions in $E^{\cdot\cdot}$ so that $\mathrm{Tot}^n\, E^{\cdot\cdot}$ equals the direct product of those $\mathrm{Map}_+(G^q \times F^p, k^\times)$ in which $p+q = n+1$, $p > 0, q > 0$.

$$D^{\cdot\cdot} = \begin{tikzcd}
\vdots & E^{\cdot\cdot} = \quad \vdots & \vdots & \\
\mathrm{Map}_+(G^2, k^\times) \arrow[u] \arrow[r] & \mathrm{Map}_+(G^2 \times F, k^\times) \arrow[u] \arrow[r] & \mathrm{Map}_+(G^2 \times F^2, k^\times) \arrow[u] \arrow[r] & \cdots \\
\mathrm{Map}_+(G, k^\times) \arrow[u] \arrow[r] & \mathrm{Map}_+(G \times F, k^\times) \arrow[u] \arrow[r] & \mathrm{Map}_+(G \times F^2, k^\times) \arrow[u] \arrow[r] & \cdots \\
k^\times \arrow[u] \arrow[r] & \mathrm{Map}_+(F, k^\times) \arrow[u] \arrow[r] & \mathrm{Map}_+(F^2, k^\times) \arrow[u] \arrow[r] & \cdots
\end{tikzcd}$$

Proposition 6.1.

(1) *Given a total 2-cocycle (σ, τ) in $E^{\cdot\cdot}$, where $\sigma \in \mathrm{Map}_+(G \times F^2, k^\times)$, $\tau \in \mathrm{Map}_+(G^2 \times F, k^\times)$, the vector space $k^G \otimes kF$ is made into a Hopf algebra, denoted by $k^{G\tau} \blacktriangleright\!\!\triangleleft_\sigma kF$, which is defined by the obvious unit and counit, and the additional structures*

$$(e_x a)(e_y b) = \delta_{x \triangleleft a, y}\, \sigma(x; a, b) e_x ab,$$

$$\Delta(e_x a) = \sum_{y \in G} \tau\big(xy^{-1}, y; a\big) e_{xy^{-1}}(y \triangleright a) \otimes e_y a,$$

$$S(e_x a) = \tau\big(x^{-1}, x; a\big)^{-1} e_{(x \triangleleft a)^{-1}}(x \triangleright a)^{-1},$$

where $a, b \in F$, $x, y \in G$. This forms in the obvious way an extension $(k^{G\tau} \blacktriangleright\!\!\triangleleft_\sigma kF)$ of kF by k^G which is associated to the fixed matched pair.

(2) $(\sigma, \tau) \mapsto (k^{G\tau} \blacktriangleright\!\!\triangleleft_\sigma kF)$ *induces an isomorphism* $H^2(\mathrm{Tot}\, E^{\cdot\cdot}) \xrightarrow{\simeq} \mathrm{Opext}(kF, k^G, \rightharpoonup, \rho)$.

Note that if σ, τ take the value 1 constantly, $k^{G\tau}\blacktriangleright\!\!\triangleleft_{\sigma} kF$ coincides with $k^{G}\blacktriangleright\!\!\triangleleft kF$ as given by Proposition 5.5.

Let $\mathrm{Aut}(k^{G}\blacktriangleright\!\!\triangleleft kF)$ denote the group of auto-equivalences of the neutral extension $(k^{G}\blacktriangleright\!\!\triangleleft kF)$.

PROPOSITION 6.2.

(1) *Given a total* 1*-cocycle* ν *in* $E^{\cdot\cdot}$, *we have an auto-equivalence* $f_\nu : e_x a \mapsto \nu(x;a)e_a$ *of* $(k^{G}\blacktriangleright\!\!\triangleleft kF)$.

(2) $\nu \mapsto f_\nu$ *gives an isomorphism* $H^1(\mathrm{Tot}\,E^{\cdot\cdot}) \xrightarrow{\simeq} \mathrm{Aut}(k^{G}\blacktriangleright\!\!\triangleleft kF)$.

THEOREM 6.3 *(Kac exact sequence). We have an exact sequence*,

$$\begin{aligned}0 &\to H^1(F\bowtie G,k^\times)\to H^1(F,k^\times)\oplus H^1(G,k^\times)\to \mathrm{Aut}\big(k^{G}\blacktriangleright\!\!\triangleleft kF\big)\\ &\to H^2(F\bowtie G,k^\times)\to H^2(F,k^\times)\oplus H^2(G,k^\times)\to \mathrm{Opext}(kF,k^G)\\ &\to H^3(F\bowtie G,k^\times)\to H^3(F,k^\times)\oplus H^3(G,k^\times).\end{aligned}$$

PROOF. $E^{\cdot\cdot}$ is regarded as a double subcomplex of $D^{\cdot\cdot}$. Their total complexes make a short exact sequence of complexes. The associated long exact cohomology sequence, combined with the isomorphisms from Propositions 6.1, 6.2, gives the Kac exact sequence. □

REMARK 6.4.

(1) In $E^{\cdot\cdot}$, we can replace the coefficients $k^\times$ with any trivial (or even non-trivial) $F\bowtie G$-module M. The resulting double complex will be denoted by $E^{\cdot\cdot}(M)$. As above, there is a long exact cohomology sequence,

$$\begin{aligned}\cdots &\to H^n(F\bowtie G,M)\to H^n(F,M)\oplus H^n(G,M)\to H^n\big(\mathrm{Tot}\,E^{\cdot\cdot}(M)\big)\\ &\to H^{n+1}(F\bowtie G,M)\to\cdots.\end{aligned}\tag{6.5}$$

(2) Suppose that $\alpha : G\times G\to k^\times$ is a 2-cocycle, and (H) represents an equivalence class in $\mathrm{Opext}(kF,k^G)$. Let $\delta : H^2(G,k^\times)\to \mathrm{Opext}(kF,k^G)$ denote the homomorphism which is induced from the horizontal differential in $D^{\cdot\cdot}$; it appears in the Kac exact sequence. Then, $(H)\cdot\delta\alpha^{-1}$ is equivalent to the extension (H^α) which is given by the twist H^α of H by the dual cocycle α in $k^G\otimes k^G$; see [M7, Proposition 3.1].

(3) Schauenburg [Sb2,Sb3] gives a categorical interpretation of the Kac exact sequence in a quite generalized context.

COROLLARY 6.6. *Suppose that* k *is algebraically closed.*

(1) *(See [M4].) The extensions of* kF *by* k^G *are finitely many up to equivalence.*

(2) *Every extension of* kF *by* k^G *is equivalent to some extension that is defined over the ring* $\mathbb{Z}[\zeta]$ *(*$\subset\mathbb{C}$*), where* ζ *is some root of* 1 *whose order does not divide* $\mathrm{ch}\,k$.

PROOF. (1) It suffices to prove that every Opext group $\mathrm{Opext}(kF,k^G)$ is finite. The assumption implies that $k^\times$ is a divisible $\mathbb{Z}$-module. By the universal coefficient theorem, we see

$$H^n(\Gamma,k^\times)\simeq \mathrm{Hom}_{\mathbb{Z}}\big(H_n(\Gamma,\mathbb{Z}),k^\times\big)\quad (n\geqslant 0)$$

for any finite group Γ, whence $H^n(\Gamma, k^\times)$ is finite if $n > 0$. The Kac exact sequence now proves the desired result.

(2) Argue as above, but replacing $k^\times$ with the group $\mu(k)$ of all roots of 1 in k. Then one sees $H^n(\Gamma, \mu(k)) \simeq H^n(\Gamma, k^\times)$ $(n \geqslant 0)$. It follows from the exact sequences (6.5) with $M = \mu(k)$ and $k^\times$ that $H^2(\operatorname{Tot} E^{\cdot\cdot}(\mu(k))) \simeq \operatorname{Opext}(kF, k^G)$. This implies the claim above. □

REMARK 6.7. (See [M4].) For arbitrary k, the group $\operatorname{Aut}(k^G {\blacktriangleright\!\!\lhd} kF)$ is finite. Moreover, if $\operatorname{Aut}(\mathbb{C}^G {\blacktriangleright\!\!\lhd} \mathbb{C}F) \simeq \mathbb{Z}/(n_1) \oplus \cdots \oplus \mathbb{Z}/(n_r)$ $(n_i > 0)$, then

$$\operatorname{Aut}(k^G {\blacktriangleright\!\!\lhd} kF) \simeq \mu_{n_1}(k) \times \cdots \times \mu_{n_r}(k),$$

where $\mu_n(k)$ denotes the group of nth root of 1 in k. On the other hand, if $\operatorname{Aut}(\mathbb{C}^G {\blacktriangleright\!\!\lhd} \mathbb{C}F)$ is non-trivial, the group $\operatorname{Opext}(\mathbb{Q}F, \mathbb{Q}^G)$ is necessarily infinite.

EXAMPLE 6.8. (See [M4].) For the matched pair $(C_n, S_{n-1}, \lhd, \rhd)$ given by Example 5.4,

$$\operatorname{Opext}\bigl(kC_n, k^{S_{n-1}}, \rightharpoonup, \rho\bigr) \simeq \begin{cases} k^\times/(k^\times)^n & (n > 4), \\ k^\times/(k^\times)^8 & (n = 4), \end{cases}$$

$$\operatorname{Aut}\bigl(k^{S_{n-1}} {\blacktriangleright\!\!\lhd} kC_n\bigr) \simeq \begin{cases} \mu_n(k) & (n > 4), \\ \mu_8(k) & (n = 4). \end{cases}$$

COROLLARY 6.9. *(See [M4].) Suppose $k = \mathbb{C}$. Every extension of $\mathbb{C}F$ by $\mathbb{C}^G$ is equivalent to some extension (H) that is given by a Kac algebra H (Definition 7.2).*

PROOF. A Hopf algebra $\mathbb{C}^{G\tau} {\blacktriangleright\!\!\lhd}_\sigma \mathbb{C}F$ of the bicrossed product is a Kac algebra if and only if σ, τ take their values in $\mathbb{T} := \{x \in \mathbb{C} \mid |x| = 1\}$. This implies the corollary since, as in the last proof, that $H^2(\operatorname{Tot} E^{\cdot\cdot}(\mathbb{T})) \simeq \operatorname{Opext}(\mathbb{C}F, \mathbb{C}^G)$. □

7. Kac and Paljutkin's example and more

In what follows we will discuss Kac algebras, assuming

ASSUMPTION 7.1. $k = \mathbb{C}$.

A finite-dimensional C^*-algebra is a $*$-algebra over $\mathbb{C}$ that is isomorphic to a finite-dimensional semisimple algebra $M_{n_1}(\mathbb{C}) \times \cdots \times M_{n_r}(\mathbb{C})$; this last one has the obvious $*$-structure. Among Kac algebras we restrict our attention only to the finite-dimensional ones.

DEFINITION 7.2. (See [KP].) By a *Kac algebra* we mean a finite-dimensional C^*-algebra H with $*$-algebra maps $\Delta : H \to H \otimes H$, $\varepsilon : H \to \mathbb{C}$ such that (H, Δ, ε) is a Hopf algebra. The antipode S necessarily commutes with $*$.

A finite group algebra $\mathbb{C}G$ (respectively, its dual $\mathbb{C}^G$) turns into a Kac algebra in the unique way when one takes $g^* = g^{-1}$ (respectively, $(e_g)^* = e_g$), where $g \in G$.

To give examples of non-trivial Kac algebras, let $K = \mathbb{C}C_2$ denote the group algebra of the cyclic group $C_2 = \langle a \mid a^2 = 1 \rangle$. It is spanned by the primitive idempotents $e_0 = \frac{1}{2}(1+a)$, $e_1 = \frac{1}{2}(1-a)$.

DEFINITION 7.3. (See [M5].) Let $m \geqslant 2$ be an integer.

(1) $\mathcal{A}_{4m}$ is the Hopf algebra including K as a central Hopf subalgebra which is generated by two elements s_+, s_- over K, and is defined by the relations

$$s_\pm^2 = 1, \qquad (s_+s_-)^m = 1 \tag{7.4}$$

and the structures

$$\Delta(s_\pm) = s_\pm \otimes e_0 s_\pm + s_\mp \otimes e_1 s_\pm, \qquad \varepsilon(s_\pm) = 1,$$
$$S(s_\pm) = e_0 s_\pm + e_1 s_\mp.$$

(2) $\mathcal{B}_{4m}$ is the Hopf algebra defined in the same way as above except that the relation $(s_+s_-)^m = 1$ is replaced by

$$(s_+s_-)^m = a.$$

Notice that these Hopf algebras are defined over any field of characteristic $\neq 2$, and their dimensions equal $4m$. Over $\mathbb{C}$, they are indeed Kac algebras with respect to the unique $*$-structure determined by $s_\pm^* = s_\pm$. Let $(D_{2m}, C_2, \mathrm{triv}, \triangleright)$ be the matched pair opposite (Remark 5.10) to the one given by Example 5.7(2); thus, $D_{2m} = \langle s_+, s_- \mid (7.4) \rangle$, $a \triangleright s_\pm = s_\mp$. One sees that $\mathcal{A}_{4m}$ is the bismash product $\mathbb{C}^{C_2} \blacktriangleright\!\!\triangleleft \mathbb{C}D_{2m}$ associated to the matched pair above, while $\mathcal{B}_{4m}$ is a bicrossed product $\mathbb{C}^{C_2}{}^{\tau}\blacktriangleright\!\!\triangleleft_{\sigma} \mathbb{C}D_{2m}$ with τ trivial. It is proved in [M5, Proposition 3.11] that the Opext group $\mathrm{Opext}(\mathbb{C}D_{2m}, \mathbb{C}^{C_2}, \mathrm{triv}, \rho)$ associated to the matched pair consists of the two extensions $(\mathcal{A}_{4m})$, $(\mathcal{B}_{4m})$. When $m = 2$, $\mathcal{B}_8$ coincides with what the famous example given by Kac and Paljutkin [KP]. $\mathcal{B}_{4m}$ coincides with a quasitriangular Hopf algebra that was previously constructed by Suzuki [Su]. Aside from the examples by Sekine [Se], the $\mathcal{B}_{4m}$ $(m \geqslant 2)$ are another series of Kac algebras containing Kac and Paljutkin's. They posses the following interesting properties.

PROPOSITION 7.5.

(1) *(See [CDMM].) $\mathcal{B}_{4m}$ is selfdual.*

(2) *(See [Su].) $\mathcal{B}_{4m}$ has exactly $2m$ quasitriangular structure, none of which is triangular.*

(3) *(See [M5].) There exists no $\mathcal{B}_{4m}$-Galois object (see Section 4) other than the trivial $\mathcal{B}_{4m}$; this implies by [Sb1] that there exist no Hopf algebras other than $\mathcal{B}_{4m}$ itself whose (co)module category is tensor-equivalent to that of $\mathcal{B}_{4m}$.*

8. Kac algebras and depth 2 subfactors

We keep Assumption 7.1. We will use the basic terminology concerning subfactors, such as used in [IK, Chapter 1]. One needs to know at least that a *factor* is a sort of central

$*$-algebras over $\mathbb{C}$. The notation in this section might feel slightly different from what was used in the preceding sections, because we will follow to some extent the customs of operator-algebraists.

The correspondence between depth 2 subfactors and Kac algebras, shown in Theorem 8.2 below, was first formulated without proof by Ocneanu [Oc], and then proved by Popa [P1,P2]; see also [Sz,Da]. From several, though equivalent, descriptions of the correspondence cited above, we choose the following one due to Kadison and Nikshych [KN].

PROPOSITION 8.1. *Let $\mathcal{P} \supset \mathcal{Q}$ be an irreducible depth 2 inclusion of II_1 factors with finite index. Let $H := (\mathcal{P} \otimes_{\mathcal{Q}} \mathcal{P})^{\mathcal{Q}}$ denote the $\mathcal{Q}$-centralizers in $\mathcal{P} \otimes_{\mathcal{Q}} \mathcal{P}$; it is obviously closed under $*$.*

(1) *The product map $\mu : \mathcal{P} \otimes H \to \mathcal{P} \otimes_{\mathcal{Q}} \mathcal{P}$, $\mu(x \otimes a) = xa$ is bijective.*

(2) *H is a finite-dimensional C^*-algebra with respect to the product*

$$\Bigl(\sum_i x_i \otimes y_i\Bigr)\Bigl(\sum_j z_j \otimes w_j\Bigr) = \sum_{i,j} z_j x_i \otimes y_i w_j.$$

(3) *H has a unique Kac-algebra structure Δ, ε which makes the following diagram commute:*

$$\begin{tikzcd}
\mathcal{P} \otimes H \otimes H \arrow[rr, "\mu^{(2)}", "\simeq"'] & & \mathcal{P} \otimes_{\mathcal{Q}} \mathcal{P} \otimes_{\mathcal{Q}} \mathcal{P} \\
\mathcal{P} \otimes H \arrow[u, "\mathrm{id}\otimes\Delta"] \arrow[rr, "\mu", "\simeq"'] \arrow[dr, "\mathrm{id}\otimes\varepsilon"'] & & \mathcal{P} \otimes_{\mathcal{Q}} \mathcal{P} \arrow[u, "\boldsymbol{\Delta}"'] \arrow[dl, "\boldsymbol{\varepsilon}"'] \\
& \mathcal{P} &
\end{tikzcd}$$

Here, $\boldsymbol{\Delta}(x \otimes y) = x \otimes 1 \otimes y$, $\boldsymbol{\varepsilon}(x \otimes y) = xy$, and $\mu^{(2)}$ denotes the composite

$$\mathcal{P} \otimes H \otimes H \underset{\mu\otimes\mathrm{id}}{\longrightarrow} \mathcal{P} \otimes_{\mathcal{Q}} \mathcal{P} \otimes H \underset{\mathrm{id}\otimes\mu}{\longrightarrow} \mathcal{P} \otimes_{\mathcal{Q}} \mathcal{P} \otimes_{\mathcal{Q}} \mathcal{P}.$$

THEOREM 8.2. *(See [P1,P2].) The correspondence $(\mathcal{P} \supset \mathcal{Q}) \mapsto H$ obtained above gives a bijection from the conjugacy classes of all irreducible depth 2 inclusions $\mathcal{P} \supset \mathcal{Q}$ of hyperfinite* II_1 *factors with finite index onto the isomorphism classes of all Kac algebras H.*

PROPOSITION 8.3. *Let $\mathcal{P} \supset \mathcal{Q}$ be an irreducible inclusion of hyperfinite II_1 factors, and let H be a Kac algebra. Then the following are equivalent.*

(a) *$\mathcal{P} \supset \mathcal{Q}$ is the depth 2 inclusion with finite index which corresponds to H.*

(b) *$\mathcal{P}/\mathcal{Q}$ is an H-Galois extension* [KT]; *this means that there exists a $*$-algebra map $\rho : \mathcal{P} \to \mathcal{P} \otimes H$ for which $\mathcal{P}$ is a right H-comodule, $\mathcal{Q} = \{x \in \mathcal{P} \mid \rho(x) = x \otimes 1\}$ and $\mathcal{P} \otimes_{\mathcal{Q}} \mathcal{P} \to \mathcal{P} \otimes H$, $x \otimes y \mapsto x\rho(y)$ is bijective.*

PROOF. (a) $\Rightarrow$ (b). Define $\rho : \mathcal{P} \to \mathcal{P} \otimes H$ so that its composite $\mu \circ \rho$ with the isomorphism μ given in part (1) of Proposition 8.1 equals $x \mapsto 1 \otimes x$, $\mathcal{P} \to \mathcal{P} \otimes_{\mathcal{Q}} \mathcal{P}$. We then see from part (3) that $\mathcal{P}/\mathcal{Q}$ is H-Galois with respect to ρ.

(b) ⇒ (a). This follows by [KN, Theorem 3.14]. □

Given $\mathcal{P} \supset \mathcal{Q}$ or H as above, it is quite difficult in general to describe explicitly the corresponding object. Following [IK,M8], we will see that such an explicit description is possible in some restricted situation.

Let $\mathcal{R}$ denote a hyperfinite II_1 factor, and let $\mathbb{U} = \{u \in \mathcal{R} \mid uu^* = 1\}$ denote the multiplicative group of unitaries. The latter includes $\mathbb{T} = \{x \in \mathbb{C} \mid |x| = 1\}$ as its center. The group $\mathrm{Aut}(\mathcal{R})$ of $*$-automorphisms of $\mathcal{R}$ includes a normal subgroup, $\mathrm{Inn}(\mathcal{R})$, consisting of all inner automorphisms $\mathrm{inn}(u) : x \mapsto uxu^{-1}$ ($u \in \mathbb{U}$). Define $\mathrm{Out}(\mathcal{R}) = \mathrm{Aut}(\mathcal{R})/\,\mathrm{Inn}(\mathcal{R})$.

Fix a matched pair $(F, G) = (F, G, \lhd, \rhd)$ of finite groups (Definition 5.1). An element (a, x) in the associated group $F \bowtie G$ will be simply denoted ax. Given $\alpha : F \to \mathrm{Aut}(\mathcal{R})$, we will write α_a for $\alpha(a)$.

Definition 8.4. (See [IK,M8].) An *outer action* of the matched pair (F, G) on $\mathcal{R}$ is a pair (α, β) of group homomorphisms $\alpha : F \to \mathrm{Aut}(\mathcal{R})$, $\beta : G \to \mathrm{Aut}(\mathcal{R})$ such that the map $\alpha\beta : F \bowtie G \to \mathrm{Aut}(\mathcal{R})$ defined by

$$(\alpha\beta)_{ax} = \alpha_a\beta_x \quad (a \in F,\ x \in G)$$

gives rise to a group monomorphism $F \bowtie G \rightarrowtail \mathrm{Out}(\mathcal{R})$, composed with the projection $\mathrm{Aut}(\mathcal{R}) \to \mathrm{Out}(\mathcal{R})$. Let $\mathrm{Out}((F, G), \mathcal{R})$ denote the set of all those outer actions.

Let $(\alpha, \beta) \in \mathrm{Out}((F, G), \mathcal{R})$. Since $\alpha\beta : F \bowtie G \to \mathrm{Aut}(\mathcal{R})$ is a group homomorphism modulo $\mathrm{Inn}(\mathcal{R})$, we have a map $\nu : G \times F \to \mathbb{U}$ such that

$$\nu(x; 1) = 1 = \nu(1; a), \qquad \beta_x\alpha_a = \mathrm{inn}\big(\nu(x; a)\big)\alpha_{x\rhd a}\beta_{x\lhd a},$$

where $x \in G$, $a \in F$. We regard $\mathbb{U}$ as if it were an $F \bowtie G$-module, and compute the coboundary $(\sigma, \tau) = (\delta\nu, \delta'\nu)$ of ν in the double complex $E^{\cdot\cdot}(\mathbb{U})$; see Remark 6.4(1). More explicitly we define $\sigma : G \times F^2 \to \mathbb{U}$, $\tau : G^2 \times F \to \mathbb{U}$ by

$$\sigma(x; a, b) = \nu(x; a)^*\nu(x; ab)\alpha_{x\rhd a}\big(\nu(x \lhd a; b)\big)^*,$$
$$\tau(x, y; a) = \nu(x; y \rhd a)\nu(xy; a)^*\beta_x\big(\nu(y; a)\big).$$

Then these have values in $\mathbb{T}$. Moreover, (σ, τ) is a total 2-cocycle in $E^{\cdot\cdot}(\mathbb{T})$, and its cohomology class $[\sigma, \tau]$ in $H^2(\mathrm{Tot}\,E^{\cdot\cdot}(\mathbb{T}))$ is independent of choice of ν. Following [M8], we denote $[\alpha, \beta]$ by $\mathrm{ik}(\alpha, \beta)$, calling it the *Izumi–Kosaki invariant* of (α, β).

The set $\mathrm{Out}((F, G), \mathcal{R})$ has an equivalence relation $\sim_{\mathrm{c}}$, called *cocycle conjugacy*, defined as follows: $(\alpha, \beta) \sim_{\mathrm{c}} (\alpha', \beta')$ if there exist $\theta \in \mathrm{Aut}(\mathcal{R})$ and systems $\{u_a\}_{a\in F}$, $\{v_x\}_{x\in G}$ of unitaries with $u_1 = v_1 = 1$, such that

$$u_{ab} = u_a\alpha_a(u_b), \qquad v_{xy} = v_x\beta_x(v_y),$$
$$\theta\alpha'_a\theta^{-1} = \mathrm{inn}(u_a)\alpha_a, \qquad \theta\beta'_x\theta^{-1} = \mathrm{inn}(v_x)\beta_x,$$

where $a, b \in F$, $x, y \in G$.

Theorem 8.5. *(See [IK,M8].) Let $\mathcal{R}$, $(F, G, \lhd, \rhd)$ be as above.*

(1) ik *induces a bijection*

$$\mathrm{Out}\big((F,G),\mathcal{R}\big)/\sim_{\mathrm{c}} \xrightarrow{\simeq} H^2\big(\mathrm{Tot}\big(E^{\cdot\cdot}(\mathbb{T})\big)\big).$$

(2) *Suppose* $(\alpha,\beta)\in\mathrm{Out}((F,G),\mathcal{R})$, *and* $\mathrm{ik}(\alpha,\beta)=[\sigma,\tau]$; *as is seen in the proof of Corollary* 6.9, *there is a Kac algebra* $\mathbb{C}^{G\tau}{}_{\blacktriangleright\!\triangleleft\sigma}\mathbb{C}F$ *of bicrossed product type. Then*

$$\mathcal{R}\rtimes_\alpha F\supset\mathcal{R}^{(\beta,G)},$$

where $\mathcal{R}\rtimes_\alpha F$ *denotes the semi-direct product associated to* α, *and* $\mathcal{R}^{(\beta,G)}$ *denotes the* G*-invariants with respect to* β, *is the irreducible depth* 2 *inclusion of hyperfinite* II_1 *factors with finite index that corresponds, via the bijection given by Theorem* 8.2, *to the Kac algebra* $\mathbb{C}^{G\tau}{}_{\blacktriangleright\!\triangleleft\sigma}\mathbb{C}F$.

PROOF. (1) See the proof of [M8, Theorem 3.6].

(2) This is proved by using Proposition 8.3, (b) ⇒ (a); see the proof of [M8, Theorem 4.1]. □

REMARK 8.6. Part (1) above is a variant of the famous result by Jones [J] which classifies all outer actions of a given finite group Γ on $\mathcal{R}$, up to conjugacy, by means of $H^3(\Gamma,\mathbb{T})$. In fact there is a close relation, as shown in [M8, p. 618], between Jones' result and ours.

EXAMPLE 8.7.

(1) For every finite group Γ, there is a unique (up to conjugacy) group homomorphism $\Gamma\to\mathrm{Aut}(\mathcal{R})$ that is outer, i.e., is monic, even composed with $\mathrm{Aut}(\mathcal{R})\to\mathrm{Out}(\mathcal{R})$. Therefore, given a matched pair $(F,G,\triangleleft,\triangleright)$, there is a unique (up to cocycle conjugacy) outer action (α,β) on $\mathcal{R}$ such that $\alpha\beta$ is a group homomorphism. For such an (α,β), the inclusion $\mathcal{R}\rtimes_\alpha F\supset\mathcal{R}^{(\beta,G)}$ corresponds to the bismash product $\mathbb{C}^G{}_{\blacktriangleright\!\triangleleft}\mathbb{C}F$.

(2) Let us describe explicitly the inclusion which corresponds to the Kac algebra $\mathcal{B}_{4m}$ ($m\geqslant 2$) defined by Definition 7.3(2). We can choose those $\chi,\theta\in\mathrm{Aut}(\mathcal{R})$ of order 2 and $u\in\mathbb{U}$, for which $\chi(u)=u^*$, $\theta(u)=-u^*$, $(\chi\theta)^{2m}=\mathrm{inn}(u)$, and $(\chi\theta)^i\notin\mathrm{Inn}(\mathcal{R})$ whenever $0<i<2m$. Set $\varphi=(\chi\theta)^2$. Let $\mathcal{P}$ denote the $\mathcal{R}$*-ring* (i.e., the algebra given by an algebra map from $\mathcal{R}$) which is generated by two elements B, Y, and is defined by the relations

$$B^m=u,\qquad Y^2=1,\qquad YB=B^{-1}Y,$$
$$Bx=\varphi(x)B,\qquad Yx=\chi(x)Y\quad(x\in\mathcal{R}).$$

Then, $\mathcal{P}$ is a hyperfinite II_1 factor including $\mathcal{R}$, with B, Y unitaries. Moreover one can prove that

$$\mathcal{P}\supset\mathcal{R}^\theta\ \big(=\big\{x\in\mathcal{R}\mid\theta(x)=x\big\}\big)$$

is the irreducible depth 2 inclusion corresponding to $\mathcal{B}_{4m}$; see [M8, Proposition 5.6].

Acknowledgements

The author thanks very much Sonia Natale and Yoshihiro Sekine for their valuable suggestions and helpful comments.

References

[A] N. Andruskiewitsch, About finite-dimensional Hopf algebras, in: Contemp. Math., vol. 294, Amer. Math. Soc., Providence, RI, 2002, pp. 1–57.

[AD] N. Andruskiewitsch, J. Devoto, Extensions of Hopf algebras, Algebra i Analiz 7 (1995) 22–61; also in St. Petersburg Math. J. 7 (1996) 17–52.

[AS] N. Andruskiewitsch, H.-J. Schneider, Lifting of quantum linear spaces and pointed Hopf algebras of order p^3, J. Algebra 209 (1998) (1996) 659–691.

[BCM] R. Blattner, M. Cohen, S. Montgomery, Crossed products and inner actions of Hopf algebras, Trans. Amer. Math. Soc. 298 (1986) 671–711.

[BD] M. Beattie, S. Dăscălescu, Hopf algebras of dimension 14, math.QA/0205243.

[BDG] M. Beattie, S. Dăscălescu, L. Grünenfelder, On the number of types of finite-dimensional Hopf algebras, Invent. Math. 136 (1999) 1–7.

[CDMM] C. Călinescu, S. Dăscălescu, A. Masuoka, C. Menini, Quantum lines over non-cocommutative cosemisimple Hopf algebras, J. Algebra 273 (2004) 753–779.

[Da] M.-C. David, Paragroupe d'Adrian Ocneanu et algèbre de Kac, Pacific J. Math. 172 (1996) 331–363.

[Doi] Y. Doi, Equivalent crossed products for a Hopf algebra, Comm. Algebra 17 (1989) 3053–3085.

[DT] Y. Doi, M. Takeuchi, Cleft comodule algebras for a bialgebra, Comm. Algebra 14 (1986) 801–818.

[D1] V. Drinfeld, Quantum groups, in: Proc. Internat. Congress Math., vol. 1, Berkeley, 1986, Amer. Math. Soc., Providence, RI, 1988, pp. 789–820.

[D2] V. Drinfeld, On almost cocommutative Hopf algebras, Leningrad Math. J. 1 (1990) 321–342.

[EG1] P. Etingof, S. Gelaki, Some properties of finite-dimensional semisimple Hopf algebras, Math. Res. Lett. 5 (1–2) (1998) 191–197.

[EG2] P. Etingof, S. Gelaki, Semisimple Hopf algebras of dimension pq are trivial, J. Algebra 210 (1998) 664–669.

[EG3] P. Etingof, S. Gelaki, On finite-dimensional semisimple and cosemisimple Hopf algebras in positive characteristic, Internat. Math. Res. Notices 16 (1998) 851–864.

[EG4] P. Etingof, S. Gelaki, On the exponent of finite-dimensional Hopf algebras, Math. Res. Lett. 6 (1999) 131–140.

[EG5] P. Etingof, S. Gelaki, The classification of triangular semisimple and cosemisimple Hopf algebras over an algebraically closed field, Internat. Math. Res. Notices 5 (2000) 223–234.

[EG6] P. Etingof, S. Gelaki, On families of triangular Hopf algebras, Internat. Math. Res. Notices 14 (2002) 757–768.

[EG7] P. Etingof, S. Gelaki, The classification of triangular Hopf algebras over an algebraically closed field of characteristic zero, Mosc. Math. J. 3 (2003) 37–43, 258.

[EG8] P. Etingof, S. Gelaki, On Hopf algebras of dimension pq, J. Algebra 277 (2004) 668–674.

[ENO] P. Etingof, D. Nikshych, V. Ostrik, On fusion categories, Ann. of Math. (2) 162 (2005) 581–642.

[GN] C. Galindo, N. Sonia, Simple Hopf algebras and deformations of finite groups, Math. Res. Lett., in press; electronic preprint, math.QA 0608734.

[G] S. Gelaki, Pointed Hopf algebras and Kaplansky's 10th conjecture, J. Algebra 209 (1998) 635–657.

[H] C. Hoffman, On some examples of simple quantum groups, Comm. Algebra 28 (2000) 1867–1873.

[IK] M. Izumi, H. Kosaki, Kac algebras arising from composition of subfactors: General theory and classification, Mem. Amer. Math. Soc. 750 (2002).

[J] V. Jones, Actions of finite groups on the hyperfinite II_1 factor, Mem. Amer. Math. Soc. 237 (1980).

[K1] G. Kac, Extensions of groups to ring groups, Math. USSR Sb. 5 (1968) 451–474.

[K2] G. Kac, Certain arithmetic properties of ring groups, Funct. Anal. Appl. 6 (1972) 158–160.

[KP] G. Kac, V. Paljutkin, Finite ring groups, Trans. Moscow Math. Soc. 5 (1966) 251–294.

[KN] L. Kadison, D. Nikshych, Hopf algebra actions on strongly separable extensions of depth two, Adv. Math. 163 (2001) 258–286.

[Kap] I. Kaplansky, Bialgebras, Lecture Notes in Mathematics Department of Mathematics, Univ. of Chicago, 1975.

[Ks1] Y. Kashina, On the antipode of Hopf algebras in ${}^H_H\mathcal{YD}$, Comm. Algebra 27 (1999) 1261–1273.

[Ks2] Y. Kashina, Classification of semisimple Hopf algebras of dimension 16, J. Algebra 232 (2000) 617–663.

[KSZ1] Y. Kashina, Y. Sommerhäuser, Y. Zhu, Self-dual modules of semisimple Hopf algebras, J. Algebra 257 (2002) 88–96.

[KSZ2] Y. Kashina, Y. Sommerhäuser, Y. Zhu, On higher Frobenius–Schur indicators, Mem. Amer. Math. Soc. 181 (2005) 855.

[KT] H.F. Kreimer, M. Takeuchi, Hopf algebras and Galois extensions of an algebra, Indiana Univ. Math. J. 30 (1981) 675–692.

[L] R.G. Larson, Characters of Hopf algebras, J. Algebra 17 (1971) 352–368.

[LR] R.G. Larson, D.E. Radford, Semisimple cosemisimple Hopf algebras, Amer. J. Math. 109 (1987) 187–195.

[LM] V. Linchenko, S. Montgomery, A Frobenius–Schur theorem for Hopf algebras, Algebr. Represent. Theory 3 (2000) 347–355.

[LYZ] J.-H. Lu, M. Yan, Y. Zhu, Quasi-triangular structures on Hopf algebras with positive bases, in: Contemp. Math., vol. 267, Amer. Math. Soc., Providence, RI, 2000, pp. 339–356.

[M1] A. Masuoka, Semisimple Hopf algebras of dimension $2p$, Comm. Algebra 23 (1995) 1931–1940.

[M2] A. Masuoka, Selfdual Hopf algebras of dimension p^3 obtained by extension, J. Algebra 178 (1995) 791–806.

[M3] A. Masuoka, The p^n theorem for semisimple Hopf algebras, Proc. Amer. Math. Soc. 124 (1996) 735–737.

[M4] A. Masuoka, Faithfully flat forms and cohomology of Hopf algebra extensions, Comm. Algebra 25 (1997) 1169–1197.

[M5] A. Masuoka, Cocycle deformations and Galois objects for some cosemisimple Hopf algebras of finite dimension, in: Contemp. Math., vol. 267, Amer. Math. Soc., Providence, RI, 2000, pp. 195–214.

[M6] A. Masuoka, Defending the negated Kaplansky conjecture, Proc. Amer. Math. Soc. 129 (2001) 3185–3192.

[M7] A. Masuoka, Hopf algebra extensions and cohomology, in: S. Montgomery, H.-J. Schneider (Eds.), New Directions in Hopf Algebras, in: Math. Sci. Res. Inst. Publ., vol. 43, Cambridge Univ. Press, 2002, pp. 167–209.

[M8] A. Masuoka, More homological approach to composition of subfactors, J. Math. Sci. Univ. Tokyo 10 (2003) 599–630.

[M9] A. Masuoka, Example of almost commutative Hopf algebras which are not coquasitriangular Hopf algebras, in: Lecture Notes Pure Appl. Math., vol. 237, Dekker, New York, 2004, pp. 185–191.

[Mi] Y. Miyashita, On Galois extensions and crossed products, J. Fac. Sci. Hokkaido Univ. Ser. I 21 (1970) 97–121.

[Mo1] S. Montgomery, Hopf Algebras and Their Actions on Rings, CBMS Reg. Conf. Ser. Math., vol. 82, Amer. Math. Soc., Providence, RI, 1993.

[Mo2] S. Montgomery, Classifying finite-dimensional semisimple Hopf algebras, in: Contemp. Math., vol. 229, Amer. Math. Soc., Providence, RI, 1998, pp. 265–279.

[Mo3] S. Montgomery, Representation theory of semisimple Hopf algebras, in: NATO Sci. Ser. II Math. Phys. Chem., vol. 28, Kluwer Acad. Publ., Dordrecht, 2001, pp. 189–218.

[MW] S. Montgomery, S. Witherspoon, Irreducible representations of crossed products, J. Pure Appl. Algebra 129 (1998) 315–326.

[Mv] M. Movshev, Twisting in group algebras of finite groups, Funct. Anal. Appl. 27 (1994) 240–244.

[Mu] E. Müller, Finite subgroups of the quantum general linear group, Proc. London Math. Soc. 81 (1) (2000) 190–210.

[N1] S. Natale, On semisimple Hopf algebras of dimension pq^2, J. Algebra 221 (1999) 242–278.

[N2] S. Natale, On semisimple Hopf algebras of dimension pq^2, II, Algebr. Represent. Theory 4 (2001) 277–291.

[N3] S. Natale, On semisimple Hopf algebras of dimension pq^r, Algebr. Represent. Theory 7 (2004) 173–188.

[N4] S. Natale, Semisolvability of semisimple Hopf algebras of low dimension, Mem. Amer. Math. Soc. 186 (2007).

[Ng1] S.-H. Ng, Non-semisimple Hopf algebras of dimension p^2, J. Algebra 255 (2002) 182–197.

[Ng2] S.-H. Ng, Hopf algebras of dimension $2p$, Proc. Amer. Math. Soc. 133 (2005) 2237–2242.

[NZ] W. Nichols, M.B. Zoeller, A Hopf algebra freeness theorem, Amer. J. Math. 111 (1989) 381–385.

[Nk] D. Nikshych, K_0-rings and twisting of finite-dimensional semisimple Hopf algebras, Comm. Algebra 26 (1998) 321–342.

[OS] U. Oberst, H.-J. Schneider, Untergruppen formeller Gruppen von endlichen Index, J. Algebra 31 (1974) 10–44.

[Oc] A. Ocneanu, Quantized group, string algebras and Galois theory for algebras, in: Operator Algebras and Applications, vol. 2, in: London Math. Soc. Lecture Note Ser., vol. 136, Cambridge Univ. Press, Cambridge, 1988, pp. 119–172.

[P1] S. Popa, Classification of subfactors: the reduction to commuting squares, Invent. Math. 101 (1990) 19–43.

[P2] S. Popa, Classification of amenable subfactors of type II, Acta Math. 172 (1994) 163–255.

[R1] D.E. Radford, The order of the antipode of a finite-dimensional Hopf algebra is finite, Amer. J. Math 98 (1976) 333–355.

[R2] D.E. Radford, The group of automorphisms of a semisimple Hopf algebra over a field of characteristic zero is finite, Amer. J. Math. 112 (1990) 331–357.

[Sb1] P. Schauenburg, Hopf bi-Galois extensions, Comm. Algebra 24 (1996) 3797–3825.

[Sb2] P. Schauenburg, Hopf bimodules, coquasibialgebras, and an exact sequence of Kac, Adv. Math. 163 (2002) 194–263.

[Sb3] P. Schauenburg, Hopf algebra extensions and monoidal categories, in: S. Montgomery, H.-J. Schneider (Eds.), New Directions in Hopf algebras, in: Math. Sci. Res. Inst. Publ., vol. 43, Cambridge Univ. Press, 2002, pp. 321–381.

[SS] P. Schauenburg, H.-J. Schneider, Galois type extensions and Hopf algebras, Banach Center Publ., in press.

[S1] H.-J. Schneider, Normal basis and transitivity of crossed products for Hopf algebras, J. Algebra 151 (1992) 289–312.

[S2] H.-J. Schneider, Some properties of factorizable Hopf algebras, Proc. Amer. Math. Soc. 129 (2001) 1891–1898.

[Se] Y. Sekine, An example of finite dimensional Kac algebras of Kac–Paljutkin type, Proc. Amer. Math. Soc. 124 (1996) 1139–1147.

[Sk] S. Skryabin, Projectivity and freeness over comodule algebras, Trans. Amer. Math. Soc. 359 (2007) 2597–2623.

[St] D. Ştefan, The set of types of n-dimensional semisimple and cosemisimple Hopf algebras is finite, J. Algebra 193 (1997) 571–580.

[Su] S. Suzuki, A family of braided cosemisimple Hopf algebras of finite dimension, Tsukuba J. Math. 22 (1998) 1–22.

[Sz] W. Szymański, Finite index subfactors and Hopf algebra crossed products, Proc. Amer. Math. Soc. 120 (1994) 519–528.

[Tf] E.J. Taft, The order of the antipode of a finite-dimensional Hopf algebra, Proc. Natl. Acad. Sci. USA 68 (1971) 2631–2633.

[T1] M. Takeuchi, Matched pairs of groups and bismash products of Hopf algebras, Comm. Algebra 9 (1981) 841–882.

[T2] M. Takeuchi, Modular categories and Hopf algebras, J. Algebra 243 (2001) 631–643.

[T3] M. Takeuchi, Survey of matched pairs of groups, Banach Center Publ. 61 (2003) 305–331.

[Z] Y. Zhu, Hopf algebras of prime dimension, Internat. Math. Res. Notes 1 (1994) 53–59.

This page intentionally left blank

Section 7
Machine Computation. Algorithms. Tables. Counting Algebraic Structures

This page intentionally left blank

Integral Representation and Algorithms for Closed Form Summation[⋆]

Georgy P. Egorychev

Krasnoyarsk State Technical University, Kirenskogo 26, *Krasnoyarsk* 660074, *Russia*
E-mail: *anott@scn.ru*

Eugene V. Zima

Physics and Computer Science, Wilfrid Laurier University, Waterloo, Canada
E-mail: *ezima@wlu.ca*

Contents

[⋆]Supported in part by the Natural Sciences and Engineering Research Council of Canada and the Russian Fund of Fundamental Researches.

HANDBOOK OF ALGEBRA, VOL. 5
Edited by M. Hazewinkel
© 2008 Elsevier B.V. All rights reserved
DOI: 10.1016/S1570-7954(07)05009-7

Abstract

The problem of computation and estimation of finite and infinite sums (generating functions) often arises in combinatorics and graph theory, theory of algorithms and computer algebra, group theory and function theory, probability theory and asymptotical analysis as well as in physics, statistical mechanics, and other areas of knowledge.

This article is intended for a wide audience including graduate students and researchers in the various applied fields. Here we present the history, main results, model examples, various applications and perspectives of investigation in two connected general approaches to summation: the integral representation and computation of combinatorial sums (“the method of coefficients”) and the modern algorithmic approach. Each of the approaches is based on classic results of mathematical analysis, function theory and computer algebra.

1. Introduction

The sources for the application of integrals for the computation of combinatorial sums go back to L. Euler, C. Gauss, A. Cauchy, G. Pólya and other classical masters of mathematical analysis and algebra. The concept of integral representation of sums has undergone profound development from estimates of sums in analytic number theory: the cyclotomic method (Ja. Uspensky, G. Hardy, J. Littlewood, S. Ramanujan, and others), the method of complex integration and the method of trigonometric sums (I. Vinogradov). Since C. Gauss the integral representation and formulae for the summation of hypergeometric series are widely used for the computation of sums with ordinary and q-binomial coefficients. Jointly with finite differences this apparatus is effectively used for computing binomial sums with the help of a computer [166]. Of particular interest are Pólya's investigations which are on the boundary of discrete and continuous mathematics and use contour integrals [170]. This approach was systematically developed by modern authors (see [22,23,48,59,76,81, 152,141]; see also [32–34,51,172,173] and others).

Here we present a general scheme and numerous applications for the method of integral representation and evaluation of combinatorial sums. This approach originates from classical results of mathematical analysis and the well-known method of generating functions in combinatorics. The development of this method known as "the method of coefficients" was carried out ever since the first publication of the first author of this article for more than 40 years (see [52–75,14,92], and also [99,147,121,159,203] and others). The method has been extended to computations with Laurent formal power series over $\mathbb{C}$. The main results and verification of the method of coefficients with the help of the theory of multidimensional residues were given in [59]. We have done the direct, short and uniform analytic computation of several thousand sums, have found asymptotics, multidimensional analogs and recurrence relations for several of them, have corrected a series of incorrect identities obtained by other investigators earlier, etc. The material of the following known books on combinatorial identities, generating functions and its applications was completely or partly reworked during these years: P.A. MacMahon (1915–1916), I. Schwatt (1924), V. Kudrjavtzev (1936), W. Feller (1957), "The Otto Dunkel Memorial Problem Book", Amer. Math. Soc., New York (1957), J. Riordan (1958, 1968), E. Netto (1958), D. Knuth (1968), J. Percus (1971), E.B. McBride (1971), V. Vilenkin (1971), G. Gould (1972), K. Rybnikov (1972), J. Kaucký (1975), M. Lucas (1979), M. Platonov (1979), D. Green and D. Knuth (1982), I. Goulden and D. Jackson (1983), R. Stanley (1997), A. Prudnikov, Yu. Brychkov and O. Marichev (1988), R. Graham, D. Knuth and O. Patashnik (1994), H. Wilf (1994), M. Petkovšek, H. Wilf and D. Zeilberger (1996) (see [94,98,99,115,118,125,133,143,155, 157,169,171,174,176,181,188,197,207,210]).

For example we considered more than 200 of the first identities in Gould's book (1972). An analysis of the singularities of the integrals has enabled us to explain why they can be computed in closed form, to write out new identities of the same type, and to obtain some asymptotic estimates for them [59]. In the same way one can obtain practically unlimited amounts of new identities with combinatorial numbers (see, for example, Section 3.3.3 below).

Here we will also solve several original summation problems, that were not solved by traditional methods and which arose in combinatorial analysis and graph theory, function theory and group theory, and other fields. In particular:

- the Riordan problem (proposed in 1968) about the classification of known invertible combinatorial relations;
- computing the closed form of Szegő and Bergman kernels for an extended class of n-circular domains in $\mathbb{C}^n$;
- the Kargapolov problem (proposed in 1972) on the enumeration of the ranks of factors in the lower central series of free k-step solvable groups;
- series of enumerative problems for groups and algebras of Lie type and new formulae involving summation by partitions;
- description of a characteristic function of the stopping height in the well-known Collatz conjecture;
- a multidimensional sum with polynomial coefficients arising in the well known Jacobian conjecture (two-dimensional case);
- the first polynomial identities for permanents.

We provide a justification of the method of coefficients with the help of the theory of one-dimensional and multidimensional residues. Solutions of problems above mentioned have lead to a better understanding of the algebraic, topological and combinatorial structures of the investigated objects: this allowed in many cases to obtain essential generalizations of those results. In this sense our approach is a part of the general approaches of this kind (see also [158]).

We will also give a brief overview of well-known summation algorithms used in computer algebra systems such as Maple [153], apply the integral representation approach to indefinite summation for the first time, and establish links between modern algorithms and methods of integral representation. This leads to some improvements of existing algorithms, obtained as the result of comparison of two approaches. Co-operative investigation of the particular summation problem gives new representation for the complementary Bell numbers.

At the end of this chapter an extensive list of open problems is presented.

2. Idea and method of integral representation of combinatorial sums

2.1. *History of the problem*

Combinatorial sums (numbers, expressions) express the characteristics of some combinatorial schemes in terms of its parameters. Some of them allow analytical representations which can be simplified. The result of computations generates a combinatorial identity. In computing combinatorial sums an extensive collection of methods and stages is used – from mathematical induction, inclusion–exclusion principle, properties of binomial coefficients and other combinatorial numbers, various operators, symbolic methods such as Blissard's calculus, summation formulae for hypergeometric terms and special functions of various types, difference–integral–differential equations – to combinatorial interpretations of sums in the framework of various combinatorial schemes: permutations, combinations, arrangements of particles in cells, counting the number of paths on various lattices and many others, – and more general methods: partially group-theoretic method of Pólya, incidence algebras and Möbius functions on partially ordered sets and others. However, the procedure of computation in each particular case usually was individual. Although the origin of application of the integrals for the computation of combinatorial sums goes back to L. Euler, C. Gauss, A. Cauchy, G. Pólya, algebras of analytic functions and multiple integrals were used only occasionally. For example, in well-known work by I. Good and F. Dyson (early 70s of the past century) generating functions adapted to a search of change of variables and the computation of some multiple contour integrals corresponding to combinatorial sums is carried out. Nevertheless the procedure for going through the most important step of the computation – the passage from a combinatorial sum to its integral representation – was still almost undeveloped. Often various integral representations arise during the initial steps of the calculation of combinatorial sums of various type. At the same time we have concluded that the general computational scheme remains practically without variations. It is important also that the concept of an integral representation of sums has undergone profound development from the computation of sums in analytic number theory and asymptotic analysis (see, for example, [48,156]).

Modern algorithmic treatments of such summations, that lead to immediate implementations on a computer, has started with work of S.A. Abramov (1971), G. Gosper (1977), D. Zeilberger (1991). Many improvements of the summation and theoretical foundations of these algorithmic approaches were proposed and implemented by M. Bronstein, F. Chyzak, J. Gerhard, M. van Hoeij, P. Paule, M. Petkovšek, B. Salvy, V. Strehl, H. Wilf, E. Zima and others.

2.2. *Algebra of analytic functions*

2.2.1. *Computational scheme* The general scheme of the method of integral representation of sums can be broken up into the following steps.

1. Assignment of a table of integral representation of certain combinatorial numbers. For example,[1]

- The binomial coefficients:

$$\begin{aligned}\binom{m}{k} &= \mathbf{res}_w (1+w)^m w^{-k-1} \\ &= \frac{1}{2\pi i}\int_{|w|=\varrho} (1+w)^m w^{-k-1}\,dw, \quad 0<\varrho<1; \end{aligned} \tag{1}$$

$$\begin{aligned}\binom{m}{k} &= \mathbf{res}_w (1+w)^m w^{-k-1} \\ &= \frac{1}{2\pi i}\int_{|w|=\varrho} (1+w)^m w^{-k-1}\,dw, \quad \varrho>0; \end{aligned} \tag{2}$$

[1] The **res** operator will be defined in Section 3.1.

$$\binom{m+k-1}{k} = \binom{-m}{k} = \mathbf{res}_w (1-w)^{-m} w^{-k-1}$$
$$= \frac{1}{2\pi i} \int_{|w|=\varrho} (1-w)^{-m} w^{-k-1}\, dw, \quad 0 < \varrho < 1. \tag{3}$$

- The Kronecker symbol $\delta(n,k)$:

$$\delta(n,k) = \mathbf{res}_w w^{-n+k-1}. \tag{4}$$

- The Stirling numbers of the second kind:

$$\left\{ {m \atop k} \right\} = \frac{m!}{k!} \mathbf{res}_w (-1 + \exp w)^k w^{-m-1}$$
$$= \frac{m!}{k!} \frac{1}{2\pi i} \int_{|w|=\varrho} (-1 + \exp w)^k w^{-m-1}\, dw, \quad \varrho > 0; \tag{5}$$

- The exponential coefficients:

$$\alpha^k / k! = \mathbf{res}_w \exp(\alpha w) w^{-k-1}$$
$$= \frac{1}{2\pi i} \int_{|w|=\varrho} \exp(\alpha w) w^{-k-1}\, dw, \quad \varrho > 0; \tag{6}$$
$$1/\alpha^j = \frac{1}{(j-1)!} \int_0^\infty u^{-j+1} \exp(-\alpha u)\, du, \quad \Re\alpha > -1. \tag{7}$$

- The multinomial (polynomial) coefficients, $\alpha \in \mathbb{R}$:

$$\binom{\alpha}{k_1, k_2, \dots, k_n}$$
$$= \mathbf{res}_w (1 + \sum_{s=1}^r w_s)^\alpha w_1^{-k_1-1} w_2^{-k_2-1} \cdots w_r^{-k_r-1}$$
$$= \frac{1}{(2\pi i)^r} \int_{\Gamma_r(\varepsilon)} \left(1 + \sum_{s=1}^r w_s\right)^\alpha w_1^{-k_1-1} w_2^{-k_2-1} \cdots w_r^{-k_r-1}\, dw. \tag{8}$$

Here $w = (w_1, w_2, \dots, w_r), dw = dw_1 \wedge dw_2 \wedge \dots \wedge dw_r, \varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_r)$, $\Gamma = \Gamma_r(\varepsilon) = \{w = (w_1, w_2, \dots, w_r) \in \mathbb{C}^r\colon |w_1| = \varepsilon_1, |w_2| = \varepsilon_2, \dots, |w_r| = \varepsilon_r\}$ is the skeleton of a polycylinder $U = U_r(\varepsilon) = \{w = (w_1, w_2, \dots, w_r)\colon |w_1| \leqslant \varepsilon_1, |w_2| \leqslant \varepsilon_2, \dots, |w_r| \leqslant \varepsilon_r\}$.

The table can be supplemented in the course of the computations.

2. Representation of the summand a_k of the original sum $\sum_k a_k$ by a sum of products of combinatorial numbers.

3. Replacement of the combinatorial numbers by their integrals.

4. Reduction of products of integrals to multiple integrals.

5. Interchange the order of summation and integration. This gives an integral representation of the original sum with the kernel represented by a series. The use of this transformation requires us to deform the domain of integration in such a way as to obtain a series under the integral which converges uniformly on this domain saving the value of the integral.

6. Summation of the series under the integral sign. As the rule, this series turns out to be a geometric progression [102]. This gives the integral representation of the original sum with the kernel in closed form.

7. The computation of the resulting integral by means of tables of integrals, iterated integration, the theory of single and multiple residues, or new methods (for example, the splitting lemma).

2.2.2. *Examples* The first two examples are illustrative.

1. Compute the Hardy sum

$$S_m = \sum_{k=0}^{\lfloor m/2 \rfloor} (-1)^k \frac{1}{m-k} \binom{m-k}{k}, \quad m = 1, 2, \ldots.$$

According to formula (1) for the summand (steps 1 and 2):

$$\begin{aligned}
&(-1)^k \frac{1}{m-k} \binom{m-k}{k} \\
&\quad = \frac{1}{m}\left(\binom{m-k}{k} + \binom{m-k-1}{k-1}\right) = \frac{1}{m}\left(\binom{m-k}{m-2k} + \binom{m-k-1}{m-2k}\right) \\
&\quad = \frac{1}{m}\left(\oint (1+w)^{m-k} w^{-m+2k-1}\, dw + \oint (1+w)^{m-k-1} w^{-m+2k-1}\, dw\right) \\
&\quad = \frac{1}{m} \oint (1+w)^{m-k-1}(2+w) w^{-m+2k-1}\, dw.
\end{aligned}$$

Then

$$\begin{aligned}
S_m &= \sum_{k=0}^{\lfloor m/2 \rfloor} (-1)^k \frac{1}{m} \oint (1+w)^{m-k-1}(2+w) w^{-m+2k-1}\, dw \\
&= \sum_{k=0}^{\infty} (-1)^k \frac{1}{m} \oint (1+w)^{m-k-1}(2+w) w^{-m+2k-1}\, dw
\end{aligned}$$

(as for each $k > \lfloor m/2 \rfloor$ the added terms are equal to zero). Choose ϱ, $0 < \varrho < \infty$, small enough, say $\varrho = 1/2$, so that the geometric series $\sum_{k=0}^{\infty}(-w^2/(1+w)^2)^k$ converges uniformly on the contour of integration $|w| = 1/2$:

$$S_m = \frac{1}{2\pi i m} \int_{|w|=1/2} (1+w)^{m-1}(2+w) w^{-m-1} \left\{ \sum_{k=0}^{\infty} \left(-w^2/(1+w)^2\right)^k \right\} dw.$$

By the formula for geometric summation (6th step)

$$S_m = \frac{1}{m}\oint (1+w)^m (2+w)\bigl(1+w+w^2\bigr)^{-1} w^{-m-1}\,dw.$$

The last step (by the residue theorem)

$$\begin{aligned} S_m &= -\frac{1}{m}\bigl[(1+w)^m(2+w)(1+2w)^{-1}w^{-m-1}\bigr]_{w=-1/2\pm i\sqrt{3}/2} \\ &= \frac{2(-1)^m}{m}\cos(2\pi m/3), \end{aligned}$$

and, finally, we get Hardy's identity

$$\sum_{k=0}^{\lfloor m/2\rfloor} (-1)^k \frac{1}{m-k}\binom{m-k}{k} = \frac{2(-1)^m}{m}\cos(2\pi m/3), \quad m=1,2,\dots.$$

2. Rational function summation. Compute the sum

$$\begin{aligned} S_m &= \sum_{k=1}^{m}\left\{\frac{14k+14}{(2k-1)(2k+1)(2k+3)} + \frac{2k+1}{k^2(k+1)^2}\right\} \\ &= \sum_{k=1}^{m}\left\{\frac{2}{(2k-1)} - \frac{3}{(2k+1)} + \frac{1}{(2k+3)} + \frac{1}{k^2} - \frac{1}{(k+1)^2}\right\}. \end{aligned}$$

SOLUTION. According to the formula (7) for applied to each term of the sum

$$\begin{aligned} S_m &= \sum_{k=1}^{m}\left\{\frac{2}{(2k-1)} - \frac{3}{(2k+1)} + \frac{1}{(2k+3)} + \frac{1}{k^2} - \frac{1}{(k+1)^2}\right\} \\ &= \sum_{k=1}^{m}\left\{2\int_0^\infty e^{-(2k+1)u}\,du - 3\int_0^\infty e^{-(2k-1)u}du + \int_0^\infty e^{-(2k+3)u}\,du \right. \\ &\quad \left. + \int_0^\infty ue^{-ku}\,du - \int_0^\infty ue^{-(k+1)u}\,du\right\} \\ &= \sum_{k=1}^{m}\int_0^\infty \bigl\{e^{-(2k+1)u}\bigl(2-3e^{-2u}+e^{-4u}\bigr) + ue^{-ku}\bigl(1-e^{-u}\bigr)\bigr\}\,du \\ &= \int_0^\infty \bigl(2-3e^{-2u}+e^{-4u}\bigr)\left\{\sum_{k=1}^{m} e^{-(2k+1)u}\right\}du \\ &\quad + \int_0^\infty u\bigl(1-e^{-u}\bigr)\left\{\sum_{k=1}^{m} e^{-ku}\right\}du \\ &\quad \text{(by the formula for geometric progressions, used twice)} \\ &= \int_0^\infty e^{-u}P_1\bigl(e^{-2u}\bigr)\bigl(1-e^{-2um}\bigr)\bigl(1-e^{-2u}\bigr)^{-1}\,du \end{aligned}$$

$$+\int_0^\infty ue^{-u}P_2\big(e^{-u}\big)\big(1-e^{-mu}\big)\big(1-e^{-u}\big)^{-1}\,du,$$

where $P_1(t)=2-3t+t^2$ and $P_2(t)=1-t$. Using that $P_1(1)=P_2(1)=0$, $P_1(e^{-2u})(1-e^{-2u})^{-1}=(2-e^{-2u})$ and $P_2(e^{-u})(1-e^{-u})^{-1}=1$, we have

$$\begin{aligned}
S_m &= \int_0^\infty e^{-u}\big(2-e^{-2u}\big)\big(1-e^{-2um}\big)\,du+\int_0^\infty ue^{-u}\big(1-e^{-mu}\big)\,du\\
&= \int_0^\infty \big\{2e^{-u}-e^{-3u}-2e^{-u(2m+1)}-2e^{-u(m+2)}+ue^{-u}-e^{-u(m+1)}\big\}\,du\\
&\qquad \text{(calculating each integral by formula (7))}\\
&= \frac{4}{3}-\frac{2}{2m+1}-\frac{2}{m+2}-\frac{1}{(m+1)^2},
\end{aligned}$$

and so, finally, we get an identity

$$\begin{aligned}
&\sum_{k=1}^{m}\left\{\frac{14k+14}{(2k-1)(2k+1)(2k+3)}+\frac{2k+1}{k^2(k+1)^2}\right\}\\
&\quad=\frac{4}{3}-\frac{2}{2m+1}-\frac{2}{m+2}-\frac{1}{(m+1)^2},\quad m=1,2,\ldots.
\end{aligned}$$

REMARK 1. It is interesting that the integral representations (another type) have been used already for summation of elementary fractions of the first degree [103].

3. Revision of the Strazdin combinatorial formula. In [201] the following combinatorial formula from graph theory is given: if $n\geqslant k$, then

$$\begin{aligned}
S_{n,k} &= \sum_{r=0}^{\infty}\sum_{j=0}^{k}\sum_{i=0}^{\lfloor j/2\rfloor}(-1)^j\binom{2k-j-2i}{j-2i}\binom{2n-2i}{r-i}\\
&= 2^{2(n-k)-1}(k+1)(k+2).
\end{aligned}$$

We would like to verify it by computing the left-hand side using the integral representation approach.

SOLUTION. As $\sum_{r=0}^{\infty}\binom{2n-2i}{r-i}=2^{2n-2i}$,

$$\begin{aligned}
S_{n,k} &= \sum_{j=0}^{k}\sum_{i=0}^{\lfloor j/2\rfloor}(-1)^j 2^{2n-2i}\binom{2k-j-2i}{j-2i}\\
&\qquad \text{(change of summation index } j=k-j)\\
&= \sum_{j=0}^{k}\sum_{i=0}^{\lfloor j/2\rfloor}(-1)^{k-j}2^{2n-2i}\binom{k+j-2i}{k-j-2i}\\
&= \sum_{j=0}^{k}\sum_{i=0}^{\lfloor j/2\rfloor}(-1)^{k-j}2^{2n-2i}\int_{|w|=\varrho}(1+w)^{k+j-2i}w^{-k+j+2i-1}\,dw
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\frac{1}{2\pi i}\cdots \\
&= \frac{(-1)^k 4^n}{2\pi i}\int_{|w|=1/3}\left\{\sum_{i=0}^{\infty}\bigl(w^2/4(1+w)^2\bigr)^i\right\} \\
&\quad\times\left\{\sum_{j=0}^{\infty}(-w(1+w))^j\right\}(1+w)^k w^{-k-1}\,dw \\
&= \frac{(-1)^k 4^n}{2\pi i}\int_{|w|=1/3}(1+w)^{k+2} \\
&\quad\times\bigl\{(w+2)(3w+2)(w^2+w+1)\bigr\}^{-1}w^{-k-1}\,dw \\
&\quad\text{(by the theorem on full sums of residues)} \\
&= \frac{(-1)^{k+1}4^{n+1}}{2\pi i}\bigl[(1+w)^{k+2}(3w+2)^{-1}\bigl(w^2+w+1\bigr)^{-1}w^{-k-1}\bigr]_{w=-2} \\
&\quad+\frac{1}{3}\bigl[(1+w)^{k+2}(w+2)^{-1}(w^2+w+1)^{-1}w^{-k-1}\bigr]_{w=-2/3} \\
&\quad+\bigl[(1+w)^{k+2}(w+2)^{-1}(3w+2)^{-1} \\
&\quad\times w^{-k-1}\bigl(w-\bigl((-1\mp i\sqrt{3})/2\bigr)^{-1}\bigr)\bigr]_{w=(-1\pm i\sqrt{3})/2} \\
&= 2^{2n+1}\left\{\frac{(-1/2)^{k+2}}{3}+\frac{3}{7}2^{-k-2}+\frac{10}{21}\cos(2k\pi/3)+\frac{2\sqrt{3}}{21}\sin(2k\pi/3)\right\},
\end{aligned}
$$

and, finally, we get a new identity

$$
\begin{aligned}
&\sum_{r=0}^{\infty}\sum_{j=0}^{k}\sum_{i=0}^{\lfloor j/2\rfloor}(-1)^j\binom{2k-j-2i}{j-2i}\binom{2n-2i}{r-i} \\
&\quad= 2^{2n+1}\left\{\frac{(-1/2)^{k+2}}{3}+\frac{3}{7}2^{-k-2}+\frac{10}{21}\cos(2k\pi/3)+\frac{2\sqrt{3}}{21}\sin(2k\pi/3)\right\}, \\
&\quad n\geqslant k.
\end{aligned}
$$

It is easy to see that in order to get right-hand side of the Strazdin formula, it is sufficient to replace $\binom{2n-2i}{r-i}$ by $\binom{2n-2j}{r-j}$ and repeat the same computational scheme. Thus we will get for $n\geqslant k$

$$
\sum_{r=0}^{\infty}\sum_{j=0}^{k}\sum_{i=0}^{\lfloor j/2\rfloor}(-1)^j\binom{2k-j-2i}{j-2i}\binom{2n-2j}{r-j}=2^{2(n-k)-1}(k+1)(k+2).
$$

4. Computation of the Bergman kernel for a 3-circular domain

Consider the following domain

$$
D=\bigl\{(z_1,z_2,z_3)\in\mathbb{C}_z^3\colon\ \bigl(|z_1|^2+|z_2|^2\bigr)^{1/q}+|z_3|^2<1\bigr\}. \tag{9}
$$

THEOREM 1. *(See [57], [59, Chapter 6].) Let $K(x) = K(x_1, x_2, x_3)$ be the Bergman kernel for 3-circular domain (9) with integration over the volume. Then $K(x)$ can be presented in some neighborhood of the origin by the following formulae*:

$$K(x) = \pi^{-3} \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \sum_{k_3=0}^{\infty} \frac{(q(k_1+k_2+2)+k_3+1)!(k_1+k_2+2)}{k_1!k_2!k_3!(q(k_1+k_2+2))!} x_1^{k_1} x_2^{k_2} x_3^{k_3}, \tag{10}$$

$$\begin{aligned} K(x) = {} & \frac{2}{\pi^3(2\pi i)^4} \int_{\Gamma(\varrho)} \frac{(1+u_3+u_4)^{2q+1}(1+u_1+u_2)^2}{u_4^2(u_1 - x_1(1+u_1+u_2)(1+u_3+u_4)^q)} \\ & \times \frac{du_1 \wedge du_2 \wedge du_3 \wedge du_4}{(u_2 - x_2(1+u_1+u_2)(1+u_3+u_4)^q)((u_3 - x_3(1+u_3+u_4))}, \end{aligned} \tag{11}$$

where $\Gamma(\varrho) = \{u = (u_1, u_2, u_3, u_4)\colon |u_j| = \varrho,\ j = 1, 2, 3, 4\}$ is the skeleton of a sufficiently small polycylinder $V(\varrho)$;

$$\begin{aligned} K(x) = {} & 2\pi^{-3}(1-x_1)^{q-2}\big((1-x_3)^q - (x_1+x_2)\big)^{-4} \\ & \times \big((2q+1)(1-x_3)^q - (q-1)(x_1+x_2)\big). \end{aligned} \tag{12}$$

PROOF. We consider only the proof of (11) and (12) using (10), which follows immediately from results of [57]. To obtain (10) we first represent the general term of the sum (10) in terms of polynomial coefficients:

$$2\pi^{-3} \sum_{|k| \geqslant 0} \binom{k_1+k_2+2}{k_1, k_2} \binom{q(k_1+k_2+2)+k_3+1}{k_3, 1} x_1^{k_1} x_2^{k_2} x_3^{k_3}. \tag{13}$$

Replacing the polynomial coefficients in (13) according to formula (6)

$$\binom{k_1+k_2+2}{k_1, k_2} = \frac{1}{(2\pi i)^2} \int_{\Gamma_1(\varrho)} (1+u_1+u_2)^{k_1+k_2+2} u_1^{-k_1-1} u_2^{-k_2-1}\, du_1 \wedge du_2,$$

$$\begin{aligned} & \binom{q(k_1+k_2+2)+k_3+1}{k_3, 1} \\ & \quad = \frac{1}{(2\pi i)^2} \int_{\Gamma_2(\varrho)} (1+u_3+u_4)^{q(k_1+k_2+2)+k_3+1} u_3^{-k_3-1} u_4^{-2}\, du_3 \wedge du_4, \end{aligned}$$

we get, after reducing the product of integrals to a multiple integral and interchanging summation and integration, that for all $|x_j| \leqslant \varepsilon$

$$\begin{aligned} K(x) = {} & \frac{2\pi^{-3}}{(2\pi i)^4} \int_{\Gamma(\varrho)} \frac{(1+u_3+u_4)^{2q+1}(1+u_1+u_2)^2}{u_1 u_2 u_3 u_4^2} \\ & \times \Bigg\{ \sum_{|k| \geqslant 0} \big(x_1(1+u_1+u_2)(1+u_3+u_4)^q/u_1\big)^{k_1} \\ & \times \big(x_2(1+u_1+u_2)(1+u_3+u_4)^q/u_2\big)^{k_2} \big(x_3(1+u_3+u_4)/u_3\big)^{k_3} \Bigg\}\, du. \end{aligned} \tag{14}$$

Here $\varepsilon > 0$ is small enough that the ratio of the progression under the integral sign is less than 1 on the cycle $\Gamma(\varrho)$. By the formula for the sum of a geometric progression, (14) gives us the integral representation (11) for the kernel $K(x)$.

To obtain (12) from (11) we make the following change of variables under the integral sign in (12) (Theorem 15):

$$\begin{aligned} v_1 &= u_1 - x_1(1+u_1+u_2)(1+u_3+u_4)^q, \\ v_2 &= u_2 - x_2(1+u_1+u_2)(1+u_3+u_4)^q, \\ v_3 &= u_3 - x_3(1+u_3+u_4). \end{aligned} \tag{15}$$

Taking into account, that the first terms on the right-hand sides in (15) for v_1, v_2, v_3 are greater in modulus on $\Gamma(\varrho)$ than the corresponding second term when $|x_j| \leqslant \varepsilon$, $j = 1, 2, 3$. Easy computation gives us

$$\begin{aligned} K(x) &= 2\pi^{-3}(1-x_3)^{q-2} \\ &\quad \times \int_{|u_4|=\varrho} (1+u_4)^{2q+1}\big((1-x_3)^q - (x_1+x_2)(1+u_4)^q\big)^{-2} u_4^{-2}\, du_4 \\ &= 2\pi^{-3}(1-x_3)^{q-2} \\ &\quad \times \left[\frac{\partial}{\partial u_4}(1+u_4)^{2q+1}\big((1-x_3)^q - (x_1+x_2)(1+u_4)^q\big)^{-2}\right]_{u_4=0}, \end{aligned}$$

which yields (12). □

3. Laurent formal power series over $\mathbb{C}$

Using the **res** concept and its properties the idea of the integral representation can be extended to sums that allow computation with the help of Laurent formal power series of one and several variables over $\mathbb{C}$. The **res** concept is directly connected with the classic concept of residue in the theory of analytic functions and which may be used with series of various types. This connection has enabled us to express properties of **res** operator analogous to properties of residue in the theory of analytic functions. This in turn allows us to unify the scheme of the method of integral representation of sums independently of what kind of series – convergent or formal – is being used (separately, or jointly) in the process of computation of a particular sum.

In this section we shall restrict ourself explaining only one-dimensional case, although in further computations the **res** concept will also be used for multiple series. Besides, the one-dimensional case is of interest both in itself and in the computation of multiple integrals (**res**) in terms of repeated integrals.

3.1. *Definition and properties of the* **res** *operator (method of coefficients)*

Let L be the set of a Laurent formal power series over $\mathbb{C}$; hence each element of L contains only a finite number of terms of negative degree. The *order* of the monomial $c_k w^k$ is k.

The *order* of the series $C(w) = \sum_k c_k w^k$ from L is minimal order of monomials with a nonzero coefficient. Let L_k denotes a set of series of order k, $L = \bigcup_{k=-\infty}^{\infty} L_k$. Two series $A(w) = \sum_k a_k w^k$ and $B(w) = \sum_k b_k w^k$ from L are equal if and only if $a_k = b_k$ for all k. We can introduce in L operations of addition, multiplication, substitution, inversion and differentiation[2] (see, [31,76,103,119,182]). Let $f(w)$, $\psi(w) \in L_0$. Further on we will use the following notations:

· $h(w) = wf(w) \in L_1$;
· $l(w) = w/\psi(w) \in L_1$;
· $z'(w) = \frac{d}{dw} z(w)$;
· $\bar{h} = \bar{h}(z) \in L_1$ – inverse series of the series $z = h(w) \in L_1$.

For $C(w) \in L$ define the *formal residue* as $\mathbf{res}_w C(w) = c_{-1}$. Let $A(w) = \sum_k a_k w^k$ be a *generating function* for a sequence $\{a_k\}$. Then

$$a_k = \mathbf{res}_w A(w) w^{-k-1}, \quad k = 0, 1, 2, \ldots. \tag{16}$$

For example, one of the possible representations of the binomial coefficients is

$$\binom{n}{k} = \mathbf{res}_w (1+w)^n w^{-k-1}, \quad k = 0, 1, \ldots, n.$$

There are several properties (rewriting rules) of the **res** operator which immediately follow from its definition and properties of operations in Laurent formal power series over $\mathbb{C}$. We list only a few of them which will be used in this article. Let $A(w) = \sum_k a_k w^k$ and $B(w) = \sum_k b_k w^k$ be generating functions from L.

RULE 1 *(**res** removal).*

$$\begin{gathered} \mathbf{res}_w A(w) w^{-k-1} = \mathbf{res}_w B(w) w^{-k-1} \quad \text{for all } k \\ \text{if and only if} \quad A(w) = B(w). \end{gathered} \tag{17}$$

RULE 2 *(Linearity).* For any α, β from $\mathbb{C}$

$$\alpha \mathbf{res}_w A(w) w^{-k-1} + \beta \mathbf{res}_w B(w) w^{-k-1} = \mathbf{res}_w \big((\alpha A(w) + \beta B(w)) w^{-k-1} \big). \tag{18}$$

RULE 3 *(Substitution).*
(a) For $f(w) \in L_k$ ($k \geqslant 1$) and $A(w)$ any element of L, or
(b) for $A(w)$ a polynomial and $f(w)$ any element of L including a constant

$$\sum_k f(w)^k \mathbf{res}_z \big(A(z) z^{-k-1} \big) = A\big(f(w)\big). \tag{19}$$

RULE 4 *(Inversion).* For $f(w)$ from L_0

$$\sum_k z^k \mathbf{res}_w \big(A(w) f(w)^k w^{-k-1} \big) = \big[A(w)/f(w) h'(w) \big]_{w=\overline{h(z)}}, \tag{20}$$

where $z = h(w) = wf(w) \in L_1$.

[2] In the combinatorial literature the phrase "Cauchy algebra of formal power series" is often used for the same purpose [176].

RULE 5 *(Change of variables).* If $f(w) \in L_0$, then

$$\mathbf{res}_w\big(A(w)f(w)^k w^{-k-1}\big) = \mathbf{res}_z\big(\big[A(w)/f(w)h'(w)\big]_{w=\overline{h(z)}} z^{-k-1}\big), \tag{21}$$

where $z = h(w) = wf(w) \in L_1$.

RULE 6 *(Differentiation).*

$$k\mathbf{res}_w A(w)w^{-k-1} = \mathbf{res}_w A'(w)w^{-k}. \tag{22}$$

3.2. *Connection between the theory of analytic functions and the lemma of completeness*

If a formal power series $A(w) \in L$ converges in a punctured neighborhood of zero, then the definition of $\mathbf{res}_w A(w)$ coincides with the usual definition of $\mathbf{res}_{w=0} A(w)$, used in the theory of analytic functions. The formula (16) is an analog of the Cauchy integral formula $a_k = \frac{1}{2\pi i}\oint_{|w|=\varrho} A(w)w^{-k-1}\,dw$ for the coefficients of the Taylor series in a punctured neighborhood of zero. The substitution rule of the **res** operator is direct analog of the famous Cauchy theorem. Similarly, it is possible to introduce the definition of a formal residue at the point of infinity, the logarithmic residue and the residue theorem (all necessary concepts and results in the residue theory of one and several complex variables, see [204,15,152,59,189]). Moreover, it is easy to see that rules of the **res** operator can be simply proved by reduction to known formulae in the theory of residues for corresponding polynomials [59].

In solving analytic problems with the help of generating functions the investigator usually encounters one of two interconnected problems.

PROBLEM A. Suppose that the series $S(w) = \sum_k s_k w^k$ from L is expressed in terms of the series $A(w) = \sum_k a_k w^k$, $B(w) = \sum_k b_k w^k, \dots, D(w) = \sum_k d_k w^k$ from L with the help of operations on Laurent formal power series over $\mathbb{C}$, i.e. a formula

$$S(w) = f\big(A(w), B(w), \dots, D(w)\big) \tag{23}$$

is given. For each k find a formula

$$s_k = f\big(\{a_k\}, \{b_k\}, \dots, \{d_k\}\big) \tag{24}$$

for the terms of sequence $\{s_k\}$ in terms of the terms of the sequences $\{a_k\}, \{b_k\}, \dots, \{d_k\}$.

DEFINITION 1. A sequence $\{s_k\}$ is called of A-*type*, if it is determined by a formula (24).

Consider, for example, the following substitution. Let $A(w) = \sum_k a_k w^k \in L_0$, $B(w) = \sum_k b_k w^k \in L_1$, and

$$A\big(B(w)\big) = D(w), \quad D(w) = \sum_k d_k w^k \in L. \tag{25}$$

Then

$$\begin{aligned}
d_n &= \mathbf{res}_w\{A\big(B(w)\big)w^{-n-1}\} = \mathbf{res}_w\left\{\sum_{k=0}^{\infty} a_k\left(\sum_{i=1}^{\infty} b_i w^i\right)w^{-n-1}\right\}\\
&= \mathbf{res}_w\left\{\sum_{k=0}^{n} a_k\left(\sum_{i=1}^{n} b_i w^i\right)w^{-n-1}\right\}\\
&= \sum_{k=0}^{n} a_k b_1^k \mathbf{res}_w\left\{\left(1+\sum_{i=2}^{n} b_i b_1^{-1} w^{i-1}\right)^k w^{-n+k-1}\right\}\\
&\quad \text{(by polynomial expansion)}\\
&= \sum_{k=0}^{n} a_k b_1^k \sum_{\substack{k_2+\cdots+(n-1)k_n=n-k,\\ k_2,\ldots,k_n=0,1,\ldots}} \binom{k}{k_2,\ldots,k_n}\prod_{i=2}^{n}(b_i/b_1)^{k_i},
\end{aligned}$$

i.e.

$$\begin{aligned}
d_n &= \mathbf{res}_w\{A\big(B(w)\big)w^{-n-1}\}\\
&= \sum_{k=0}^{n} a_k b_1^k \sum_{\substack{k_2+\cdots+(n-1)k_n=n-k,\\ k_2,\ldots,k_n=0,1,\ldots}} \binom{k}{k_2,\ldots,k_n}\prod_{i=2}^{n}(b_i/b_1)^{k_i},\\
&\quad n = 0, 1, 2, \ldots. \qquad (26)
\end{aligned}$$

Problem B. For each k a formula $s_k = f(\{a_k\}, \{b_k\}, \ldots, \{d_k\})$ for the terms of the sequence $\{s_k\}$ of A-type is given. Find a formula $S(w) = f(A(w), B(w), \ldots, D(w))$ for generating functions of the sequences $\{s_k\}, \{a_k\}, \{b_k\}, \ldots, \{d_k\}$.

Definition 2. A set of rules for **res** operator is called *complete*, if it allows one to solve Problem B.

Lemma 1 *(On completeness [64]). The set of Rules* 1–6 *for the* **res** *operator is complete.*

Proof. We will use induction on the number of sequence operations in formula (24) generating the given sequence $\{s_k\}$. At the first step of induction the series $S(w)$ is obtained with the help of series $A(w)$ and $B(w)$ from L by one operation over Laurent formal power series (addition, multiplication, etc.). We should give the solution to recursive relations that corresponds to each of the operations. These calculations were made in [59, pp. 31–35], and we present here only one of them. For example, by formula (26) for substitution operation we have:

$$d_n = \sum_{k=0}^{n} a_k b_1^k \sum_{\substack{k_2+\cdots+(n-1)k_n=n-k,\\ k_2,\ldots,k_n=0,1,\ldots}} \binom{k}{k_2,\ldots,k_n}\prod_{i=2}^{n}(b_i/b_1)^{k_i}, \quad n = 0, 1, 2, \ldots.$$

As

$$\binom{k}{k_2,\dots,k_n} = \mathbf{res}_{z_2,\dots,z_n}(1+z_2+\cdots+z_n)^k \prod_{i=2}^{n} z^{-k_i-1},$$

it follows that

$$d_n = \sum_{k=0}^{n} a_k b_1^k \Biggl\{ \sum_{k_2,\dots,k_n=0,1,\dots} \mathbf{res}_{w,z_2,\dots,z_n} \\ \Biggl((1+z_2+\cdots+z_n)^k w^{-n+k+k_2+\cdots+k_n-1} \prod_{i=2}^{n} (b_i/b_1)^{k_i} z^{-k_i-1} \Biggr) \Biggr\},$$

where the operator $\mathbf{res}_w w^{-n+k+k_2+\cdots+k_n-1}$ is added to take care of the linear restriction

$$k_2+\cdots+(n-1)k_n = n-k$$

on the summation indexes $k_2,\dots,k_n$. By the substitution rule for the **res** operator and the following change of variables $z_2 = wb_2/b_1,\dots,z_n = wb_n/b_1$ we get

$$\begin{aligned} d_n &= \sum_{k=0}^{n} a_k \mathbf{res}_w (b_1 w^1 + b_2 w^2 + \cdots + b_n w^n)^k w^{-n-1} \\ &= \sum_{k=0}^{\infty} a_k \mathbf{res}_w \left(\left(\sum_{j=1}^{\infty} b_j w^j \right)^k w^{-n-1} \right) = \mathbf{res}_w \left\{ \left(\sum_{k=0}^{\infty} a_k B^k(w) \right) w^{-n-1} \right\} \\ &= \mathbf{res}_w \bigl\{ A\bigl(B(w)\bigr) w^{-n-1} \bigr\}, \quad n = 0,1,2,\dots, \end{aligned}$$

i.e.,

$$D(w) = A\bigl(B(w)\bigr).$$

If the statement of the lemma is valid for $n-1$ operations, then the next inductive step is similar to the initial step. □

REMARK 2. This lemma supports the thesis according to which it should be possible to find with the help of the method of coefficients an operational (integral) representation for those sums, which can be calculated via Laurent formal power series.

3.3. *Examples*

The first example is illustrative.

3.3.1. *Computation of the Halmos sum*

$$S = S_{n,m} = \sum_{k=0}^{m} (-1)^k \binom{n}{k}, \quad m \leqslant n.$$

SOLUTION. As $\binom{n}{k} = \mathbf{res}_w (1+w)^n w^{-k-1}$, according to the general scheme of the method of coefficients we have

$$\begin{aligned}
S &= \sum_{k=0}^{m} (-1)^k \mathbf{res}_w (1+w)^n w^{-k-1} \\
&\quad \text{(by the linearity rule)} \\
&= \mathbf{res}_w \left\{ (1+w)^n \left[\sum_{k=0}^{m} (-1)^k w^{-k-1} \right] \right\} \\
&\quad \text{(by the formula of geometric progression)} \\
&= \mathbf{res}_w \left\{ (1+w)^n w^{-1} \bigl(1-(-w)^{-m-1}\bigr)\bigl(1-(-w)^{-1}\bigr)^{-1} \right\} \\
&\quad \text{(after division by } (1+w)\text{)} \\
&= \mathbf{res}_w \left\{ (1+w)^{n-1} \bigl(1+(-1)^m w^{-m-1}\bigr) \right\} \\
&\quad \text{(by the linearity rule)} \\
&= \mathbf{res}_w (1+w)^{n-1} + (-1)^m \mathbf{res}_w \left\{ (1+w)^{n-1} w^{-m-1} \right\} \\
&\quad \text{(by the definition of the } \mathbf{res} \text{ operator)} \\
&= 0 + (-1)^m \binom{n-1}{m} = (-1)^m \binom{n-1}{m}.
\end{aligned}$$

Thus we have a new proof of the well-known Halmos identity [101]

$$\sum_{k=0}^{m} (-1)^k \binom{n}{k} = (-1)^m \binom{n-1}{m}, \quad m \leqslant n.$$

3.3.2. *Calculation of the Lyamin–Selivanov sum from graph theory* Compute the following multiple sum [134] of Stirling numbers

$$\begin{aligned}
S = \sum_{a=0}^{n-k} \sum_{d=0}^{a} \sum_{p=0}^{a} \sum_{c=d}^{p+d} (-1)^{p+a-c} \binom{k-d}{a-d} \binom{d}{p+d-c} \\
\times \frac{n! n^{n-k-a-1}}{(k-d)!(n-k-p)!c!} \left\{ {c \atop d} \right\} \left\{ {n-k-p \atop n-k-a} \right\}.
\end{aligned}$$

SOLUTION. Since

$$\binom{a}{d} = \mathbf{res}_w (1+w)^a w^{-d-1}, \qquad \binom{d}{p+d-c} = \mathbf{res}_u (1+u)^d w^{-p-d+c-1},$$

$$\left\{ {c \atop d} \right\} = \frac{c!}{d!} \mathbf{res}_x \bigl(e^x - 1\bigr)^d x^{-c-1},$$

$$\left\{ {n-k-p \atop n-k-a} \right\} = \frac{(n-k-p)!}{(n-k-a)!} \mathbf{res}_y \bigl(e^y - 1\bigr)^{n-k-a} y^{-n+k+p-1},$$

applying three times the substitution rule we have successively

$$\begin{aligned}
S = {} & \sum_{a,d,p,c} (-1)^a \frac{n!n^{n-k-a-1}}{(n-k-a)!k!}\binom{k}{a}\Bigg\{(-1)^{p-c}\binom{a}{d}\binom{d}{p+d-c} \\
& \times \mathbf{res}_x \big(e^x-1\big)^d x^{-c-1} \times \mathbf{res}_y \big(e^y-1\big)^{n-k-a} y^{-n+k+p-1}\Bigg\} \\
= {} & \sum_{a,d,p,c} (-1)^a \frac{n!n^{n-k-a-1}}{(n-k-a)!k!}\binom{k}{a}(-1)^{p-c} \\
& \times \mathbf{res}_{xywu} \big(e^x-1\big)^d x^{-c-1} \big(e^y-1\big)^{n-k-a} y^{-n+k+p-1} \\
& \times (1+u)^d u^{-p-d+c-1} (1+w)^a w^{-d-1} \\
= {} & \sum_{a,d,p} \Bigg\{ \dots \Bigg\{ \sum_{c=0}^{\infty} (-u)^c \mathbf{res}_x \big(e^x-1\big)^d x^{-c-1} \Bigg\}\Bigg\} \\
& \text{(change of variables } x=-u \in L_1) \\
= {} & \sum_{a,d} \Bigg\{ \dots \Bigg\{ \sum_{p=0}^{\infty} (-y)^p \mathbf{res}_u (1+u)^d \big(e^{-u}-1\big)^d u^{-p-d-1} \Bigg\}\Bigg\} \\
& \text{(change of variables } u=-y \in L_1) \\
= {} & \sum_{a} \Bigg\{ \dots \Bigg\{ \sum_{d=0}^{\infty} \big((1-y)\big(1-e^y\big)/y\big)^d \mathbf{res}_w (1+w)^a w^{-d-1} \Bigg\}\Bigg\} \\
& \text{(change of variables } w=(1-y)(1-e^y)/y) \\
= {} & \sum_{a} (-1)^a \frac{n!n^{n-k-a-1}}{(n-k-a)!k!}\binom{k}{a} \\
& \times \mathbf{res}_y \big\{\big(e^y-1\big)^{n-k-a}\big(y-(1-y)\big(e^y-1\big)\big)^a y^{-n+k-a-1}\big\}.
\end{aligned}$$

It is easy to see that the **res** with respect to y is equal to $(1/2)^a$. Thus we get a new identity

$$\begin{aligned}
& \sum_{d=0}^{a}\sum_{p=0}^{a}\sum_{c=d}^{p+d}(-1)^{p+c}\binom{a}{d}\binom{d}{p+d-c}\frac{d!(n-k-a)!}{c!(n-k-p)!}\left\{ c \atop d \right\}\left\{ n-k-p \atop n-k-a \right\} \\
& \quad = (1/2)^a, \quad a=0,1,\dots,n-k.
\end{aligned}$$

3.3.3. *Sums with linear constraints on the summation indices* It is not uncommon to meet multidimensional sums with constraints on the summation indices when solving enumeration problems, especially in graph theory. In order to take into account linear constraints on the summation indices the well-known trick of introducing extra weighting variables or δ-functions can be used (see also Section 6.1). This enables us, by increasing the dimension of the integrals used, to pass to the solution of a summation problem with independent summation indices. The following result, obtained with the help of this approach, is given in [58].

THEOREM 2. *Let*

$$S_m = -\sum_{J\in H_{nm}(A)} \frac{(j_1+\cdots+j_n-1)!}{j_1!\ldots j_n!}\prod_{r=1}^{n}(-\beta_r)^{j_r}, \tag{27}$$

where $m = (m_1,\ldots,m_k)$, $A = (a_{rl})$ is an $n\times k$ matrix, $m_1,\ldots,m_k$ and a_{rl} are positive integers, $H_{nm}(A) = \{J = (j_1,\ldots,j_n)\colon j_1,\ldots,j_n = 0,1,\ldots$ and $JA \leqslant m\}$, and $\beta_1,\ldots,\beta_n$ are complex numbers. Then the series $\sum_{m_1,\ldots,m_k=1}^{\infty} S_{m_1,\ldots,m_k} t_1^{m_1}\ldots t_k^{m_k}$ in the variables $(t_1,\ldots,t_k) = t$, which is the generating function for the numbers $S_{m_1,\ldots,m_k}$, $m_1,\ldots,m_k = 1,2,\ldots$, converges in some neighborhood of the origin, and its sum is

$$F(t) = F(t_1,\ldots,t_k) = \ln f(t_1,\ldots,t_k)\prod_{l=1}^{k}(1-t_l)^{-1}, \tag{28}$$

where $f(t_1,\ldots,t_k) = 1+\sum_{r=1}^{n}\beta_r\prod_{l=1}^{k}t_l^{a_{rl}}$, and $\ln x$ is understood to be the principal branch of the logarithm.

From (27), (28) with $k = 1$ we get the following:

COROLLARY 1. *Suppose that m and $\alpha_1,\ldots,\alpha_n$ are positive integers, with $\alpha_1 \leqslant \alpha_2 \leqslant \cdots \leqslant \alpha_n$,*

$$K_{nm}(\alpha_1,\ldots,\alpha_n) = \big\{J = (j_1,\ldots,j_n)\colon j_1,\ldots,j_n = 0,1,\ldots;\ j_1,\ldots,j_n\colon \\ \alpha_1 j_1+\cdots+\alpha_n j_n = m\big\},$$

q and $\beta_1,\ldots,\beta_n$ are complex numbers, p is a positive real number, B_m $(m = 0,1,\ldots)$ are the Bernoulli numbers, and $(\beta)^j$ is taken equal to 1 when $\beta = j = 0$. Then the following identities hold:

$$\sum_{J\in K_{nm}(\alpha_1,\ldots,\alpha_n)} \frac{(j_1+\cdots+j_n-1)!}{j_1!\ldots j_n!}\prod_{r=1}^{n}(-\beta_r)^{j_r} = \frac{1}{m}\sum_{r=1}^{\alpha_n} x_r^{-m},$$

where x_r, $r = 1,\ldots,\alpha_n$, are the roots of the equation $1+\sum_1^n \beta_r x^{\alpha_r} = 0$;

$$\sum_{J\in K_{nm}(1,\ldots,n)} \frac{(j_1+\cdots+j_n-1)!}{j_1!\ldots j_n!}\prod_{r=1}^{n}\left\{-q^r\binom{p+r-1}{r}\right\}^{j_r} = -\frac{p}{m}q^m,$$

$m = 1,\ldots,n$;

$$\sum_{J\in K_{nm}(1,\ldots,n)} \frac{(j_1+\cdots+j_n-1)!}{j_1!\ldots j_n!}\prod_{r=1}^{n}\left\{-q^r\binom{n}{r}\right\}^{j_r} = \frac{n}{m}(-q)^m,$$

$m = 1,\ldots,n$;

$$\sum_{J\in K_{nm}(1,\ldots,n)} \frac{(j_1+\cdots+j_n-1)!}{j_1!\ldots j_n!}\prod_{r=1}^{n}\{-q^r/r!\}^{j_r} = \begin{cases}-q, & m = 1,\\ 0, & m = 2,\ldots,n.\end{cases}$$

For $m = 1, \dots, n$, $\beta_r = (-1)^{r/2}/r!$ *for* $r = 2, 4, \dots$, *and* $\beta_r = 0$ *for* $r = 1, 3, \dots$ *the identity*

$$\sum_{J \in K_{nm}(1,\dots,n)} \frac{(j_1 + \cdots + j_n - 1)!}{j_1! \dots j_n!} \prod_{r=1}^{n} (-\beta_r)^{j_r} = \begin{cases} 0, & \textit{for odd } m, \\ \frac{(-1)^{m/2+1} 2^m (2^m - 1) B_m}{m \cdot m!}, & \textit{for even } m. \end{cases}$$

is valid.

For $m = 1, \dots, n$, $\beta_1 = q$, $\beta_{2k+1} = q(q^2 - 1^2)(q^2 - 3^2) \cdots (q^2 - (2k-1)^2)/(2k+1)!$ *and*

$$\beta_{2k+2} = q^2 \bigl(q^2 - 2^2\bigr) \cdots \bigl(q^2 - (2k)^2\bigr)/(2k+2)!, \quad k = 1, 2, \dots,$$

the identity

$$\sum_{J \in K_{nm}(1,\dots,n)} \frac{(j_1 + \cdots + j_n - 1)!}{j_1! \dots j_n!} \prod_{r=1}^{n} (-\beta_r)^{j_r} = \begin{cases} 0, & \textit{for even } m, \\ \frac{(-1)^{\lfloor m/2 \rfloor + 1} q (m-1)!}{m 2^{m-1} (\lfloor m/2 \rfloor !)^2}, & \textit{for odd } m, \end{cases}$$

is valid.

For $m = 1, \dots, n$, $\beta_r = 0$ *for* $r = 1, 3, \dots$, *and* $\beta_r = |E_r|/r!$ *for* $r = 2, 4, \dots$ *the identity*

$$\sum_{J \in K_{nm}(1,\dots,n)} \frac{(j_1 + \cdots + j_n - 1)!}{j_1! \dots j_n!} \prod_{r=1}^{n} (-\beta_r)^{j_r} = \begin{cases} 0, & \textit{for odd } m, \\ \frac{(-1)^{m/2} 2^m (2^m - 1) B_m}{m \cdot m!}, & \textit{for even } m, \end{cases}$$

is valid for the Euler numbers E_m, $m = 0, 1, \dots$.

The Sheehan identity [191]

$$\sum_{J \in K_{nn}(1,\dots,n)} (-1)^{j_1 + \cdots + j_n - 1} \frac{(j_1 + \cdots + j_n - 1)!}{j_1! \dots j_n!} = 1/n$$

follows from (27), (28) *with* $k = 1$, $m = n$, *and* $\beta_r = 1$ *and* $\alpha_r = r$ *for* $r = 1, \dots, n$.

4. Some applications in combinatorial analysis and group theory

4.1. *Solution of the Riordan problem (1968) on the characterization of known pairs of inverse combinatorial relations and their algebraic characterization*

In 1968 J. Riordan posed the problem of characterizing of known pairs of inverse combinatorial relations of the form

$$a_m = \sum_{k=0}^{\infty} c_{mk} b_k, \qquad b_m = \sum_{k=0}^{\infty} d_{mk} a_k, \quad m = 0, 1, 2, \ldots, \tag{29}$$

where $C = (c_{ij})$ is an invertible infinite lower triangular matrix whose general term is a linear combination of known combinatorial numbers, and $D = (d_{ij})$ is its inverse ([176, Introduction], [175]). Each pair of relations of this form generates a combinatorial identity

$$\sum_{k} c_{mk} d_{kn} = \delta(m, n), \quad m, n = 0, 1, 2, \ldots, \tag{30}$$

where $\delta(m, n)$ is the Kronecker symbol.

A large part of Riordan monograph [176] on combinatorial identities is concerned with pairs of inverse relations with binomial coefficients in the one-dimensional case. The first complete solution to the Riordan problem was given in 1973–1974 [53,54,59] by studying properties of a special type of matrices, defined by a certain integral construction and a certain 5-tuple $F = (\{\alpha_m\}, \{\beta_k\}, \varphi(x), f(x), \psi(x))$, where $\{\alpha_m\}_{m=0,1,2,\ldots}$, $\{\beta_k\}_{k=0,1,2,\ldots}$ are sequences of non-zero numbers and $\varphi(x)$, $f(x)$, $\psi(x)$ are Laurent formal power series over $\mathbb{C}$. In [70] we presented a new solution to Riordan's problem, gave a series of the new algebraic results, and demonstrated how the integral representation can be used in a unified approach for generating new types of combinatorial identities. In conclusion we compared these results with other known classification approaches.

Let $A(w) = \sum_k a_k w^k$ be a generating function for the sequence $\{a_k\}$.

Definition 3. We say that a matrix $C = (c_{mk})_{m,k=0,1,2,\ldots}$ as in (29) is of type E or $E^q(\{\alpha_m\}, \{\beta_k\}; \varphi, f, \psi)$ if its general term is defined by the formula

$$c_{mk} = \frac{\beta_k}{\alpha_m} \mathbf{res}_z \big(\varphi(z) f^k(z) \psi^m(z) z^{-m+qk-1}\big),$$

where q is a positive integer, $\alpha_m, \beta_k \neq 0$; $\varphi(z)$, $f(z)$, $\psi(z) \in L_0$.

In particular, for $q = 1$, the matrix (c_{mk}) is infinite lower triangular with general term

$$c_{mk} = \frac{\beta_k}{\alpha_m} \mathbf{res}_z \big(\varphi(z) f^k(z) \psi^m(z) z^{-m+k-1}\big). \tag{31}$$

The relations (29) are completely defined by the matrix $C = (c_{mk})_{m,k=0,1,2,\ldots}$. For this reason, we attach the type of this matrix to the relation, and use terms as "a relation of type E" when necessary. Occasionally,we will omit the superscript in the type (for example, when $q = 1$ or when it is not important in the context).

For example, the binomial coefficients $\binom{n}{k}$, $n, k = 0, 1, 2, \ldots$, admit integral representations of the following types:

(a) type $E(\{1\}, \{1\}; (1+w), 1, 1)$:

$$\binom{n}{k} = \mathbf{res}_w (1+w)^n w^{-n+k-1};$$

$$\binom{n}{k} = \mathbf{res}_w (1+w)^n w^{-k-1}, \qquad \binom{n}{k} = \mathbf{res}_w (1-w)^{-n+k-1} w^{-k-1};$$

(b) type of $E(\{1\}, \{1\}; 1, (1-w)^{-1}, (1-w)^{-1})$:

$$\binom{n}{k} = \mathbf{res}_w (1-w)^{-k-1} w^{-n+k-1};$$

(c) ordinary:

$$\binom{n}{k} = \mathbf{res}_w (1+w)^n w^{-k-1}, \qquad \binom{n}{k} = \mathbf{res}_w (1-w)^{-n+k-1} w^{-k-1}.$$

4.1.1. *Main results*

LEMMA 2. *A matrix (c_{mk}) of the type $E(\{\alpha_m\}, \{\beta_k\}; \varphi(z), f(z), \psi(z))$ can be uniquely represented as a matrix of type $E(\{\alpha_m\}, \{\beta_k\}; z\bar{h}'(z)\varphi(\bar{h}(z))/\bar{h}(z), 1, \psi(\bar{h}(z))\bar{h}(z)/z)$, or as a matrix of the type $E(\{\alpha_m\}, \{\beta_k\}; z\bar{l}'(z)\varphi(\bar{l}(z))/\bar{l}(z), f(\bar{l}(z))\bar{l}(z)/z, 1)$, where $h(z) = zf(z)$, $l(z) = z/\psi(z)$.*

THEOREM 3 *(On inverses [59]). Relations of the type E^1 are equivalent to functional relations between the series $\widetilde{A}(w) = \sum_{m\geqslant 0} \alpha_m a_m w^m$ and $\widetilde{B}(w) = \sum_{m\geqslant 0} \beta_m b_m w^m$ as follows*:

$$\widetilde{A}\big(l(w)\big)l'(w)\psi(w) = \varphi(w)\widetilde{B}\big(h(w)\big).$$

The matrix (d_{mk}), the inverse of the matrix (c_{mk}) of type E^1 exists, is unique, is of the type E^1, and has the following general term

$$d_{mk} = \frac{\alpha_k}{\beta_m} \mathbf{res}_w \big(\varphi^{-1}(w) l'(w) h'(w) f^{-m-1}(w) \psi^{-k-1}(w) w^{-m+k-1}\big). \tag{32}$$

THEOREM 4 *(On classification [53,54,59]). The pairs of inverse relations of simplest type, of Gould type, of Tchebycheff type, of Legendre type, of Legendre–Tchebycheff type, of Abel type, of ordinary and exponential types* [176, Tables 2.1–2.5, 3.1–3.3] *and of Lagrange type* [176, Chapter 4] *all belong to the type $E^q(\{\alpha_m\}, \{\beta_k\}; \varphi, f, \psi)$.*

THEOREM 5 *(On combinatorial examples). Matrices of binomial coefficients, Stirling numbers (usual and generalized) of the first and second kind, the Lax numbers and many others numbers* [193] *belong to the type $E^q(\{\alpha_m\}, \{\beta_k\}; \varphi, f, \psi)$.*

THEOREM 6 *(On probabilistic examples [59]). A number of well known one and multidimensional discrete probability distributions, viz.: Poisson and negative binomial distributions (usual and generalized), Borel–Tanner distribution, Haight distribution, general Consul–Shenton discrete distribution* [79,40–43] *all belong to the type $E^q(\{\alpha_m\}, \{\beta_k\}; \varphi, f, \psi)$.*

THEOREM 7 *(On products [59]). Let the sequence* $\{\alpha_m\}$, $\alpha_m \neq 0$, *be fixed. Then the product of two matrices* $A = (a_{mk})$ *and* $B = (b_{mk})$ *of type* $E(\{\alpha_m\}, \{\alpha_k\}; \varphi, f, \psi)$ *is a matrix the same type.*

THEOREM 8 *(On decomposition [59]). A matrix* (d_{mk}) *of the type* $E(\{\alpha_m\}, \{\beta_k\}; \varphi, f, \psi)$ *splits into the product of three matrices* (a_{mk}), (b_{mk}) *and* (c_{mk}) *of the following types*: $E(\{\alpha_m\}, \{\gamma_k\}; \varphi_1, 1, \psi)$, $E(\{\gamma_m\}, \{\zeta_k\}; \varphi_2, 1, 1)$ *and* $E(\{\zeta_m\}, \{\beta_k\}; \varphi_3, f, 1)$, *where* $\varphi = \varphi_1\varphi_2\varphi_3$, $\varphi_1(0)\varphi_2(0)\varphi_3(0) \neq 0$ *and the sequences of non-zero numbers* $\{\gamma_m\}$, $\{\zeta_k\}$ *are arbitrary.*

PROOF. We have

$$\begin{aligned}
d_{mk} &= \sum_{r=0}^{m}\sum_{s=0}^{r} a_{mr}b_{rs}c_{sk}\\
&= \sum_{r=0}^{m}\sum_{s=0}^{r} \frac{\gamma_r}{\alpha_m}\mathbf{res}_w\varphi_1(w)\psi^m(w)w^{-m+r-1} \times \frac{\zeta_s}{\gamma_r}\mathbf{res}_v\varphi_2(v)v^{-r+s-1}\\
&\quad \times \frac{\beta_k}{\zeta_s}\mathbf{res}_u\varphi_3(u)f^k(u)u^{-s+k-1}\\
&= \sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\cdots\\
&\quad \text{(for } r > m, \text{ or } s > r \text{ each added term of the sum is equal 0 according}\\
&\quad \text{to definition of the } \mathbf{res} \text{ operator)}\\
&= \frac{\beta_k}{\alpha_m}\mathbf{res}_w\Bigg\{\varphi_1(w)\psi(w)^m w^{-m-1}\Bigg[\sum_{r=0}^{\infty} w^r\,\mathbf{res}_v\varphi_2(v)v^{-r-1}\\
&\quad \times\Bigg(\sum_{s=0}^{\infty} v^s\,\mathbf{res}_u\varphi_3(u)f^k(u)u^{-s+k-1}\Bigg)\Bigg]\Bigg\}.
\end{aligned}$$

To complete the proof it suffices to sum over s and r by the substitution rule for the **res** operator in the variables u and v with variable changes $u = v$ and $v = w$, respectively. □

REMARK 3. In [59] this theorem was proved for $\gamma_m = \zeta_k = 1$ and $\varphi_1 = \varphi$, $\varphi_2 = \varphi_3 = 1$. Although the statement of Theorem 5 is stronger, the scheme of the proof remains almost the same. The presence of new weighting coefficients γ_m and $\zeta_k, m, k = 0, 1, 2, \ldots,$ and the representation $\varphi = \varphi_1\varphi_2\varphi_3$ gives algebraic completeness to the formulation of the theorem on decomposition. It also allows us to formulate new results on algebraic and probabilistic characterizations of pairs of inverse relations and introduces new objects (methods), such as e.g. the Lagrange summation matrix below.

THEOREM 9 *(On algebraic characterization [70]). Let a sequence of nonzero numbers* $\{\alpha_m\}$ *be fixed. Then*:

(a) *The set of all matrices of the type* $E(\{\alpha_m\}, \{\alpha_k\}; \varphi, f, \psi)$ *forms a subgroup of the group* $TN(C)$ *of all lower triangular matrices. If* $\varphi(0) = f(0) = \psi(0) = 1$ *then the group* $E(\{\alpha_m\}, \{\alpha_k\}; \varphi, f, \psi)$ *is unitriangular.*
(b) *The groups* $E(\{\alpha_m\}, \{\alpha_k\}; \varphi, 1, \psi)$ *and* $E(\{\alpha_m\}, \{\alpha_k\}; \varphi, f, 1)$ *are isomorphic.*
(c) *Every set of all matrices of type* $E(\{\alpha_m\}, \{\alpha_k\}; \varphi, 1, 1)$, $E(\{\alpha_m\}, \{\alpha_k\}; 1, f, 1)$ *or* $E(\{\alpha_m\}, \{\alpha_k\}; 1, 1, \psi)$ *forms a subgroup of the group* $E(\{\alpha_m\}, \{\alpha_k\}; \varphi, f, \psi)$.
(d) *The subgroups* $E(\{\alpha_m\}, \{\alpha_k\}; \varphi, 1, 1)$, $E(\{\alpha_m\}, \{\alpha_k\}; 1, f, 1)$ *and* $E(\{\alpha_m\}, \{\alpha_k\}; 1, 1, \psi)$ *pairwise have only the element* $I = (\delta(m,k))_{m,k=0,1,2,\dots}$ *in common.*
(e) *The group* $E(\{\alpha_m\}, \{\alpha_k\}; \varphi, f, \psi)$ *is a product of its subgroups*

$$E\bigl(\{\alpha_m\}, \{\alpha_k\}; \varphi, 1, 1\bigr), \qquad E\bigl(\{\alpha_m\}, \{\alpha_k\}; 1, f, 1\bigr) \quad \text{and}$$
$$E\bigl(\{\alpha_m\}, \{\alpha_k\}; 1, 1, \psi\bigr).$$

DEFINITION 4. We say that the matrix (31) is a matrix summation method of type $E(\{a_m\}, \{b_k\}; \varphi, f, \psi)$, if all c_{mk} are nonnegative and $\lim_{m\to\infty} c_{mk} = 0$ for any k. We call the matrix summation methods of types $E(\{a_m\}, \{c_k\}; 1, 1, \psi)$, $E(\{c_m\}, \{d_k\}; \varphi, 1, 1)$ and $E(\{d_m\}, \{b_k\}; 1, f, 1)$ the summation methods of Lagrange, of Voronoy and analytic, respectively.

THEOREM 10. *Let* $A(w)$, $B(w)$, $C(w)$ *and* $D(w)$ *be generating functions for the sequences* $\{a_k\}, \{b_k\}, \{c_k\}$ *and* $\{d_k\}$, *respectively. For a matrix summation method of type E to be regular* [102], *it is necessary and sufficient that* $A(l(w))l'(w)\psi(w) = \varphi(w)B(h(w))$. *Similarly, the summation methods of Lagrange, Voronoy and analytic will be regular, if and only if respectively* $\psi(w)l'(w)A(l(w)) = C(w)$, $C(w) = \varphi(w)D(w)$ *and* $D(w) = B(h(w))$.

THEOREM 11 *(On functional-theoretical characterization [59]). The well-known matrix summation methods for divergent series of de la Vallée-Poussin, Obreshkov, Cesàro, Euler,* $P(q, r, s)$, *and the general methods of Lagrange, Voronoy, analytic, Gronwall, etc. are all particular cases of regular methods of type* E. *A regular summation method of type E splits into the product of regular summation methods of Lagrange, Voronoy and analytic.*

REMARK 4. The Lagrange summation method is introduced here for the first time. The classic Gronwall, Voronoy and analytic summation methods of divergent series are methods of the following types: $E(\{b_m\}, \{1\}; g(w)(1 - wf(w)), f(w), 1)$, $E(\{b_m\}, \{1\}; g(w)(1 - w), 1, 1)$ and $E(\{1\}, \{1\}; 1 - wf(w), f(w), 1)$, where $g(w), f(w) \in L_0$ and $g(w) = 1 + \sum_{m=1}^{\infty} b_m w^m$. The last statement of Theorem 11 is an extension of a known result in divergent summation theory that the Gronwall matrix splits into the product of matrices of summing divergent series of Voronoy and analytic types.

4.1.2. *Riordan arrays and Riordan group* A particular case of matrices of type E (see Theorem 12) and some results of the previous section appear also in the concept of the Riordan group and Riordan array of regular and exponential type. These concepts were introduced in 1991 by Shapiro et al. [190]. The group is quite easily described but unifies many themes in enumeration, including a generalized concept of a renewal array defined by Rogers in 1978 [195,20,147]. Their basic idea was to define a class of infinite lower triangular arrays with properties analogous to those of the Pascal triangle.

A Riordan array is an infinite lower triangular array $D = \{d_{nk}\}_{n,k\in\mathbb{N}}$, defined by a pair of formal power series $d(t), h(t) \in L_0$, such that the generic element is the nth coefficient in the series $d(t)(th(t))^k$:

$$d_{nk} = \left[t^n\right]d(t)\bigl(th(t)\bigr)^k, \quad n,k = 0,1,2,\ldots, \qquad d_{nk} = 0 \quad \text{for } k > n. \tag{33}$$

Here it is always assumed that $d(0) \neq 0$; if we also have $h(0) \neq 0$ then the Riordan array is said to be *proper*; in the proper-case the diagonal elements d_{nn} are different from zero for all $n \in \mathbb{N}$. Proper Riordan arrays are characterized by the following basic property [195], that immediately follows from Lemma 2: a matrix $\{d_{nk}\}_{n,k\in\mathbb{N}}$ is a proper Riordan array iff there exists a sequence $A = \{a_i\}_{i\in\mathbb{N}}$ with $a_0 \neq 0$ s.t. every element $d_{n+1,k+1}$ can be expressed as a linear combination with coefficients in A of elements in the preceding row, starting from the preceding column, i.e.

$$d_{n+1,k+1} = a_0 d_{n,k} + a_1 d_{n,k+1} + a_2 d_{n,k+2} + \cdots.$$

The Riordan group is the set of infinite lower triangular matrices of type (33). Shapiro [190] and others often denote a Riordan matrix D by $D = (g(x), f(x))$, $g(x) \in L_0$, $f(x) \in L_1$. We denote by R^* the set of Riordan matrices. R^* is a group under matrix multiplication with the following properties: $(g(x), f(x)) * (u(x), v(x)) = (g(x)u(f(x)), \underline{v(f(x))})$, $I = (1, x)$ is the identity element. The inverse of D is given by $D^{(-1)} = (1/g(\overline{f}(x)), \overline{f}(x))$.

It was shown in 2000 [161] that the elements of the Riordan group of the form $(xf'(x)/f(x), f(x))$ form a subgroup denoted by PW, as proved by the following:

(i) The identity $(1, x) = x(x)'/x, x) \in PW$.

(ii) The product

$$\begin{aligned}
&\bigl(xf'(x)/f(x), f(x)\bigr) * \bigl(xh'(x)/h(x), h(x)\bigr)\\
&\quad = \left(\frac{xf'(x)}{f(x)}\frac{f(x)h'(f(x))}{h(f(x))}, h\bigl(f(x)\bigr)\right)\\
&\quad = \left(\frac{x(h(f(x)))'}{h(f(x))}, h\bigl(f(x)\bigr)\right) \in PW.
\end{aligned}$$

(iii) The inverse of $(xf'(x)/f(x), f(x))$ is

$$\left(\overline{f}(x), \overline{f}(x)\frac{f'(\overline{f}(x))}{f(\overline{f}(x))}\right)^{-1} = \left(x\frac{(\overline{f})'(x)}{\overline{f}(x)}, \overline{f}(x)\right) \in PW.$$

Other similar early examples involving the *Bell subgroup* $(g(x), xg(x))$ are by Jabotinsky ([111], [190, p. 238]). Let $f(x), g(x) \in L_0$ as usual. An operation called *Lagrange product* is defined in [196]: $f(x) \otimes g(x) = f(x)g(xf(x))$. This product is associative, distributive and has an identity: $f(x) \otimes 1 = f(x) = 1 \otimes f(\underline{x})$. Let $y = xf(x) \in L_1$. The inverse element of the series $f(x) \in L_0$ is denoted by $\overline{f}(y) = 1/f(\overline{l}(y))$, where $x = l(y) = yf(y)$. The group $(L_0, \otimes)$ is called the *Lagrange group*.

The following theorem gives a unified classification of known results about the Riordan groups from the point of view of groups of type E. It also provides new information about the structure of groups of type E and R^*.

THEOREM 12 *(On classification and structure [70]). Let $\varphi(0) = f(0) = \psi(0) = 1$. Then*

(a) *The matrix Riordan group of exponential type* [47] *coincides with the group $E(\{n!\}, \{k!\}; \varphi, f, 1)$.*

(b) *The Riordan group coincides with the group $E(\{1\}, \{1\}; \varphi, f, 1)$; the Bell group coincides with the group $E(\{1\}, \{1\}; f, f, 1)$; the PW group coincides with the group $E(\{1\}, \{1\}; (wf(w))'/f(w), f, 1)$.*

(c) *The PW group is isomorphic to the group $E(\{1\}, \{1\}; 1, 1, \psi)$ (see* Lemma 2*); the Bell group is isomorphic to the Lagrange group.*

(d) *The group $E(\{1\}, \{1\}; \varphi, f, \psi)$ decomposes into a product of three its subgroups, the PW, Voronoy and Bell subgroups.*

(e) *The Riordan group decomposes into a product of two its subgroups, the Voronoy and PW subgroups, or into a product of its two Voronoy and Bell subgroups.*

4.1.3. *Inverse identities of Legendre–Tshebyshev type* As in [176, Table 2.6, Relation 5] let $p > 0$, $r > 0$, and

$$C = (c_{mk}) = \left(\binom{rm+p}{m-k} - (r-1)\binom{rm+p}{m-k-1}\right), \quad m, k = 0, 1, 2, \ldots. \tag{34}$$

Then:

(a) the matrix (34) is of type $E^{(1)}(\{1\}, \{1\}; (1+z)^p(1-(r-1)z), 1, (1+z)^r)$;

(b) the inverse identities defined by matrix (34) are equivalent to the following functional relations:

$$A\big(z(1+z)^{-r}\big) = (1+z)^{p+1}B(z),$$
$$B(z) = (1+z)^{-p-1}A\big(z(1+z)^{-r}\big);$$

(c) the matrix $D = (d_{mk})$ (inverse of the matrix (34)) is given by

$$D = (d_{mk}) = \left((-1)^{m-k}\binom{m+p+rk-k}{m-k}\right);$$

(d) the matrices C and D can be represented as $C = ABI$, $D = IB^{-1}A^{-1}$, where I is the identity matrix,

$$A = (a_{mk}) = \left(\binom{mr}{m-k}\right),$$
$$B = (b_{mk}) = \left(\binom{p}{m-k} - (r-1)\binom{p}{m-k-1}\right),$$
$$A^{-1} = \big(a_{mk}^{(-1)}\big) = \left((-1)^{m-k}\binom{m+rk-k-1}{m-k}\right),$$
$$B^{-1} = \big(b_{mk}^{(-1)}\big) = \left((-1)^{m-k}\binom{p+m-k}{m-k}\right);$$

(e) the matrix relations

$$CD = I, \qquad C = ABI, \qquad D = IB^{-1}A^{-1},$$
$$ABB^{-1}A^{-1} = I, \qquad BB^{-1}A^{-1} = A^{-1}, \qquad ABB^{-1} = A$$

generate the following combinatorial identities:

$$\sum_{s=k}^{m}(-1)^{s-k}\left\{\binom{rm+p}{m-s} - (r-1)\binom{rm+p}{m-s-1}\right\}\binom{s+p+rk-k}{s-k}$$
$$= \delta(m,k), \quad m,k = 0,1,\ldots,$$

$$\sum_{s=k}^{m}\binom{mr}{m-s}\left\{\binom{p}{s-k} - (r-1)\binom{p}{s-k-1}\right\}$$
$$= \left(\binom{rm+p}{m-k} - (r-1)\binom{rm+p}{m-k-1}\right), \quad m,k = 0,1,\ldots,$$

$$\sum_{s=k}^{m}\binom{p+m-s}{m-s}\binom{s+rk-k-1}{s-k}$$
$$= \binom{m+p+rk-k}{m-k}, \quad m,k = 0,1,2,\ldots,$$

$$\sum_{n=k}^{m}\sum_{t=n}^{m}\sum_{s=t}^{m}(-1)^{t-k}\binom{mr}{m-s}\left(\binom{p}{s-t}\right.$$
$$\left. - (r-1)\binom{p}{s-t-1}\right)\binom{p+t-n}{t-n}\binom{n+rk-k-1}{n-k}$$
$$= \delta(m,k), \quad m,k = 0,1,\ldots,$$

$$\sum_{t=k}^{m}\sum_{s=t}^{m}\left(\binom{p}{m-s} - (r-1)\binom{p}{m-s-1}\right)$$
$$\times (-1)^{s-k}\binom{p+s-t}{s-t}\binom{t+rk-k-1}{t-k}$$
$$= (-1)^{m-k}\binom{m+rk-k-1}{m-k}, \quad m,k = 0,1,\ldots,$$

$$\sum_{t=k}^{m}\sum_{s=t}^{m}\binom{mr}{m-s}\left(\binom{p}{s-t} - (r-1)\binom{p}{s-t-1}\right)(-1)^{t-k}\binom{p+t-k}{t-k}$$
$$= \binom{mr}{m-k}, \quad m,k = 0,1,\ldots.$$

PROOF. From the integral representation of the binomial coefficients and from (34), it follows that

$$c_{mk} = \mathbf{res}_z(1+z)^{rm+p}z^{-m+k-1} - (r-1)\mathbf{res}_z(1+z)^{rm+p}z^{-m+k}$$
$$= \mathbf{res}_z\big((1+z)^{rm+p}\big(1-(r-1)z\big)z^{-m+k-1}\big).$$

Comparison of this expression for c_{mk} with (31) proves claim a) of this example, if we let

$$\alpha_m = \beta_k = 1, \qquad \varphi(z) = (1+z)^p\bigl(1-(r-1)z\bigr),$$
$$f(z) = 1, \qquad \psi(z) = (1+z)^r.$$

The other claims follow from properties of the operator **res** and of the relations of type E^1. □

4.1.4. *Conclusions* To our knowledge the results of Theorems 8, 9 and 12 are new in the theory of Riordan arrays and Riordan groups, and it is easy to find many various combinatorial interpretations and applications for them (see, [20,145,146,161] and etc.).

The results of Theorems 4–6, 10 and 11 are natural and not surprising for several reasons. First of all, the representation of combinatorial numbers a_n, $n = 0, 1, 2, \dots$, as well as their generating functions $A(w) = \sum_{n=0}^{\infty} a_n w^n$ by a infinite triangular (semicirculant) matrices is a routine procedure in combinatorial analysis (see, for example, [103,116] and a remark of S. Roman [180, p. 43]). Second, an integral representation of type E typically appears in the evaluation of many concrete combinatorial sums of different kinds (see [59], main theorem). This allows one to give a combinatorial interpretation to summation formulae, related to matrices of the type E. Weighting coefficients like α_m and β_k could be interpreted for example as the number of terms or the value of the sum under investigation. Finally, operations of multiplication, substitution and inversion with Laurent formal power series, hidden in the construction of matrices of E type, also have a combinatorial interpretation (see [96,103,177–179,37,114,19,89,44,97] and many others), including combinatorial interpretations and various proofs of one and multidimensional inversion Lagrange formulae ([86,90,107,127,154,208,205,91], etc.). It allows to explain in every particular case the algebraic structure of the enumeration object under investigation. The result of Theorem 8 plays a similar role (compared to the results of Theorems 11 and 12). The example from the previous section can be viewed as an extension of the approach in [176].

Note that the construction (31) and the results of Theorems 4–12 can be easily extended in several variants in the multidimensional case with the help of the main theorem in [59] including the case of fractional powers of the variables. Also the results of this section can be extended to a wide class of difference and q-difference relations using nice results of Ch. Krattenchaler, G. Andrews, I. Gessel and many others (see [121–124,17,86–88,95,28], etc.). Further generalization of the idea of pairs of inverse relations in the algebra of multivariate formal power series can be made based on an interesting work of V. Stepanenko [200]. These results and their various applications are the direction of our future investigations. In particular we are interested in their applications to asymptotic results and inverse combinatorial relations of different kinds (see [24–26,50,78,80,83,144,213,162,156] and [108,148,149,151,139,163], etc.).

4.2. *Solution of the Kargapolov problem on the enumeration of the ranks of factors in the lower central series of free k-step solvable groups with q generators*

The ranks of the factors for the lower central series of a free group Φ with q generators were computed in [100, Theorem 11.2.2]:

$$M_q(n) = \frac{1}{n}\sum_{d/n} \mu(d) q^{n/d}.$$

In [120, Question 2.18] M.I. Kargapolov posed the problem of computing the ranks of the factors for the lower central series of a free solvable group. Since the bases of a free solvable group and a free solvable Lie algebra coincide [192], the desired formula can be obtained by counting the basic commutators (regular words) in a basis for the free solvable Lie algebra. In [194] the formulae (4.21)–(4.24) were obtained for computing the ranks $R_q^{(3)}(n)$ of the factors of the free class three solvable group with q generators. These formulae were used in combination with the method of coefficients to find a considerably simpler formula for $R_q^{(3)}(n)$, and this enables us to solve the Kargapolov problem for a free solvable group in a combinatorial way [52].

The following formulae for the case $k = 3$ were obtained by V. Sokolov [194]:

$$R_q^{(3)}(n) = \begin{cases} \theta_q^{(1)}(n), & n = 1, 2, 3, \\ \theta_q^{(1)}(n) + \theta_q^{(2)}(n), & n \geqslant 4; \end{cases} \tag{35}$$

$$\theta_q^{(1)}(n) = \begin{cases} 0, & n = 0, \\ q, & n = 1, \\ (n-1)\binom{n+q+2}{q-2}, & n > 1; \end{cases} \tag{36}$$

$$\begin{aligned} \theta_q^{(2)}(n) = {} & \sum_{j=3}^{n-2} \theta_q^{(1)}(j)\big[\psi(n-j, 2) - \psi(n-j, j)\big] \\ & + \sum_{l=2}^{\lfloor n/2 \rfloor} \sum_{i=2}^{\lfloor n/2 \rfloor} (l-1)\binom{\theta_q^{(1)}(i) + l - 2}{l} \psi(n - li, i + 1), \end{aligned} \tag{37}$$

where

$$\psi(m, i) = \begin{cases} \sum_{\forall m \in \Psi(m,i)} \theta_q^{(1)}(\overline{m}), & m \geqslant i, \\ 1, & i > m = 0, \\ 0, & i > m > 0 \text{ or } m < 0, \end{cases} \tag{38}$$

$$\theta_q^{(1)}(\overline{m}) = \prod_{j=1}^{s} \binom{\theta_q^{(1)}(i_j) + k_j - 1}{k_j}.$$

The function $\theta_q^{(1)}(\overline{m})$ is defined on the set $\Psi(m, i)$ of i-partitions of the positive integer m. Any partition of m in $\Psi(m, i)$ has the form

$$m = \overbrace{i_1 + i_1 + \cdots + i_1}^{k_1 \text{ times}} + \overbrace{i_2 + i_2 + \cdots + i_2}^{k_2 \text{ times}} + \cdots + \overbrace{i_s + i_s + \cdots + i_s}^{k_s \text{ times}},$$

where all the terms are not less than i, and $i_j \neq i_l$ for $j \neq l$.

THEOREM 13. *(See [52].)*

$$R_q^{(3)}(n) = \begin{cases} \theta_q^{(1)}(n), \quad n = 1, 2, 3, \\ \theta_q^{(1)}(n) + \mathbf{res}_z\{(qz-1)(\prod_{j=1}^{n}(1-z^j)^{-\theta_q^{(1)}(j)})z^{-n-1}\}, \\ \quad n \geqslant 4; \end{cases} \tag{39}$$

$$R_2(w) = \sum_{n=0}^{\infty} {}_q^{(3)}\theta_q^{(1)}(n)w^n = (1+wq) + (1-w)^{-q}(wq-1),$$

$$\begin{aligned} R_3(w) &= \sum_{n=0}^{\infty} R_q^{(3)}(n)w^n \\ &= R_2(w) - \sum_{n=0}^{3} w^n \mathbf{res}_z\left\{(qz-1)\exp\left(-\sum_{j=1}^{3} R_2(z^j)/j\right)z^{-n-1}\right\} \\ &\quad + (qw-1)\exp\left(-\sum_{j=1}^{\infty} R_2(w^j)/j\right), \quad 0 \leqslant w < 1. \end{aligned} \tag{40}$$

The proof of this theorem follows from results in Lemmas 3–5 below, where integral representations are found successively for some of the above defined quantities $R_q^{(3)}(n)$, $\theta_q^{(1)}(n)$, $\theta_q^{(2)}(n)$ and $\psi(m,i)$. It is interesting to observe how the "superfluous" summation signs "disappear" either after summation operations or in the process of collecting similar terms.

LEMMA 3.

$$\theta_q^{(1)}(n) = \mathbf{res}_w\big(1 + wq + (1-w)^{-q}(wq-1)\big)w^{-n-1}, \quad n = 0, 1, 2, \ldots. \tag{41}$$

LEMMA 4.

$$\psi(m,j) = \mathbf{res}_z\left\{\left(\prod_{i=j}^{n}(1-z^j)^{-\theta_q^{(1)}(i)}\right)z^{-m-1}\right\}. \tag{42}$$

A proof this formulae is immediately obtained by comparing the expressions for the values of $\theta_q^{(1)}(n)$ and $\psi(m,j)$ as given in their definitions and Lemmas 3, 4.

LEMMA 5.

$$S_1 = \sum_{j=0}^{n} \theta_q^{(1)}(j)\psi(n-j,2) = \psi(n,2) + \psi(n,2) + \psi(n,2) + \psi(n,2). \tag{43}$$

PROOF. According to the formulae of Lemmas 3, 4 and the substitution rule

$$S_1 = \sum_{j=0}^{n} \mathbf{res}_w\big(1 + wq + (1-w)^{-q}(wq-1)\big)w^{-j-1}$$

$$
\begin{aligned}
&\times \mathbf{res}_z \left(\prod_{k=2}^{\infty} \left(1 - z^k\right)^{-\theta_q^{(1)}(k)} \right) z^{-n+j-1} \\
&= \mathbf{res}_z \left\{ \left(\prod_{k=2}^{\infty} \left(1 - z^k\right)^{-\theta_q^{(1)}(k)} \right) \right. \\
&\qquad \left. \times z^{-n-1} \left(\sum_{j=0}^{n} z^{-j} \mathbf{res}_w \left(1 + wq + (1-w)^{-q}(wq-1)\right) w^{-j-1} \right) \right\} \\
&= \mathbf{res}_z \left\{ \left(1 + zq + (1-z)^{-q}(zq-1)\right) \left(\prod_{k=2}^{\infty} \left(1 - z^k\right)^{-\theta_q^{(1)}(k)} \right) z^{-n-1} \right\}.
\end{aligned}
$$

By (42) and relation $q = \theta_q^{(1)}(1)$ this gives us (43). □

LEMMA 6. *If $i \geqslant 2$, then*

$$
\begin{aligned}
S_2 &= \sum_{i=0}^{\lfloor n/2 \rfloor} (l-1) \binom{\theta_q^{(1)}(i) + l - 2}{l} \psi(n - li, i+1) \\
&= \theta_q^{(1)}(i) \psi(n-i, i) - \psi(n, i).
\end{aligned}
$$

PROOF. Since $(l-1)\binom{\alpha+l-2}{l} = \mathbf{res}_w (1-w)^{-a}(w\alpha - 1) w^{-l-1}$, the result of Lemma 4 and the substitution rule imply that

$$
\begin{aligned}
S_2(i) &= \sum_{l} \mathbf{res}_z \left(\prod_{j=i+1}^{\infty} (1 - z^j)^{-\theta_q^{(1)}(j)} \right) z^{-n+li-1} \\
&\qquad \times \mathbf{res}_w (1-w)^{-\theta_q^{(1)}(i)} \left(w\theta_q^{(1)}(i) - 1\right) w^{-l-1} \\
&= \mathbf{res}_z \left\{ \left(\prod_{j=i+1}^{\infty} \left(1 - z^j\right)^{-\theta_q^{(1)}(j)} \right) \right. \\
&\qquad \left. \times z^{-n-1} \left(\sum_{l=0}^{\infty} z^{li} \mathbf{res}_w (1-w)^{-\theta_q^{(1)}(i)} \left(w\theta_q^{(1)}(i) - 1\right) w^{-l-1} \right) \right\} \\
&= \mathbf{res}_z \left\{ \left(\prod_{j=i+1}^{\infty} \left(1 - z^j\right)^{-\theta_q^{(1)}(j)} \right) \left(1 - z^i\right)^{-\theta_q^{(1)}(i)} \left(z^i \theta_q^{(1)}(i) - 1\right) z^{-n-1} \right\} \\
&= \theta_q^{(1)}(i) \psi(n-i, i) - \psi(n, i).
\end{aligned}
$$

□

LEMMA 7.

$$
\theta_q^{(2)}(n) = q\psi(n-1, 1) - \psi(n, 1).
$$

Proof. From (36)–(38) we get, after elementary transformations,

$$\theta_q^{(2)}(n) = S_1 - \sum_{j=0}^{\lfloor n/2 \rfloor} \theta_q^{(1)}(j)\psi(n-j,j) + \sum_{i=2}^{\lfloor n/2 \rfloor} S_2(i) \\ - q\bigl(\psi(n-1,2) - \psi(n-1,1)\bigr) + \sum_{i=2}^{\lfloor n/2 \rfloor} \psi(n,i+1). \tag{44}$$

The required result is a consequence of the substitution of the expressions S_1 and $S_2(i)$ into (44) and of the condition $\psi(n, \lfloor n/2 \rfloor + 1) = 0$. □

From last lemma and from (35) and (42) we get formula (39) in Theorem 13. Formula (40) follows from the relations

$$\prod_{j=1}^{\infty}\bigl(1-w^j\bigr)^{-\theta_q^{(1)}(j)} = \exp\left(-\sum_{j=1}^{\infty}\theta_q^{(1)}(j)\ln\bigl(1-w^j\bigr)\right) \\ = \exp\left(-\sum_{n=1}^{\infty}\sum_{j=1}^{\infty} w^{nj}/n\right) = \sum_{n=1}^{\infty} R_2\bigl(w^n\bigr)/n, \quad 0 \leqslant w < 1.$$

Theorem 14. *(See [52].) The rank of the factor $\gamma_n(F)/\gamma_{n+1}(F)$ $(n = 1, 2, \ldots)$ of a free class $(k+1)$ solvable group F with q generators $(q \geqslant 2)$ can be computed according to the following recursion formula*:

$$R_q^{(0)}(n) = \begin{cases} 0, & n \neq 1, \\ q, & n = 1; \end{cases}$$

$$R_q^{(1)}(n) = \begin{cases} 0, & n = 0, \\ q, & n = 1, \\ (n-1)\binom{n+q-2}{q-2}, & n > 1 \end{cases}$$

and for $k \geqslant 1$,

$$R_q^{(k+1)}(n) = \begin{cases} R_q^{(k)}(n), & n < n^{(k)}, \\ R_q^{(k)}(n) + \mathbf{res}_z\{(qz-1)(\prod_{j=1}^{\infty}(1-z^j)^{-R_q^{(k)}(j)})z^{-n-1}\}, \\ \quad n \geqslant n^{(k)}, \end{cases}$$

where $n^{(k)}$ is the minimal possible weight of an arbitrary R_k-word:

$$n^{(k)} = \begin{cases} (2^{k+2}-1)/3, & q = 2 \text{ and } k = 0, 2, 4, \\ (2^{k+2}-2)/3, & q = 2 \text{ and } k = 1, 3, \\ 21 \times 2^{k-4}, & q = 2 \text{ and } k \geqslant 5, \\ 2^k, & q \geqslant 3. \end{cases}$$

Remark 5. The simplicity of formulae (39), (40) for $R_q^{(3)}(n)$ allows to understand the structure of the set of objects being enumerated (regular words). This, in turn, enables us to solve the Kargapolov problem for a free solvable group [52], and then also for a free polynilpotent group [92] and for free groups in varieties [59, pp. 215–221]. Independently, an asymptotic solution of the same problem was given later in [167].

4.3. *Enumeration of ideals for some matrix rings*

In [62,63,65,71,75] with the help of the method of coefficients several enumerative problems for groups and algebras of Lie type including enumeration of ideals of some matrix rings were solved. In general, these investigations may be considered as part of the program that is formulated in [206].

In [126] a constructive description of various classes of ideals of some matrix rings has been given. This description was done by considering certain sets of matrix positions which allow enumeration with the help of a combinatorial scheme of paths with diagonal steps on a rectangular lattice [71]. In particular, we proved, that the number of some ideals of the ring $R_n(K, J)$ is equal to

$$\sum_{i,j=1}^{n} \binom{n-i+j-1}{n-i}\binom{i-1+n-j}{n-j}.$$

PROPOSITION 1. *Let $n \in \mathbb{N}$. Then the following identity is valid*:

$$\sum_{i,j=1}^{n} \binom{n-i+j-1}{n-i}\binom{i-1+n-j}{n-j} = (2n-1)\binom{2n-2}{n-1}.$$

PROOF. According to the properties on **res** operator we have

$$\begin{aligned}
&\sum_{i,j=1}^{n} \binom{n-i+j-1}{n-i}\binom{i-1+n-j}{n-j}\\
&\quad= \sum_{i,j=1}^{n} \mathbf{res}_x\big((1-x)^{-j}x^{-n+i-1}\big)\mathbf{res}_y\big((1-y)^{-i}x^{-n+j-1}\big)\\
&\quad= \sum_{i,j=1}^{\infty} \cdots\\
&\quad= \mathbf{res}_{xy}\left\{(1-x)^{-1}(1-y)^{-1}(xy)^{-n-1}\left[\sum_{i,j=1}^{\infty}\big(x/(1-y)\big)^{i}\big(y/(1-x)\big)^{j}\right]\right\}\\
&\quad= \mathbf{res}_{xy}\big\{(1-x)^{-1}(1-y)^{-1}(xy)^{-n-1}\\
&\qquad\times\big[\big(1-\big(x/(1-y)\big)\big)^{-1}\big(1-\big(y/(1-x)\big)\big)^{-1}\big]\big\}\\
&\quad= \mathbf{res}_{xy}\big\{(1-x-y)^{-2}(xy)^{-n-1}\big\}\\
&\quad= \mathbf{res}_{xy}\left\{\left[1+\sum_{k=1}^{\infty}\binom{k+1}{k}(x+y)^{k}\right](xy)^{-n-1}\right\}
\end{aligned}$$

$$= \sum_{k=1}^{\infty}(k+1)\mathbf{res}_{xy}\bigl\{(x+y)^k(xy)^{-n-1}\bigr\}$$
$$= (k+1)\mathbf{res}_{xy}\bigl\{(x+y)^k(xy)^{-n-1}\bigr\}\big|_{k=2(n-1)} = (2n-1)\binom{2n-2}{n-1}.$$

The last sum has all terms with $k \neq n$ equal to zero. □

The following example of a triple summation involving summation by partitions and the integer part operator appeared recently in an enumerative algebra application: namely in enumerative formulae for projectively congruent quadrics and symmetric forms of moduli over local rings [132,198]. Let $N(n,s)$ denote the number of classes of projective space RP_{n-1} ($n > 2$) over a local ring R with principal maximal ideal of nilpotency step s, where $2 \in R^*$ and $|R^* : R^{*2}| = 2$. Then

$$\begin{aligned} N(n,s) &= \sum_{m=1}^{n}\sum_{q=1}^{\min(m,s)}\binom{s}{q}2^{q-1}\biggl\{\binom{-1+m/2}{q-1}' + \binom{m-1}{q-1}\biggr\}, \\ &\quad \text{if } R^* \cap (1+R^2) \nsubseteq R^{*2}; \end{aligned} \tag{45}$$

$$\begin{aligned} N(n,s) &= \sum_{m=1}^{n}\sum_{q=1}^{\min(m,s)}\binom{s}{q}\biggl\{\binom{-1+m/2}{q-1}' \\ &\quad + \sum_{(n_1,\dots,n_q)\in\Omega_q(m)}\biggl\lfloor 1/2\prod_{j=1}^{q}(n_j+1)\biggr\rfloor\biggr\}, \quad \text{if } 1+R^{*2} \subset R^{*2}. \end{aligned} \tag{46}$$

Here $\binom{p}{q}'$ is equal to $\binom{p}{q}$ for nonnegative integers p and q, and 0, otherwise; $\Omega_q(m)$ denotes the set of all ordered partitions of the number m in q parts: $n_1 + \cdots + n_q = m$. These formulae can be simplified using the integral representation technique.

PROPOSITION 2. *Let*

$$S(n,s) = -1/2 + \sum_{q=0}^{s}2^{q-1}\binom{s}{q}\binom{n}{q}$$
$$= -1/2 + 1/2\,\mathbf{res}_z(1+z)^s(1-z)^{-s-1}z^{-n-1}.$$

If $R^* \cap (1+R^2) \nsubseteq R^{*2}$ *then*

$$N(n,s) = S(n,s) + S\bigl(\lfloor n/2\rfloor, s\bigr), \tag{47}$$

and for fixed s *and* $n \to \infty$

$$N(n,s) \asymp -1 + 2^{-1}\bigl(2^s+1\bigr)n^s/s!.$$

PROPOSITION 3. *Let*

$$T(n,p) = -1/2 + 1/2\binom{n+p}{p}.$$

If $1 + R^{*2} \subset R^{*2}$ *then*

$$N(n,s) = T(n,2s) + T\big(\lfloor n/2 \rfloor, s\big). \tag{48}$$

REMARK 6. A simplification of formulae usually brings new information on the structure of the objects of enumeration. For example, the simplification of the known formulae for $R_q^{(3)}(n)$ from [194] led to a better understanding of the structure of the enumerable regular words (commutators) known Shirshov bases of a free Lie algebra. This led to a solution of the Kargapolov problem of computing the ranks $R_q^{(k)}(n)$ of the factors for the lower central series of a free solvable group of step k with q generators for arbitrary k [52]. This, in turn, made is possible to solve an analogous problem for a free polynilpotent group [92] and for free groups in varieties [59, p. 215–221]. Another answer to the problem of Kargapolov was suggested in [167].

The simplicity found in formulae (47), (48) poses the following problem: give an independent algebraic proof and interpretation of these formulae for the number of quadrics the in projective space RP_{n-1}, $n > 2$, over a local ring R with its maximal ideal nilpotent of class s.

5. Multidimensional case and applications in the theory of holomorphic functions of several complex variables

5.1. *General construction and its particular cases (MacMahon's master theorem and Good's theorem)*

In this section we consider integrals of meromorphic forms over cycles (over chains, in the general case). Such integrals are usually difficult to compute directly except for the case of separating cycles [214, p. 145]. A separating cycle is a composite co-boundary for an isolated point of intersection of the singular surfaces determined by the different factors in the denominator of the integrand form.

The following general construction frequently arises in computing one-dimensional and multidimensional combinatorial sums by means of integral representations, and it can be regarded as a generalization of the MacMahon master theorem and a known theorem of Good [135,90,91].

THEOREM 15. *(See* [59, *pp.* 175–176], [55,56].*) Suppose that* $\psi(z)$, $f(z)$, $\varphi_j(z)$ *and* $f_j(z)$, $j = 1, \dots, n$, *are holomorphic functions in a domain* $D \subset \mathbb{C}^n$, *where*

$$f(z) = \prod_{j=1}^{n} f_j(z), \qquad \varphi(z) = \prod_{j=1}^{n} \varphi_j(z),$$

and the system of functions $\varphi_1(z), \dots, \varphi_n(z)$ *has a finite set* Ω *of simple* (*of multiplicity* 1) *zeros in* D, *the Jacobian* $\partial\varphi/\partial z$ *is not identically zero*, G *is one or several of the connected components of the set*

$$\big\{z \in D\colon \big|\varphi_j(z)\big| < \varepsilon_j,\ j = 1, \dots, n\big\}, \quad G \subset D,$$

and $\gamma = \Gamma_n = \{z \in \overline{G}\colon |\varphi_j(z)| = \varepsilon_j,\ j = 1, \dots, n\}$ *is the skeleton of* $\overline{G}$. *Then*

$$\begin{aligned} S(k) &= \frac{1}{(2\pi i)^n} \int_{\Gamma_n} \psi(z) f^k(z) \varphi^{-k-1}(z)\, dz \\ &= \frac{1}{(2\pi i)^{2n}} \int_{\Gamma_n \times \gamma} \psi(z) \left(\prod_{j=1}^{n} \bigl(\varphi_j^{m_j}(z) - t_j f^{m_j}(z)\bigr)^{-1} t_j^{-k-1} \right) dz \wedge dt \\ &= \frac{1}{(2\pi i)^{2n}} \int_{\Gamma_n \times \gamma_2} \psi(z) \qquad (49) \\ &\quad \times \left(\prod_{j=1}^{n} \varphi_j^{k-m_j}(z) \bigl(\varphi_j^k(z) - t_j f^k(z)\bigr)^{-1} t_j^{-m_j-1} \right) dz \wedge dt, \end{aligned}$$

where the skeleton $\gamma = \{t = (t_1, \dots, t_n)\colon |t_j| = \varepsilon_j,\ j = 1, \dots, n\}$ *and* ε_j *are chosen small enough so that for* $t \in \gamma$ *and* $z \in \Gamma$ *the following conditions are satisfied*: $|\varphi_j(z)| < |t_j f(z)|$, $j = 1, \dots, n$. *In particular*:

(a) *If* $m_1 = \cdots = m_n = 1$, *then* [90,56]

$$S(k) = \frac{1}{(2\pi i)^n} \int_{\gamma} \left(\sum_{z \in A} \left[\psi(z) / \frac{\partial w}{\partial z} \right]_{z=z(t)} \right) \left(\prod_{j=1}^{n} t_j^{-k-1} \right) dz \wedge dt,$$

where A *is the set of zeros* $z(t)$ *of the system*

$$w_j = \varphi_j(z) - t_j f(z), \quad j = 1, \dots, n, \tag{50}$$

in G *when* $t \in \gamma$, *and* $\partial w / \partial z$ *is the Jacobian of the transformation* (50).

(b) *If* $f_j(z_0) \neq 0$, $j = 1, \dots, n$, *then* [56]

$$S(k) = \frac{1}{(2\pi i)^n} \int_{\gamma} \left(\sum_{z \in B} \left[\psi(z) / \frac{\partial w}{\partial z} \right]_{z=z(t)} \right) \left(\prod_{j=1}^{n} t_j^{-k-1} \right) dz \wedge dt,$$

where B *is the set of zeros* $z(t)$ *of the system*

$$w_j = \varphi_j^{m_j}(z) - t_j f_j^{m_j}(z), \quad j = 1, \dots, n, \tag{51}$$

in G *when* $t \in \gamma$, *and* $\partial w / \partial z$ *is the Jacobian of the transformation* (51).

(c) *If* $g(z) = \psi(z)(\prod_{j=1}^{n} \varphi_j^{k-m_j}(z))$ *is a holomorphic function in* D *and* $f_j(z_0) \neq 0$, $j = 1, \dots, n$, $z_0 \in \Omega$, *then*

$$S(k) = \frac{1}{(2\pi i)^n} \int_{\gamma} \left(\sum_{z \in C} \left[\psi(z) / \frac{\partial w}{\partial z} \right]_{z=z(t)} \right) \left(\prod_{j=1}^{n} t_j^{-m_j-1} \right) dz \wedge dt,$$

where C *is the set of zeros* $z(t)$ *of the system*

$$w_j = \varphi_j^k(z) - t_j f_j^k(z), \quad j = 1, \dots, n, \tag{52}$$

in G *when* $t \in \gamma$, *and* $\partial w / \partial z$ *is the Jacobian of the transformation* (52).

COROLLARY 2. *MacMacon's master theorem* [135] *follows from part* (c) *of the previous theorem with* $k = 1$, $\psi(z) = \varphi(z)$, $\varphi_j(z) = z_j$, *and* $f_j(z) = \sum_{i=1}^{n} a_{ij} z_j$, $j = 1, \dots, n$.

Although the denominator of the function $g(z)/(\partial w/\partial z)$ can vanish in G, the indicated residue theorem is applicable because for $t \in \gamma$ the integrand

$$g(z)\left(\prod_{j=1}^{n}\left(\varphi_j^k(z) - t_j f_j^k(z)\right)^{-1}\right) = \frac{g(z)}{w(z,t)}\,dz = \frac{g(z)}{\partial w/\partial z}\frac{dw(z,t)}{w(z,t)}$$

is holomorphic in $D \setminus \{z\colon\ w(z,t) = 0\}$, and $w(z,t) \neq 0$ at the points $z(t) \in A$.

The formulae in (a)–(c) of the theorem express the integral (49) in terms of multidimensional residues. It is not possible to find residue formulae in the general case. However, in concrete cases the computation of the integral (49) does not cause any fundamental difficulties. In particular, it can be computed by iterated integration or with the help of the splitting lemma (see, for example, [59, §5.4]). The procedure of this lemma enables us to find the necessary change of variables and in the case for separating cycles to reduce the integral to the computation of local **res**'s (Grothendieck residues) at these points, where the special surfaces intersect. If the case will be successful then using this construction generates a combinatorial formula of summation having greater multiplicity than the original sum. It is not difficult to see that this is of analogous type as the Wilf transformation "snake oil method" [209]. There a scalar (diagonal) product of two number sequences is replaced by a Cauchy product for sequences of larger dimension.

5.2. *Computation of Szegő and Bergman kernels for functions holomorphic in certain bounded and unbounded n-circular domains in $\mathbb{C}^n$*

The computation of the Szegő and Bergman kernels for various domains in $\mathbb{C}^n$ in closed form is a hard and interesting problem in function theory ([189,15] and many others). Surprisingly, it has been solved for a long time by many authors including specialists in function theory for the separate cases with the help of different ad hoc methods (see, for example, [216,27,36]).

Here the method of integral representation of sums is applied to the solution of this important problem for a wide class of bounded and unbounded n-circular domains. In Section 2.2.2 we obtained a closed form formula for the Bergman kernel of a certain domain in $\mathbb{C}_z^3$. In [57] we found in closed form the Szegő and Bergman kernels of certain bounded n-circular domains in $\mathbb{C}^n$ which depend on a family of real parameters. Our formulae enabled us (by a passage to the limit) to formulate a conjecture for those Szegő kernels with concrete unbounded domains, that has been justified in [14]. There we obtained integral representations with Szegő kernels for functions that are holomorphic in certain unbounded complete domains in $\mathbb{C}^n$, and found in closed form the Szegő kernels for a class of the concrete unbounded n-circular domains in $\mathbb{C}^n$ (see [14, Theorem 2]). The conditions of applicability for those integral representations do not have analogues for holomorphic functions of single complex variable. Here is the notation needed to formulate these results.

Let $p = (p_1, \dots, p_n)$, $\gamma = (\gamma_1, \dots, \gamma_n)$, and $\alpha = (\alpha_1, \alpha_2)$, with p_j, γ_i and α_l positive integers, and consider the n-circular domain

$$D^{p,\gamma,\alpha} = \bigl\{z = (z_1, \dots, z_n)\bigr\} \in \mathbb{C}^n :$$
$$\left(\sum_{i=1}^{s}\left(\sum_{j=1+\nu_{i-1}}^{\nu_i} |z_j|^{2/p_j}\right)^{1/\gamma_i}\right)^{1/\alpha_1}$$
$$< \exp\left(-\left(\sum_{i=1+s}^{m}\left(\sum_{j=1+\nu_{i-1}}^{\nu_i} |z_j|^{2/p_j}\right)^{1/\gamma_i}\right)^{1/\alpha_2}\right),$$

where s and ν_i are certain fixed numbers determining the number of terms in the parentheses, with $0 < s < m$ and $0 = \gamma_0 < \gamma_1 < \gamma_2 < \cdots < \gamma_m = n$. Furthermore, let $\zeta = (\zeta_1, \dots, \zeta_n) \in \mathbb{C}^n_\zeta$, $x_j = z_j\overline{\zeta_j}$, $j = 1, \dots, n$, and consider the differential form

$$\frac{1}{(2\pi i)^n} d|\zeta|^2[n] \wedge \frac{d\zeta}{\zeta} = \frac{1}{(2\pi i)^n} d|\zeta_1|^2 \wedge \ldots \wedge d|\zeta_{n-1}|^2 \wedge \frac{d\zeta_1}{\zeta_1} \wedge \ldots \wedge \frac{d\zeta_n}{\zeta_n}$$

and the Szegő kernel $h(x) = h(x_1, \dots, x_n)$ for the domain $(\ldots)$ when the integration measure corresponds to this differential form. Moreover, let $\varGamma^{(1)}_{(\rho)} = \{z = (z_1, \dots, z_n) \in \mathbb{C}^n_z\colon |z_j| = \rho, j = 1, \dots, n\}$, $\varGamma^{(2)}_{(\rho)} = \{v = (v_1, \dots, v_m) \in \mathbb{C}^m_v\colon |v_j| = \rho, j = 1, \dots, m\}$, and $\varGamma^{(3)}_{(\rho)} = \{w \in \mathbb{C}_w\colon |w| = \rho\}$, with ρ be sufficiently small. We set $\varGamma_{(\rho)} = \varGamma^{(1)}_{(\rho)} \times \varGamma^{(2)}_{(\rho)} \times \varGamma^{(3)}_{(\rho)}$ and define the numbers $\delta(\beta, p) = \exp(2\pi i\beta/p)$, $\beta = 1, \dots, p$.

THEOREM 16. *(See [57].) The following relations hold*:

$$h(x) = \sum_{|k|\geqslant 0} \frac{(\alpha_1 Q_1)^{\alpha_2 Q_2}}{(\alpha_2 Q_2)!}(Q_1)!(Q_2)!$$
$$\times\left(\prod_{i=1}^{m}\frac{(T_i)!}{(\gamma_i T_i)!}\right)\frac{x_n^{k_n}}{(p_n k_n)!}\left(\prod_{j=1}^{n-1}\frac{x_j^{k_j}}{(p_j(k_j+1)-1)!}\right),$$

where

$$T_i = \sum_{j=1+\nu_{i-1}}^{\nu_i} p_j(k_j+1), \quad i = 1, \dots, m-1,$$

$$T_m = p_n k_n + \sum_{j=1+\nu_{m-1}}^{n-1} p_j(k_j+1),$$

$$Q_1 = \sum_{i=1}^{s}\gamma_i T_i, \qquad Q_2 = \sum_{i=1+s}^{m}\gamma_i T_i;$$

$$
\begin{aligned}
h(x) = \frac{\partial^{n-1}}{\partial x_1 \dots \partial x_{n-1}} \Bigg\{ & \frac{1}{(2\pi i)^{n+m+1}} \int_{\Gamma_{(\rho)}} w^{-1} \left(\prod_{i=1}^{m} v_i \right)^{-1} \left(\prod_{j=1}^{n} z_j \right)^{-1} \\
& \times \left(\prod_{j=1}^{s} \prod_{j=1+\nu_{i-1}}^{\nu_i} z_j^{p_j} v_i^{\gamma_i p_j} \big(z_j^{p_j} v_i^{\gamma_i p_j} - x_j V_1^{-\gamma_i p_j} Z_i^{p_j} e^{w\alpha_1 \gamma_i p_j} \big)^{-1} \right) \\
& \times \Bigg(\prod_{j=1+s}^{m} \prod_{j=1+\nu_{i-1}}^{\nu_i} z_j^{p_j} v_i^{\gamma_i p_j} w^{\alpha_2 \gamma_i p_j} \\
& \times \big(z_j^{p_j} v_i^{\gamma_i p_j} w^{\alpha_2 \gamma_i p_j} - x_j V_2^{-\gamma_i p_j} Z_i^{p_j} e^{w\alpha_1 \gamma_i p_j} \big)^{-1} \Bigg) dz \wedge dv \wedge dw \Bigg\},
\end{aligned}
$$

where $|x|$ *is sufficiently small, and*

$$
Z_i = 1 + \sum_{j=1+\nu_{i-1}}^{\nu_i} z_j, \quad i = 1, \dots, m,
$$

$$
V_1 = 1 + \sum_{i=1}^{s} v_i, \qquad V_2 = \sum_{i=1+s}^{m} v_i;
$$

$$
\begin{aligned}
h(x) = {} & \alpha_2^{-1} \left(\prod_{j=1}^{n} p_j^{-1} \right) \left(\prod_{i=1}^{m} \gamma_i^{-1} \right) \\
& \times \Bigg\{ \sum_{t_1=1}^{p_1} \cdots \sum_{t_n=1}^{p_n} \sum_{r_1=1}^{\gamma_1} \cdots \sum_{r_m=1}^{\gamma_m} \sum_{l=1}^{\alpha_2} \Bigg\{ 1 - \left(\sum_{i=1}^{s} \delta(r_i, \gamma_i) F_i^{1/\gamma_i}(x) \right) \\
& \times \exp\left(\alpha_1 \delta(l, \alpha_2) \left(\sum_{i=1+s}^{m} \delta(r_i, \gamma_i) F_i^{1/\gamma_i}(x) \right)^{1/\alpha_2} \right) \Bigg\}^{-1} \Bigg\}
\end{aligned}
$$

where

$$
F_i(x) = \sum_{j=1+\nu_{i-1}}^{\nu_i} \delta(t_j, p_j) x_j^{1/p_j}, \quad j = 1, \dots, m.
$$

The following special cases of the kernel $h(x)$ are of particular interest.

(a) Let $\alpha_1 = \alpha_2 = \gamma_1 = \cdots = \gamma_m = p_1 = \cdots = p_m = 1$. Then

$$
D = \left\{ z = (z_1, \dots, z_n) \in \mathbb{C}^n;\ \sum_{j=1}^{\nu} |z_j|^2 < \exp\left(- \sum_{j=1+\nu}^{n} |z_j|^2 \right) \right\},
$$

$$
h(x) = \frac{\partial^{n-1}}{\partial x_1 \dots \partial x_{n-1}} \left\{ 1 - \left(\sum_{j=1}^{\nu} x_j \right) < \exp\left(- \sum_{j=1+\nu}^{n} x_j \right) \right\}^{-1}.
$$

(b) Let $\alpha_2 = \gamma_{1+s} = \cdots = \gamma_m = p_{1+\nu} = \cdots = p_m = 1$. Then

$$D = \Bigg\{ z = (z_1, \ldots, z_n) \in \mathbb{C}^n;$$
$$\left(\sum_{i=1}^{s} \left(\sum_{j=1+\nu_{i-1}}^{\nu_i} |z_j|^{2/p_j} \right)^{1/\gamma_i} \right)^{1/\alpha_1} < \exp\left(- \sum_{j=1+\nu}^{n} |z_j|^2 \right) \Bigg\},$$

$$h(x) = \left(\prod_{j=1}^{s} \gamma_i^{-1} \left(\prod_{j=1+\nu_{i-1}}^{\nu_i} p_j^{-1} \right) \right)$$
$$\times \frac{\partial^{n-1}}{\partial x_1 \ldots \partial x_{n-1}} \left\{ \sum_{t_1=1}^{p_1} \cdots \sum_{t_s=1}^{p_s} \sum_{r_1=1}^{\gamma_1} \cdots \sum_{r_m=1}^{\gamma_m} \left\{ 1 - \left(\sum_{i=1}^{s} \delta(r_i, \gamma_i) F_i^{1/\gamma_i}(x) \right) \right. \right.$$
$$\left. \left. \times \exp\left(\alpha_1 \sum_{j=1+\nu}^{n} x_j \right) \right\}^{-1} \right\}.$$

For fixed $x_{1+\nu}, \ldots, x_n$ the fact that $h(x)$ is a single-valued algebraic function of each variable $x_1, \ldots, x_\nu$ ([109, p. 293], and [189, Problem 16b, p. 302]) implies that $h(x)$ is rational in these variables. Moreover, it is rational with respect to $\exp\big(\alpha_1 \sum_{j=1+\nu}^{n} x_j\big)$. In particular, if $D = \{z = (z_1, z_2);\ |z_1|^{2/p} < \exp(-|z_2|^2)\}$, with p a natural number, then

$$h(x) = \frac{\partial}{\partial x_2} \big\{ 1/\big(1 - x_1 \exp(p x_2)\big) \big\}^{-1}.$$

(c) Suppose that at least one p_j, $j = 1 + \nu, \ldots, n$, is greater than 1.
For example, if $D = \{z = (z_1, z_2);\ |z_1|^2 < \exp(-|z_2|)\}$, then

$$h(x) = \frac{\partial}{\partial x_2} \left\{ \frac{1 - x_1 \cosh x_2^{1/2}}{1 - 2x_1 \cosh x_2^{1/2} + x_1^2} \right\}.$$

The above kernel only appears not to be holomorphic; however, since the function $\cosh x_2$ is even, and consequently $\cosh(x_2^{1/2})$ is holomorphic.

REMARK 7. An even more general statement of the last theorem is formulated in [59, Chapter 6]. However due to inconvenience of the calculations we completely omit the general case here and the important stage of computations, connected with finding the coefficients of monomials of multiple degree series that are Szegő and Bergman kernels of the type studied. These coefficients are expressed as multiple definite integrals from a certain family with integer parameters, that are connected by complex recurrence relations, and can be computed effectively using integral representation techniques [59]. Previously, the analogous integrals were computed by other authors in closed form for some special cases by means of different ad hoc methods. Solving this problem [57,14] we successfully used the method of coefficients and integral representation for the β-function.

5.3. *Integral representation and polynomial identities for the computation of permanents*

The concept of integral representation applies to obtaining new computation formulae for special combinatorial sums such as permanents. The permanent of an $n \times n$ matrix $A = (a_{ij})$ over the field of complex numbers is defined by the expression

$$per(A) = \sum_{\sigma \in S_n} a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)}, \tag{53}$$

where S_n is a symmetric group of order n.

The permanent has the following known integral representation:

$$per(A) = \mathbf{res}_{z_1,\ldots,z_n} \left(\prod_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} z_j \right) z_i^2 \right). \tag{54}$$

According to the main MacMahon theorem we have

$$per(A) = \mathbf{res}_{t_1 \ldots t_n} \{1/\Delta\} t_1^{-2} \ldots t_n^{-2},$$

where

$$\Delta = \det(\delta_{ij} - a_{ij} t_j).$$

This formula is sometimes used in combinatorial computations and in the computation of the permanent [164]. Moreover we can obtain from the integral formula (54) another useful formula for computing the permanent, if we note that

$$per(A) = \mathbf{res}_{z_1,\ldots,z_n} \left(\prod_{i=1}^{n} f_i(z) z_i^2 \right), \tag{55}$$

where

$$f_i(z) = \sum_{j=1}^{n} a_{ij} z_j + \cdots \tag{56}$$

are series whose coefficients of monomials with degrees >1 are arbitrary.

THEOREM 17 *(Polynomial identities for the permanent [60,61]). If $\lambda_1, \ldots, \lambda_n$ and $\gamma_1, \ldots, \gamma_n$ are independent variables then*

$$per(A) = \sum_{k=0}^{n} (-1)^k \sum_{1 \leqslant j_1 < \cdots < j_k \leqslant n} \left(\prod_{i=1}^{n} (\lambda_j a_{ij} - a_{ij_1} - \cdots - a_{ij_k}) \right) \tag{57}$$

and

$$per(A) = \sum_{k=0}^{n} (-1)^k \sum_{1 \leqslant j_1 < \cdots < j_k \leqslant n} \left(\prod_{i=1}^{n} (\gamma_i - a_{ij_1} - \cdots - a_{ij_k}) \right). \tag{58}$$

PROOF. If we substitute in (56)

$$f_i(z) = \exp\left(\lambda_i\left(\sum_{j=1}^{n} a_{ij}z_j\right)\right) - \exp\left((1-\lambda_i)\left(\sum_{j=1}^{n} a_{ij}z_j\right)\right) = \sum_{j=1}^{n} a_{ij}z_j + \cdots$$

then after removing the parentheses in the product $\prod_{i=1}^{n} f_i(z)$ we have

$$\prod_{i=1}^{n} f_i(z) = \prod_{i=1}^{n}\left(\exp\left(\lambda_i\left(\sum_{j=1}^{n} a_{ij}z_j\right)\right) - \exp\left((1-\lambda_i)\left(\sum_{j=1}^{n} a_{ij}z_j\right)\right)\right)$$

$$= \sum_{k=0}^{n}(-1)^k \sum_{1\leqslant j_1<\cdots<j_k\leqslant n} \prod_{i=1}^{n} \exp\left(z_i\left(\sum_{j=1}^{n} \lambda_j a_{ij} - a_{ij_1} - \cdots - a_{ij_k}\right)\right).$$

According to the last formula and formula (55) it is easy to obtain formula (57), if we note that

$$\mathbf{res}_{z_i} \exp\left(z_i\left(\sum_{j=1}^{n} \lambda_j a_{ij} - a_{ij_1} - \cdots - a_{ij_k}\right)\right) z_i^{-2}$$

$$= \sum_{j=1}^{n} \lambda_j a_{ij} - a_{ij_1} - \cdots - a_{ij_k}, \quad i = 1, \ldots, n.$$

Formula (58) is equivalent to formula (57). Let us show that (58) follows from (57). Let the vector components $\widetilde{\gamma} = (\gamma_1, \ldots, \gamma_n)$ be small, and the matrix A be close to a unit matrix. Due to the continuity of the function $\det(A)$ with respect to the entries of the matrix A this implies that $\det(A) \neq 0$. By $\widetilde{\lambda} = (\lambda_1, \ldots, \lambda_n) = A^{(-1)}\widetilde{\gamma}$ the validity of formula (58) follows from formula (57): for $\widetilde{\gamma}$ near the origin and A close to the unit matrix. Since the expression in right part of (57) is a polynomial in its variables then by virtue of the uniqueness theorem the identity (58) holds for arbitrary values of the variables. Further, since $\gamma_i = \sum_j \lambda_j a_{ij}$, (57) follows from (58), and so these formulae are equivalent. □

REMARK 8. The polynomial identity (58) for the computation of the permanent for particular values of the free parameters $\gamma_i = 0$ and $\gamma_i = (\sum_{j=1}^{n} a_{ij})/2$, $i = 1, \ldots, n$, implies the well-known formulae of Ryser and Nijenhuis–Wilf [150] of inclusion–exclusion type. These formulae are the most effective ones at the present time for the calculation of permanents.

6. Combinatorial relations in the Collatz and Jacobian conjectures

6.1. *Characteristic function of the stopping height in the Collatz conjecture*

The $3x + 1$ problem is known under different names. It is often called *Collatz's problem*, *Ulam's problem*, *the Syracuse problem*, *Kakutani's problem*, *and Hasse's algorithm* [129].

Consider the Collatz sequence of iterations $(n, f(n), f(f(n)), \dots)$ where

$$f(n) = \begin{cases} (3n+1)/2, & \text{for odd } n, \\ n/2, & \text{for even } n. \end{cases} \tag{59}$$

The $3x+1$ conjecture states that for any natural number n this sequence will contain the number 1. For example, $n = 7$ will generate the sequence (7, 11, 17, 26, 13, 20, 10, 5, 8, 4, 2, 1, 2, 1, ...) The index of the first element equal to 1 in this sequence is called the *stopping height* of the instance of Collatz problem. The following arithmetic reformulation of the Collatz problem is given in [140]:

THEOREM 18. *The* $3x+1$ *conjecture is true if and only if for every positive integer a there are natural numbers w and v such that* $a \leqslant w$ *and*

$$\begin{aligned} &\binom{2w+1}{w}\binom{4(w+1)v+1}{v}\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\sum_{t=0}^{\infty}\binom{v}{r}\binom{w(v-r)}{s}\binom{wr}{t} \\ &\quad\times\binom{2s+2t+r+(4w+3)v+1}{3((4w+4)t+a)+2(4w+4)r+(4w+4)s} \\ &\quad\times\binom{3((4w+4)t+a)+2(4w+4)r+(4w+4)s}{2s+2t+r+(4w+3)v+1} \equiv 1 \ (\textit{modulo}\ 2). \end{aligned} \tag{60}$$

REMARK 9. Careful investigation of this result along with computer experiments shows that this formula and analogous statements [140, Theorem 1, Corollaries 1–3] are not valid. The following correction is required: the term a has to be replaced by $a/3$ in order to make it work. We will use the corrected version of (60) below.

Here we give the characteristic function of the stopping height of Collatz problem and some of its properties. Let H, $H \subset L$, be the set of formal Laurent power series with integer coefficients.

DEFINITION 5. Two series $A(w) = \sum_k a_k w^k$ and $B(w) = \sum_k b_k w^k$ from H are congruent, i.e. $A(w) \equiv B(w)$, if and only if $a_k \equiv b_k$ (modulo 2) for all k.

For example, if $\alpha = 2^x$, $x \in \mathbb{N}$, the following congruences hold:

$$(1+u)^{\alpha} \equiv 1 + u^{\alpha}, \qquad (1+u)^{\alpha-1} \equiv \sum_{s=0}^{\alpha-1} u^s, \tag{61}$$

$$\bigl(1-(\alpha-1)^2 u\bigr)^{-1/(\alpha-1)} \equiv \prod_{s=0}^{\infty}\bigl(1+u^{\alpha^s}\bigr). \tag{62}$$

The following proposition gives a characterization of the free parameters in (60) with the help of a well-known theorem of Kummer (1852).

PROPOSITION 4. *(See* [140,68].*)*

(1) $\binom{2w+1}{w} \equiv 1$ (*modulo* 2) $\Leftrightarrow w = 2^x - 1$, $x \in \mathbb{N}$.

(2) *If* $x, v \in \mathbb{N}$ *and* $w = 2^x - 1$, *then*

$$\binom{4(w+1)v+1}{v} \equiv 1 \ (\textit{modulo}\ 2)$$

$$\Leftrightarrow \quad v = \left(2^{r(x+2)} - 1\right)/\left(2^{(x+2)} - 1\right) = \sum_{i=0}^{r-1}\left(2^{i(x+2)}\right).$$

The following theorem from [68] contains an integral representation of (60) using the last proposition:

THEOREM 19. *Let* $a, v, w \in \mathbb{N}$ *and write*

$$\begin{aligned} S = \sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} & \binom{v}{r}\binom{w(v-r)}{s}\binom{wr}{t} \\ & \times \binom{2s+2t+r+(4w+3)v+1}{3(4w+4)t+2(4w+4)r+(4w+4)s+a} \\ & \times \binom{3(4w+4)t+2(4w+4)r+(4w+4)s+a}{2s+2t+r+(4w+3)v+1}. \end{aligned} \tag{63}$$

Then

$$S = \mathbf{res}_u\left\{g(u)u^{-(4w+3)v+a-2}\right\} \tag{64}$$

where

$$g(u) = \left(\left(1+u^{-2+(4w+4)}\right)^w + u^{-1+2(4w+4)}\left(1+u^{-2+3(4w+4)}\right)^w\right)^v. \tag{65}$$

PROOF. The product of the last two binomial coefficients in (63) is equal to

$$\delta\left(3(4w+4)t+2(4w+4)r+(4w+4)s+a, 2s+2t+r+(4w+3)v+1\right),$$

where $\delta(n,k)$ is the Kronecker symbol. Using the integral representation (1) for each of the first three binomial coefficients in (63), and $\delta(n,k) = \mathbf{res}_w w^{-n+k-1}$ for the product of the last two of them we get

$$\begin{aligned} S = \sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} & \mathbf{res}_x \frac{(1+x)^v}{x^{r+1}} \mathbf{res}_y \frac{(1+y)^{w(v-r)}}{y^{s+1}} \mathbf{res}_z \frac{(1+z)^{wr}}{z^{t+1}} \\ & \times \mathbf{res}_u u^{(-2+(4w+4))s+(-2+3(4w+4))t+(-1+2(4w+4))r-(4w+3)v+a-2}. \end{aligned}$$

Further, using the method of coefficients, we sum the last expression with respect to r, s and t. For this, using the linearity and substitution rules and successive changes of variables $x = (1+y)^{-w}(1+z)^w u^{(-1+2(4w+4))}$, $y = u^{(-2+(4w+4))}$ and $z = u^{(-2+3(4w+4))}$ we obtain

$$\begin{aligned} S = \sum_{s=0}^{\infty}\sum_{t=0}^{\infty} & \mathbf{res}_{y,z,u} \frac{((1+y)^w + (1+z)^w u^{(-1+2(4w+4))})^v}{y^{s+1}z^{t+1}} \\ & \times u^{(-2+(4w+4))s+(-2+3(4w+4))t-(4w+3)v+a-2} \end{aligned}$$

$$
\begin{aligned}
&= \cdots = \mathbf{res}_u \bigl\{ \bigl(\bigl(1 + u^{-2+(4w+4)}\bigr)^{w} \\
&\quad + u^{-1+2(4w+4)} \bigl(1 + u^{-2+3(4w+4)}\bigr)^{w} \bigr)^{v} u^{-(4w+3)v+a-2} \bigr\}. \qquad \square
\end{aligned}
$$

From Theorems 18, 19, Proposition 4 and the congruences (61), (62) we have

THEOREM 20. *The* $3x+1$ *conjecture is true if and only if for every positive integer* a *there are natural numbers* r *and* $\alpha = 2^{x+2}$, *where* $x \in \mathbb{N}$, *such that* $a \leqslant -1 + \alpha/4$, *and one of the equivalent congruences is true*:

$$
\begin{aligned}
&\mathbf{res}_u u^{-\alpha^r + a - 1} \bigl(\bigl(1 + u^{-2+\alpha}\bigr)^{-1+\alpha/4} \\
&\quad + u^{-1+2\alpha} \bigl(1 + u^{-2+3\alpha}\bigr)^{-1+\alpha/4} \bigr)^{(\alpha^r - 1)/(\alpha - 1)} \equiv 1 \quad (\mathit{modulo}\ 2), \\
&\mathbf{res}_u u^{-\alpha^r + a - 1} \prod_{t=0}^{\infty} \bigl(\bigl(1 + u^{(-2+\alpha)\alpha^t}\bigr)^{-1+\alpha/4} \\
&\quad + u^{(-1+2\alpha)\alpha^t} \bigl(1 + u^{(-2+3\alpha)\alpha^t}\bigr)^{-1+\alpha/4} \bigr) \equiv 1 \quad (\mathit{modulo}\ 2), \\
&\mathbf{res}_u u^{-\alpha^r + a - 1} \prod_{t=0}^{\infty} \Biggl(\sum_{s=0}^{-1+\alpha/4} \bigl(u^{s(-2+\alpha)\alpha^t} + u^{(-1+2\alpha+s(-2+3\alpha))\alpha^t} \bigr) \Biggr) \\
&\quad \equiv 1 \quad (\mathit{modulo}\ 2), \\
&\mathbf{res}_u u^{-\alpha^r + a - 1} \bigl[\bigl(1 - (\alpha - 1)^2 u^{-2+\alpha}\bigr)^{-1+\alpha/4} \\
&\quad - (\alpha - 1)^2 u^{-1+2\alpha} \bigl(1 - (\alpha - 1)^2 u^{-2+3\alpha}\bigr)^{-1+\alpha/4} \bigr]^{-1/(\alpha-1)} \\
&\quad \equiv 1 \quad (\mathit{modulo}\ 2).
\end{aligned} \tag{66}
$$

We call the function

$$
Q(u) = \prod_{t=0}^{\infty} \Biggl(\sum_{s=0}^{-1+\alpha/4} \bigl(u^{s(-2+\alpha)\alpha^t} + u^{(-1+2\alpha+s(-2+3\alpha))\alpha^t} \bigr) \Biggr)
$$

a *characteristic function* of the stopping height in the Collatz conjecture. It is shown in [68] that the coefficients of the function $Q(u) = \sum_k q_k u^k$ from H are equal to either 0 or 1. Therefore, the congruence (66) is a theoretical functional reformulation of the Collatz conjecture. It was also proven there, that (66) is equivalent to a known number theoretical reformulation of the problem [212].

It is well known that knowledge of the generating functions means a lot in combinatorial analysis. Therefore, the study of the properties of $Q(u)$ is important. Some useful properties and a recurrence relation for the coefficients of $Q(u)$ are to be found in [68,73] and a functional equation for $Q(u)$ is found in [69].

6.2. *Combinatorial identities in the Jacobian problem*

In 1939 O. Keller has stated the following conjecture [117]: *a the polynomial mapping* $f : \mathbb{C}^m \to \mathbb{C}^m$ *is polynomially invertible*, *if its Jacobian is a non-zero constant*.

A complete review of the current status of this problem including a description of various generalizations and attempts for its solution was given in 2000 by Arno van den Essen [77].

Using the classical [35,202,183] Cayley–Sylvester–Sack formula for two bivariate polynomials V. Stepanenko [199] has conjectured that the proof of the Jacobian conjecture can be reduced to the verification of a series of identities of the following type:

$$\begin{aligned}&\frac{(2n)!}{(n+1)!n!}-\frac{1}{(n+1)!}\sum_{k=1}^{\lfloor\frac{n+1}{2}\rfloor}(2n-2k+1)!(2k-2)!\sum_{m=1}^{k}\bigl(S_1^{(1)}+S_2^{(1)}\bigr)\\&\quad+\frac{1}{(n+1)!}\sum_{k=1}^{\lfloor\frac{n+1}{2}\rfloor}(2n-2k+2)!(2k-3)!\sum_{m=1}^{k}\bigl(S_1^{(2)}+S_2^{(2)}\bigr)\equiv 0,\\&\quad\forall n\geqslant 2,\end{aligned}\tag{67}$$

where

$$\begin{aligned}S_1^{(1)}&=*\sum_{j=2m-1}^{2k-1}\ \sum_{i=0}^{(2k-j-1)/2}2^{2k-2i-2m+1}\\&\quad\times\frac{(2k-j)}{(n-4k+i+j+1)!(2k-2i-j)!i!(2k+m-j-1)!(j-2m+1)!(m-1)!},\\S_2^{(1)}&=*\sum_{j=2m-1}^{2k-3}\ \sum_{i=0}^{(2k-j-3)/2}2^{2k-2i-2m+1}\\&\quad\times\frac{(2k-j-1)}{(n-4k+i+j+2)!(2k-2i-j-1)!i!(2k+m-j-2)!(j-2m+2)!(m-1)!},\\S_1^{(2)}&=*\sum_{j=2m-1}^{2k-1}\ \sum_{i=0}^{(2k-j-1)/2}2^{2k-2i-2m}\\&\quad\times\frac{(2k-j)}{(n-4k+i+j+2)!(2k-2i-j)!i!(2k+m-j-1)!(j-2m)!(m-1)!},\\S_2^{(2)}&=*\sum_{j=2m-1}^{2k-3}\ \sum_{i=0}^{(2k-j-3)/2}2^{2k-2i-2m}\\&\quad\times\frac{(2k-j-1)}{(n-4k+i+j+3)!(2k-2i-j-1)!i!(2k+m-j-2)!(j-2m+1)!(m-1)!}.\end{aligned}$$

Here we define $*\sum_{j=2m-1}^{2k-1}\cdots=\sum_{j=2m-1,2m-3,\dots}^{2k-1}$. If factorials with a negative arguments arise under the summation sign $\sum$ then we suppose these summands equal to 0.

In [72] Egorychev has stated the validity of (67) by means of the method of integral representation and computation of combinatorial sums.

THEOREM 21. *The identity (67) is valid.*

The proof consists, as usual, of the series of assertions and finds the integral representation for the intermediate combinatorial quantities.

PROPOSITION 5. *The sum $S_1^{(1)}$ admits the following integral representation*:

$$\frac{1}{(2\pi i)^3(n-1)!}\int_{\Gamma_3(\varepsilon)}\Big\{2^{2k-2m}w_1^{-2k+2m-1}w_3^{-2k+m-1}w_5^{-m}$$
$$\times\left(1+\frac{w_1^2}{4}+w_1w_3+\sum_{s=1,3,4,5}w_s\right)^{n-1}$$
$$\times\left(1+\frac{w_1}{2}\right)+2^{2k-2m}w_1^{-2k+2m-1}w_3^{-2k+m-1}w_5^{-m}$$
$$\times\left(1+\frac{w_1^2}{4}-w_1w_3+\sum_{s=1,3,4,5}w_s\right)^{n-1}\left(1+\frac{w_1}{2}\right)\Big\}\,dw,$$

where the skeleton $\Gamma_3(\varepsilon)=\{w=(w_1,w_3,w_5)\colon |w_1|=\varepsilon_1, |w_3|=\varepsilon_2, |w_5|=\varepsilon_5\}$.

PROPOSITION 6. *The sum $S_2^{(1)}$ admits the following integral representation*:

$$\frac{1}{(2\pi i)^3(n-1)!}\int_{\Gamma_3(\varepsilon)}\Big\{2^{2k-2m}w_1^{-2k+2m-1}w_3^{-2k+m-1}w_5^{-m}$$
$$\times\left(1+\frac{w_1^2}{4}+w_1w_3\sum_{s=1,3,4,5}w_s\right)^{n-1}\left(1+\frac{w_1}{2}\right)$$
$$-\,2^{2k-2m}w_1^{-2k+2m-1}w_3^{-2k+m-1}w^{-m}$$
$$\times\left(1+\frac{w_1^2}{4}-w_1w_3+\sum_{s=1,3,4,5}w_s\right)^{n-1}\left(1+\frac{w_1}{2}\right)\Big\}\,dw.$$

PROPOSITION 7. *The sum $S_1^{(2)}$ admits the following integral representation*:

$$\frac{1}{(2\pi i)^3(n-1)!}\int_{\Gamma_3(\varepsilon)}\Big\{2^{2k-2m-1}w_1^{-2k+2m}w_3^{-2k+m}w_5^{-m}$$
$$\times\left(1+\frac{w_1^2}{4}+w_1w_3+\sum_{s=1,3,4,5}w_s\right)^{n-1}\left(1+\frac{w_1}{2}\right)$$
$$+\,2^{2k-2m-1}w_1^{-2k+2m}w_3^{-2k+m}w_5^{-m}$$
$$\times\left(1+\frac{w_1^2}{4}-w_1w_3+\sum_{s=1,3,4,5}w_s\right)^{n-1}\left(1+\frac{w_1}{2}\right)\Big\}\,dw.$$

PROPOSITION 8. *The sum $S_2^{(2)}$ admits the following integral representation*:

$$\frac{1}{(2\pi i)^3(n-1)!}\int_{\Gamma_3(\varepsilon)}\Big\{2^{2k-2m-1}w_1^{-2k+2m}w_3^{-2k+m}w_5^{-m}$$
$$\times\left(1+\frac{w_1^2}{4}+w_1w_3+\sum_{s=1,3,4,5}w_s\right)^{n-1}\left(1+\frac{w_1}{2}\right)$$

$$- 2^{2k-2m-1} w_1^{-2k+2m} w_3^{-2k+m} w_5^{-m}$$
$$\times \left(1 + w_1 + \frac{w_1^2}{4} + w_3 - w_1 w_3 + w_5\right)^{n-1} \left(1 + \frac{w_1}{2}\right)\Bigg\} dw.$$

PROPOSITION 9. *For $n \geqslant 2$ the following formula is valid*:

$$\sum_{m=1}^{k} \left(S_1^{(1)} + S_2^{(1)}\right) = \frac{2}{(n-1)!} \binom{2n-1}{2k-2} \binom{n-1}{2k-1}.$$

PROPOSITION 10. *For $n \geqslant 2$ the following formula is valid*:

$$\sum_{m=1}^{k} \left(S_1^{(2)} + S_2^{(2)}\right) = \frac{2}{(n-1)!} \binom{2n-1}{2k-3} \binom{n-1}{2k-2}.$$

PROPOSITION 11. *For $n \geqslant 2$ the following identity is valid*:

$$\sum_{k=1}^{\lfloor \frac{n+1}{2} \rfloor} \left\{ -\binom{n-1}{2k-1} + \binom{n-1}{2k-2} \right\} = 0.$$

The standard completion of the proof is left to the reader [72].

7. Computer algebra algorithms for indefinite and definite summation

In this section we give a brief overview of some well-known summation algorithms used in computer algebra systems such as Maple [153], and establish links between them and methods of integral representation.

7.1. *Preliminaries*

Let K be a field of characteristic 0, and the E_k denote the shift operator with respect to k: $E_k F(k) = F(k+1)$. A closed form function $F(k)$ is said to have a closed form sum $G(k)$ (usually written as $G(k) = \sum_k F(k)$) if

$$(E_k - 1) G(k) = F(k). \tag{68}$$

REMARK 10. The term "closed form" can have different meanings depending on context. For example, it can mean that the function $F(k)$ is written in terms of elementary or special functions not using the summation sign. However this definition is too broad and weak. In computer algebra texts [84] it is sometimes narrowed to something like the following: $F(k)$ is in closed form if $E_k F(k)/F(k)$ is a rational function over K. In this case we are dealing with so called "hypergeometric terms" and this definition might be to narrow. However, often "closed form" summability means an ability to express the sum in the same elementary terms as the summand [59,128], and this obviously depends on the type of summand. We will used closed form in this broader context, making additional assumptions when necessary.

Note, that the **res** operator defined in Section 3.1, commutes not only with the definite summation operator, but also with the shift operator E_k, the difference operator $\Delta_k = E_k - 1$ and the indefinite summation operator $\sum_k$. This immediately allows one to apply the ideas of integral representation techniques to indefinite summation problems. If a summation problem under the **res** sign becomes a geometric summation problem, then the usual indefinite summability condition has a straightforward analog in the integral representation.

Let $F(k)$ have an integral representation $\mathbf{res}_w f(k,w)$ with a geometric kernel $f(k,w)$ with respect to k, i.e.

$$E_k f(k,w) = q(w) f(k,w).$$

Then

$$(E_k - 1) f(k,w) = \bigl(q(w) - 1\bigr) f(k,w),$$

and

$$\sum_k f(k,w) = \frac{1}{q(w)-1} f(k,w). \tag{69}$$

In other words, in an integral representation with a kernel that is geometric in k the application of the operator $E_k - 1$ corresponds to the multiplication of the kernel by $(q(w)-1)$, and the application of the operator $\sum_k$ corresponds to the multiplication of the kernel by $\frac{1}{q(w)-1}$. This last value is sometimes referred to as a "summation unit" in the integral representation literature. An operator form of the summability equality

$$(E_k - 1)\sum_k = \sum_k (E_k - 1) = 1$$

corresponds to the trivial "summation unit" cancelation equality

$$\bigl(q(w) - 1\bigr)\frac{1}{q(w)-1} = 1.$$

If $F(k) = \mathbf{res}_w f(k,w)$ and $G(k) = \mathbf{res}_w g(k,w)$, then the summability condition $(E_k - 1)G(k) = F(k)$ translates into divisibility of $f(k,w)$ by $(q(w) - 1)$. Observe, that $F(k)$, $G(k)$ do not need to be geometric or hypergeometric in order for $f(k,w)$, $g(k,w)$ to be geometric. The reasonable question here is, how often is the kernel of an integral representation of a given closed form function $F(k)$ is geometric? In our opinion often enough in order to try to apply integral representations in an algorithmic fashion.

Computer algebra traditionally considers several problems related to summation. If an indefinite sum $G(k)$ (solution of the first order recurrence (68)) cannot be found, one can try to solve the *additive decomposition problem*: construct two functions $R(k)$ and $H(k)$, such that

$$F(k) = (E_k - 1)R(k) + H(k), \tag{70}$$

where $H(k)$ is simpler than $F(k)$ is some sense. Or equivalently construct two functions $R(k)$ and $H(k)$, such that

$$\sum_k F(k) = R(k) + \sum_k H(k), \tag{71}$$

where $H(k)$ is simpler than $F(k)$. The measure of simplicity can be different for different classes of functions. For example if $F(k)$ is a rational function we can require both $R(k)$ and $H(k)$ to be rational with $H(k)$ having a denominator of lowest possible degree.

The definite summation problem is: find a closed form expression for

$$\sum_{k=m}^{n} F(k), \quad m \leqslant n,$$

in the case when the summand does not depend on the summation bounds m and n is usually solved by first computing an indefinite sum $G(k)$ in (68), and then using a discrete analog of the Newton–Leibnitz formula

$$\sum_{k=m}^{n} F(k) = G(n+1) - G(m). \tag{72}$$

The conditions under which (72) can be used are obvious in the case of rational summand $F(k)$. Recently [11–13], necessary and sufficient conditions for the applicability of (72) were obtained for the case of a hypergeometric (and, more generally, a P-recursive) summand $F(k)$.

7.2. *Summability criterion for rational functions*

The algorithmic treatment of rational summation and decomposition problems started with the work of S.A. Abramov [1,2]. There were a number of algorithms and improvements developed over the following years, see e.g. [5,136,160,168,104] (in particular [168] gives a complete overview of these algorithms and improvements to them). Most of the description of these algorithms explicitly avoid polynomial factorization in $K[k]$. Before discussing these approaches we recall a criterion of rational summability as found in [3,5] (we essentially quote the definitions and the criterion from [5] here).

Consider (71), assuming without loss of generality that $F(k)$ is a proper rational function. Temporarily replace the coefficient field K by its algebraic closure $\overline{K}$. The partial fraction decomposition of $F(k)$ has the form

$$F(k) = \sum_{i=1}^{m} \sum_{j=1}^{t_i} \frac{\beta_{ij}}{(k - \alpha_i)^j}. \tag{73}$$

Write $\alpha_i \sim \alpha_j$ if $\alpha_i - \alpha_j$ is an integer. Obviously, $\sim$ is an equivalence relation on the set $\{\alpha_1, \ldots, \alpha_m\}$. Each of the corresponding equivalence classes has a largest element in the sense that all the other elements of the class are obtained by subtracting positive integers from it. Let $\alpha_1, \ldots, \alpha_v$ be the largest elements of all the classes ($v \leqslant m$). Then (73) can be rewritten as

$$F(k) = \sum_{i=1}^{v} \sum_{j=1}^{l_i} M_{ij}(E_k) \frac{1}{(k-\alpha_i)^j}. \tag{74}$$

Here $M_{ij}(E_k)$ is a linear difference operator with constant coefficients (a polynomial in E_k over $\overline{K}$). Let $F(k)$ have the form (74) and suppose that (68) possesses a solution $R(k) \in K(k)$. The rational function $R(k)$ can be written in a form analogous to (74):

$$\sum_{i=1}^{v} \sum_{j=1}^{l_i} L_{ij}(E_k) \frac{1}{(k-\alpha_i)^j}. \tag{75}$$

This presentation is unique and therefore

$$(E_k - 1)L_{ij}(E_k) = M_{ij}(E_k). \tag{76}$$

From here we read of the

RATIONAL SUMMABILITY CRITERION. *A necessary and sufficient condition for existence of a rational solution of* (68) *is that for all* $i = 1, \ldots, v$; $j = 1, \ldots, l_i$ *there is an operator* $L_{ij}(E_k)$ *such that* (76) *holds.*

Then (68) has the solution (75) and all other rational solutions of (68) can be obtained by adding arbitrary constants. If at least one operator $M_{ij}(E_k)$ is not divisible by $E_k - 1$ then (68) has no rational solution. We want then to construct (71). Consider one term from (74) writing it for simplicity in the form

$$M(E_k) \frac{1}{(k-\alpha)^j}, \quad j \geqslant 1,$$

compute the left quotient $L(E_k)$ and left remainder w:

$$M(E_k) = (E_k - 1)L(E_k) + w, \quad w \in \overline{K}, \tag{77}$$

and write the right-hand side of (71) in the form

$$L(E_k) \frac{1}{(k-\alpha)^j} + \sum_k \frac{w}{(k-\alpha)^j}. \tag{78}$$

This gives a solution to the decomposition problem for this single term, since the denominator of the rational function under the sign of the indefinite sum has obviously the lowest possible degree.

Note that instead of (77) one can consider a reduction modulo $E_k - 1$ of the form

$$M(E_k) = (E_k - 1)V(E_k) + vE_k^c, \quad v \in \overline{K}, \tag{79}$$

where c is some convenient nonnegative integer. Different choices of reductions of the form (79) will leave the degree of the denominator of nonrational part intact, but can vary the degree of the denominator in the rational part of the decomposition (71).

7.3. *Analog of the criterion in integral representation terms*

Now we will describe an analog of this criterion in the integral representation framework. For this we will use the following representation:

$$\begin{aligned}\frac{1}{(k-\alpha)^j} &= \frac{1}{(j-1)!}\int_0^\infty e^{-(k-\alpha)u}u^{j-1}\,du\\ &= \frac{1}{(j-1)!}\int_0^\infty e^{-ku}\bigl(e^{\alpha u}u^{j-1}\bigr)\,du.\end{aligned} \tag{80}$$

Consider again just one term $M(E_k)\frac{1}{(k-\alpha)^j}$ $(j \geqslant 1)$, from (74). For clarity, let

$$\begin{aligned}M(E_k) &= \beta_t E_k^t + \cdots + \beta_1 E_k + \beta_0,\\ L(E_k) &= \gamma_{t-1}E_k^{t-1} + \cdots + \gamma_1 E_k + \gamma_0.\end{aligned}$$

Consider the two commutating polynomials with the same coefficients as M and L

$$P(X) = \beta_t X^t + \cdots + \beta_1 X + \beta_0, \qquad Q(x) = \gamma_{t-1}X^{t-1} + \cdots + \gamma_1 X + \gamma_0.$$

Noting that

$$\beta_l E_k^l \frac{1}{(k-\alpha)^j} = \frac{1}{(j-1)!}\int_0^\infty \beta_l e^{-lu}e^{-ku}\bigl(e^{\alpha u}u^{j-1}\bigr)\,du, \quad 0 \leqslant l \leqslant t,$$

write

$$M(E_k)\frac{1}{(k-\alpha)^j} = \frac{1}{(j-1)!}\int_0^\infty P\bigl(e^{-u}\bigr)e^{-ku}\bigl(e^{\alpha u}u^{j-1}\bigr)\,du,$$

i.e. the kernel of the integral representation of the summand is geometric in k with the base e^{-u}. Now (as in (69))

$$\sum_k M(E_k)\frac{1}{(k-\alpha)^j} = \frac{1}{(j-1)!}\int_0^\infty \frac{P(e^{-u})e^{-ku}}{e^{-u}-1}\bigl(e^{\alpha u}u^{j-1}\bigr)\,du.$$

If $M(E_k) = L(E_k)(E_k - 1)$ then $P(e^{-u}) = (e^{-u}-1)Q(e^{-u})$,

$$\sum_k M(E_k)\frac{1}{(k-\alpha)^j} = \frac{1}{(j-1)!}\int_0^\infty Q\bigl(e^{-u}\bigr)e^{-ku}\bigl(e^{\alpha u}u^{j-1}\bigr)\,du,$$

and we can use (80) backwards to get a closed form expression for the result of the rational summation. In other words, the condition of divisibility of $M(E_k)$ by $E_k - 1$ is equivalent to the divisibility of $P(e^{-u})$ by $(e^{-u} - 1)$ (or $P(X)$ by $(X - 1)$). The last in turn means that $P(1) = 0$ or $\sum_{l=0}^t \beta_l = 0$ and this condition has to hold for all terms (for all i, j) in (74) in order for $F(k)$ to be rationally summable.

Note that the property of coefficients of the full partial fraction decomposition proven in [142] is an immediate corollary of this summability criterion (divisibility of $M_{ij}(E_k)$ by $E_k - 1$ in (74)).

If $P(1) = w \neq 0$, then $P(X) - w$ is divisible by $X - 1$. In this case (77) corresponds to $P(X) = Q(X)(X - 1) + w$, and

$$\frac{1}{(j-1)!}\int_0^\infty Q\bigl(e^{-u}\bigr)e^{-ku}\bigl(e^{\alpha u}u^{j-1}\bigr)\,du,$$

provides the rational part of the decomposition (78) while

$$\frac{1}{(j-1)!}\int_0^\infty \frac{we^{-ku}}{e^{-u}-1}\bigl(e^{\alpha u}u^{j-1}\bigr)\,du$$

provides the nonrational part of the decomposition (78).

It is easy to see, that $P(X) - wX^c$ for $1 \leqslant c \leqslant t$ will be also divisible by $X - 1$, which leaves us with the same amount of flexibility in expressing the rational decomposition result as equation (79) does. Needles to say that very similar observations hold in the case of the definite rational summation.

7.4. *Rational summation algorithms*

Let $F(k) = \frac{f(k)}{g(k)}$. Define the dispersion of $F(k)$ ($\mathbf{dis}F(k)$) [1] to be the maximal integer distance between roots of the denominator $g(k)$. It can be computed e.g. as the largest nonnegative integer root of the resultant of polynomials $g(k)$ and $g(k + h)$. Write $\rho = \mathbf{dis}F(k)$. If $\rho = 0$ than we can take in (71) $R(k) = 0$ and $H(k) = F(k)$ (see [1,5,168]).

Now, let $\rho > 0$. All algorithms mentioned above do not directly use the summability criterion carefully avoiding factorization in $K[k]$, and fall into one of two categories.

- *Iterative* (Hermite reduction like) algorithms will start with $R(k) = 0$ and $H(k) = F(k)$ and decrease the dispersion of $H(k)$ by 1 at each iteration, reducing the nonrational part $H(k)$ and letting grow the rational part $R(k)$. The number of iterations is equal to ρ.
- *Non-iterative* (analogous to the Ostrogradsky algorithm) algorithms first build universal denominators $u(k)$ and $v(k)$ such that denominator of $R(k)$ will divide $u(k)$, denominator of $H(k)$ will divide $v(k)$, and then reduce the problem to linear algebra, solving a system of linear equations of size $\sim \deg u(k)$ (see [160,5,168]). In turn, usually $\deg u(k) = \Theta(\rho)$. The choice of $u(k)$ of the lowest possible degree is obviously crucial here. In [104] an algorithm which gives sharp bound $u(k)$ in the case when $F(k)$ is rational summable ($H(k) = 0$) is presented.

In both these classes of algorithms if $\rho = \mathbf{dis}F(k) \gg \deg g(k)$ the complexity of the rational function decomposition is defined by the value of ρ. Consider the following examples:

$$\sum_k \frac{-2k+999}{(k+1)(k-999)k(k-1000)} = \frac{1}{k(k-1000)},$$

$$\sum_k \frac{k^3-1998k^2+996999k+999999}{(k+1)(k-999)k(k-1000)} = \frac{1}{k(k-1000)} + \sum_k \frac{1}{k}.$$

The dispersion of the summand in both of these examples is 1001. On the one hand, iterative algorithms will require about 1001 steps of polynomial gcd computations. On the other hand, the universal denominator constructed by linear algebra based algorithms will have degree about 1001. In general, ρ can be as large as the magnitude of the trailing coefficient of the denominator of the summand. Thus, the cost of the iterative and linear algebra based algorithms for computing a decomposition as in (71) may be exponential in the size of the input.

It was already observed that while solving the rational decomposition problem (71) factorization in $K[k]$ should not be avoided. It is shown in [137] how the use of a factorization $g(k) = g_1(k)g_2(k)\dots g_t(k)$ improves the time of computing $\mathbf{dis}F(k)$. In [67] it was shown how after computing $\mathbf{dis}F(k)$ factorization (which is already performed and which is effective [105]) can be used to easily split $F(k)$ into shift equivalence classes and directly build (71) using several simple observations:

(a) if $\deg g_i(k) \neq \deg g_j(k)$ then linear factors over $\overline{K}$ of the denominators $g_i(k)$ and $g_j(k)$ fall into different shift equivalence classes in (74);

(b) finding among given polynomials of equal degree those which are shift equivalent is an easy task [137];

(c) full partial fraction decomposition over $\overline{K}$ is effective [29] at least when $K = Q$;

(d) computing a quotient and remainder in (77) is trivial (e.g., the remainder is obtained by substitution of 1 instead of E_k into $M(E_k)$ in (77));

(e) treating different shift equivalence classes separately will allow one to minimize the degree of the denominator of the rational part in each class as in [168].

The complexity of partial fraction decomposition does not depend on the dispersion of the given rational function, which means that at least in the case $\mathbf{dis}F(k) \gg \deg g(k)$ this leads to practical and efficient algorithms as shown in [67]. The efficiency of this straightforward solution was confirmed by a prototype implementation in Maple.

Later in [85] a new polynomial time algorithm was proposed that does not use factorization in $K[k]$ but instead applies *shiftless* factorization in order to directly build (71).

There are even more reasons not to avoid partial fraction decomposition in the case of the definite rational summation. We refer here to summation problems of the form $\sum_{k=0}^{n} f(k,n)$, where partial fraction decomposition of $f(k,n)$ with respect to k has pairs of terms as, e.g.

$$\frac{1}{nk+1} - \frac{1}{n(n-k)+1} \quad \text{or} \quad \frac{1}{k^2+1} - \frac{1}{(n-k)^2+1}.$$

These kinds of problems are not directly treatable by known summation algorithms, which avoid factorization. For example, the “W–Z” method is not applicable here, because such terms are not proper hypergeometric. The usual answer from computer algebra systems for this type of summation involves a linear combination of values of the Ψ function, which is equivalent to 0 but is not recognized as such. After performing full partial fraction decomposition it takes little effort to find such cases. The use of an integral representation is even more advisable here, because after performing geometric definite summation under integral sign the terms in the kernel of the integral representation corresponding to the terms above of the input expression will cancel each other. Note, that problem of this kind was recently solved in [106] but the validity of the algorithm is still to be proven.

7.5. *Gosper summation of hypergeometric terms*

A function $F(k)$ is called a *hypergeometric term* if the ratio $E_k F(k)/F(k)$ is a rational function. This rational function is called a *certificate* of $F(k)$. The Gosper algorithm [93] for the indefinite summation problem (68) is based on the observation that, given a hypergeometric term $F(k)$, if the hypergeometric term $G(k)$ in (68) exists, it is a rational multiple of $F(k)$, i.e. $G(k) = V(k)F(k)$, where $V(k) \in K(k)$. An outline of the Gosper algorithm is as follows:

1. Compute the certificate $r(k) = E_k F(k)/F(k)$.
2. Write $r(k)$ in Gosper–Petkovšek form [166]
$$r(k) = \frac{EP(k)}{P(k)} \frac{Q(k)}{R(k)}, \tag{81}$$
where P, Q, R are polynomials and $\gcd(Q, E^j R) = 1$ for $j \in \mathbb{N}$.
3. Find a polynomial solution $y(k)$ of the recurrence
$$QE_k y(k) - Ry(k) = P. \tag{82}$$
4. If that solution exists, write
$$G(k) = \frac{y(k)R(k)}{P(k)} F(k). \tag{83}$$

For example, let $F(k) = k!$, than the Gosper algorithm proceeds as follows:

1. $r(k) = (k+1)!/k! = (k+1)$.
2. $r(k) = \frac{1}{1}\frac{k+1}{1}$. Thus, $P = 1$, $Q = (k+1)$, $R = 1$.
3. Find a polynomial solution of the recurrence equation
$$(k+1)Ey(k) - y(k) = 1. \tag{84}$$
4. No solution: $\sum_k k!$ has no closed form.

If instead we take $F(k) = k \cdot k!$ then

1. $r(k) = (k+1)!(k+1)/(k!k) = (k+1)^2/k$.
2. $r(k) = \frac{(k+1)}{k}\frac{k+1}{1}$, so that $P = k$, $Q = (k+1)$, $R = 1$.
3. Find a polynomial solution of the recurrence equation
$$(k+1)Ey(k) - y(k) = k \tag{85}$$
4. By inspection $y(k) = 1$ solves (85). Write
$$\sum_k k \cdot k! = \frac{y(k)R(k)}{P(k)} f(k) = \frac{1 \cdot 1}{k} k \cdot k! = k!.$$

7.6. *Two approaches to a summation of a particular hypergeometric term*

Consider the following type of the indefinite summation problem:

$$\sum_k k^\alpha k!, \tag{86}$$

where α is nonnegative integer. For which values of α the sum (86) has closed form? As we have seen in the previous section, for $\alpha = 0$ there is no closed form sum, and for $\alpha = 1$ there is a closed form sum.

GOSPER APPROACH. Applying the Gosper algorithm to $F(k) = k^\alpha k!$ we get: $r(k) = \frac{(k+1)^\alpha}{k^\alpha}\frac{k+1}{1}$. Hence, $P = k^\alpha$, $Q = (k+1)$, $R = 1$, and we need to find a polynomial solution of the recurrence equation

$$(k+1)E_k y(k) - y(k) = k^\alpha \tag{87}$$

with degree n at most equal to $\alpha - 1$. Substituting $y(k) = a_n k^n + a_{n-1}k^{n-1} + \cdots + a_1 k + a_0$ into (87), we obtain the following $(n+2) \times (n+1)$ system of linear equations

$$\begin{pmatrix} 0 & \dots & 0 & 0 & 0 & 1 \\ 0 & \dots & 0 & 0 & 1 & \binom{n+1}{n} \\ 0 & \dots & 0 & 1 & \binom{n}{n-1} & \binom{n+1}{n-1} \\ 0 & \dots & 1 & \binom{n-1}{n-2} & \binom{n}{n-2} & \binom{n+1}{n-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & \binom{2}{1} & \dots & \binom{n-1}{1} & \binom{n}{1} & \binom{n+1}{1} \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} 1 \\ a_n \\ a_{n-1} \\ \vdots \\ a_0 \end{pmatrix}. \tag{88}$$

Observe, that the truncated system

$$\begin{pmatrix} 0 & \dots & 0 & 0 & 0 & 1 \\ 0 & \dots & 0 & 0 & 1 & \binom{n+1}{n} \\ 0 & \dots & 0 & 1 & \binom{n}{n-1} & \binom{n+1}{n-1} \\ 0 & \dots & 1 & \binom{n-1}{n-2} & \binom{n}{n-2} & \binom{n+1}{n-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & \binom{2}{1} & \dots & \binom{n-1}{1} & \binom{n}{1} & \binom{n+1}{1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} 1 \\ a_n \\ a_{n-1} \\ \vdots \\ a_1 \end{pmatrix} \tag{89}$$

always has a solution, and that last equation of (88): $a_0 + a_1 + \cdots + a_n = a_0$ is equivalent to

$$\sum_{i=1}^{n} a_i = 0. \tag{90}$$

REGULAR AND COMPLEMENTARY BELL NUMBERS. The Bell numbers (the number of all partitions of a set of n objects) are defined by

$$B_n = \sum_{k=0}^{n} \left\{ {n \atop k} \right\},$$

where the $\left\{ {n \atop k} \right\}$ are the Stirling numbers of the second kind (Stirling subset numbers)

$$\left\{ {n \atop k} \right\} = k \left\{ {n-1 \atop k} \right\} + \left\{ {n-1 \atop k-1} \right\}, \qquad \left\{ {n \atop 0} \right\} = \delta_{0n}.$$

An alternative definition is

$$B_n = \frac{1}{e}\sum_{k=0}^{\infty}\frac{k^n}{k!},$$

and the exponential generating function is

$$b(x) = e^{e^x-1} = \sum_{n=0}^{\infty}\frac{B_n}{n!}x^n.$$

The complementary Bell numbers (the difference between the number of partitions of a set of n objects into an even and odd number of subsets) are defined by

$$\widetilde{B}_n = \sum_{k=0}^{n}(-1)^k\left\{ {n \atop k} \right\}, \tag{91}$$

where, again, the $\left\{ {n \atop k} \right\}$ are the Stirling numbers of the second kind. An alternative definition is

$$\widetilde{B}_n = e\sum_{k=0}^{\infty}\frac{(-1)^k k^n}{k!},$$

with the exponential generating function

$$\tilde{b}(x) = e^{1-e^x} = \sum_{n=0}^{\infty}\frac{\widetilde{B}_n}{n!}x^n.$$

Note that these numbers are also called Uppuluri–Carpenter numbers. However, the term complimentary Bell numbers better reflects the relation between the generating functions: $b(x)\tilde{b}(x) = 1$.

INTEGRAL REPRESENTATION APPROACH. Let α be a natural number. Write

$$k^\alpha = \sum_{l=0}^{\alpha}(-1)^{\alpha+l}\left\{ {\alpha+1 \atop l+1} \right\}(k+1)^{\bar{l}}.$$

From here using the integral representation of the Γ function, that is, $\Gamma(k+1) = k! = \int_0^\infty e^{-t}t^k\,dt$, write

$$(-1)^\alpha k^\alpha k! = \sum_{l=0}^{\alpha}(-1)^l(k+l)!\left\{ {\alpha+1 \atop l+1} \right\} = \int_0^\infty e^{-t}t^k L(t)\,dt,$$

where

$$L(t) = \sum_{l=0}^{\alpha}(-1)^l\left\{ {\alpha+1 \atop l+1} \right\}t^l.$$

Now

$$(-1)^{\alpha}\sum_k k^{\alpha}k! = \sum_k \int_0^{\infty} e^{-t}t^k L(t)\,dt = \int_0^{\infty} e^{-t}L(t)\Bigl(\sum_k t^k\Bigr)\,dt$$
$$= \int_0^{\infty} e^{-t}L(t)\frac{t^k}{t-1}\,dt = \sum_{l=0}^{\alpha} r_l\bigl((n+l)!\bigr) + L(1)\sum_k k!,$$

where $r_l = \sum_{i=l+1}^{\alpha}(-1)^i \left\{ {\alpha+1 \atop i} \right\}$, $l = 0, 1, \dots, \alpha - 1$. Note that

$$L(1) = \sum_{l=0}^{\alpha}(-1)^l \left\{ {\alpha+1 \atop l+1} \right\} = -\widetilde{B}_{\alpha+1}.$$

Therefore,

PROPOSITION 12. *The following equality holds*:

$$\sum_k k^{\alpha}k! = (-1)^{\alpha}\sum_{l=0}^{\alpha} r_l\bigl((n+l)!\bigr) + (-1)^{\alpha+1}\widetilde{B}_{\alpha+1}\sum_k k!, \tag{92}$$

and the existence of a closed form expression for $\sum_k k^{\alpha}k!$ *is equivalent to* $\widetilde{B}_{\alpha+1} = 0$, *where* $\widetilde{B}_{\alpha+1}$ *is a complementary Bell number.*

The number $\widetilde{B}_2 = 0$, which means that $k^1k!$ is summable.

REMARK 11. It is not known whether all complementary Bell numbers but the second one are nonzero [130].

REMARK 12. Let a_i, $i = 0, \dots, n$, be solutions of (88). It comes as no surprise that

$$\sum_{i=1}^{n} a_i = (-1)^{n+1}\widetilde{B}_{n+2} = (-1)^{\alpha}\widetilde{B}_{\alpha+1}$$

Consider (89) again and solve it for different values of α. The left-hand side of (90) gives a new representation of a complementary Bell number different from the representation (91). This representation (which comes from the Gosper algorithm) might have some interesting properties and deserves additional study.

Note that the algorithm from [9] will produce the same decomposition as (92) for every particular value α. Still, formula (92) is very attractive, since it has explicit appearance of the complementary Bell number, is defined for all nonnegative integer α, and can be easily generalized to a summand of the form $P_{\alpha}(k)k!$ where $P_{\alpha}(k)$ is an arbitrary polynomial of degree α. Direct use of the Gosper algorithm for large values of α will lead to the large linear systems to be solved.

7.7. *Zeilberger algorithm of creative telescoping*

Consider now a summand of the form $F(n,k)$ in the definite summation problem

$$G(n,k)=\sum_{k=a(n)}^{b(n)}F(n,k). \tag{93}$$

Let $F(n,k)$ be hypergeometric in both variables n and k, i.e. both $\frac{E_nF(n,k)}{F(n,k)}$ and $\frac{E_kF(n,k)}{F(n,k)}$ are rational functions. If the Gosper algorithm fails to find $G(n,k)$ such that $G(n,k+1)-G(n,k)=F(n,k)$ then one can try to solve (93) using Zeilberger's [215] algorithm of *creative telescoping*. This algorithm will try to construct a linear difference operator in E_n with coefficients that are polynomials in n and $G(n,k)$ such that

$$LF(n,k)=G(n,k+1)-G(n,k). \tag{94}$$

Applying $\sum_{k=a(n)}^{b(n)}$ to both sides of (94) and using commutation properties of linear operators we will get

$$L\left\{\sum_{k=a(n)}^{b(n)}F(n,k)\right\}=G\big(n,b(n)+1\big)-G\big(n,a(n)\big), \tag{95}$$

and the summation problem is reduced to the problem of solving inhomogeneous linear difference equation with polynomial coefficients and generally a d'Alembertian right-hand side. For the last problem one can use an algorithm from [6] or [166]. For example, let $F(n,k)=\frac{(-1)^kk^2\binom{n}{k}}{k-a}$, then in (94)

$$L=(a-1-n)E_n+n+1$$

and $G(n,k)$ is equal to

$$\frac{(n+1)k(-1)^k(-n-1+kn)(k-1)\binom{n}{k}}{(-k+a)n(n-1)} + \frac{(a-1-n)k(-1)^k(-n-1+kn)(k-1)\binom{n+1}{k}}{(-k+a)n(n-1)}.$$

Solving (95) with right-hand side $G(n,n+1)-G(n,0)$ gives

$$\sum_{k=0}^{n}\frac{(-1)^kk^2\binom{n}{k}}{k-a}=-\frac{\Gamma(3-a)a\Gamma(n+1)}{(2-3a+a^2)\Gamma(n+1-a)}, \tag{96}$$

which can be further simplified to $-a/\binom{n-a}{n}$.

Zeilberger's algorithm, originally developed for hypergeometric terms $F(n,k)$, was later generalized to holonomic functions [38,39]. Sufficient conditions for the applicability of this algorithm to a hypergeometric term can be found in [166], and necessary conditions in [10]. Note, that although this algorithm is general and can be applied to a wide variety of problems of the form (93), the construction of L and $G(n,k)$ in (94) can be very expensive even for the hypergeometric case. The order of L can be arbitrary high and algorithm starts trying to construct L of the lowest possible order r, $r+1$, etc. When L is finally constructed it can also be the case that solving of a high-order equation (95) is very time-consuming.

The existence of a general algorithm does not exclude the existence of more efficient methods for particular cases of the problem, each of which can have some level of generality. For example, definite summation problem in (96) is a particular case of the following problem:

$$\sum_{k=0}^{n}(-1)^k \frac{P(k)}{Q(k)}\binom{n}{k}, \tag{97}$$

where $P(k)$ and $Q(k)$ are polynomials. Using the method of integral representations one can obtain the following formula:

$$\sum_{k=0}^{n}(-1)^k \frac{P(k)}{Q(k)}\binom{n}{k} = \sum_{\{\alpha:\ Q(\alpha)=0\}} \frac{u(\alpha)P(\alpha)}{\alpha}\binom{n+\alpha}{n}^{-1}, \tag{98}$$

where $u(\alpha)$ is the appropriate coefficient in the full partial fraction decomposition of $1/Q(k)$ and (as usual) $Q(k)$ has no integer zeros on the summation interval. This particular method was implemented [67] in Maple using pattern matching and the direct use of (98). It returns the following answer in a fraction of a second

$$\begin{aligned}
&\sum_{k=0}^{n} \frac{(-1)^k k^7 \binom{n}{k}}{(k+a)(k+b)(k+c)(k+d)} \\
&\quad = \frac{a^6\binom{n+a}{n}^{-1}}{(-b+a)(-c+a)(-d+a)} - \frac{b^6\binom{n+b}{n}^{-1}}{(-b+a)(-c+b)(-d+b)} \\
&\qquad - \frac{d^6\binom{n+d}{n}^{-1}}{(-d+a)(-d+b)(-d+c)} + \frac{c^6\binom{n+c}{n}^{-1}}{(-c+a)(-c+b)(-d+c)}.
\end{aligned}$$

At the same time applying the general Zeilberger algorithm to this summand leads to the construction of a 4th order operator L and $G(n,k)$ containing polynomials up to degree 10, and solving the difference equation (95) does not produce an answer.

7.8. *Accurate summation*

The algorithm for accurate summation [8] generalizes the Gosper algorithm. Given $F(k)$ let the minimal annihilating difference operator L of order v with coefficients from $K(k)$ be known, i.e. $LF(k) = 0$. If $\sum_k F(k)$ has a minimal annihilating operator $\tilde{L}$ of the same order v, then it is easy to construct a difference operator V of order $v-1$, such that $\sum_k F(k) = VF(k)$. Given L and $F(k)$ the accurate summation algorithm checks for the existence of an $\tilde{L}$ of the same order as L and returns the expression for the sum $VF(k)$ in case of success.

For example, $F(k) = \Gamma(k+1) - \Gamma(k) - \Psi(k)$ has a minimal annihilating operator

$$L = E_k^3 - \frac{k^5 + 8k^4 + 22k^3 + 34k^2 + 27k + 5}{(k^3 + 3k^2 + 5k + 6)k} E_k^2$$

$$
\begin{aligned}
&+\frac{12k^4+31k^3+42k^2+26k+5+2k^5}{(k^3+3k^2+5k+6)k}E_k\\
&-\frac{5k^3+12k^2+13k+5+k^4}{k^3+3k^2+5k+6}.
\end{aligned}
$$

The accurate summation algorithm computes

$$
\begin{aligned}
\tilde{L} = &-\frac{(k+1)(k^3-5k^2+4k-2)}{(k^2+k+3)k}E_k^3\\
&+\frac{k^5-14k^3-10k^2-4+2k}{(k^2+k+3)k}E_k^2\\
&-\frac{2k^5-2k^4-16k^3-15k^2-5k-2}{(k^2+k+3)k}E_k\\
&+\frac{k^4-k^3-5k^2-5k-2}{k^2+k+3}
\end{aligned}
$$

of the same order and

$$
\begin{aligned}
V = &\frac{(k+1)(k^3-5k^2+4k-2)}{(k^2+k+3)k}E_k^2\\
&-\frac{k^5-k^4-10k^3-9k^2-2}{(k^2+k+3)k}E_k\\
&+\frac{k^4-k^3-6k^2-6k-5}{k^2+k+3}, \quad \text{such that}\\
\sum_k F(k) = {}& VF(k)\\
= {}& \frac{(k^4-k^3-6k^2-6k-5)(\Gamma(k+1)-\Gamma(k)-\Psi(k))}{k^2+k+3}\\
&-\frac{(k^5-k^4-10k^3-9k^2-2)(\Gamma(k+2)-\Gamma(k+1)-\Psi(k+1))}{(k^2+k+3)k}\\
&+\frac{(k+1)(k^3-5k^2+4k-2)(\Gamma(k+3)-\Gamma(k+2)-\Psi(k+2))}{(k^2+k+3)k}.
\end{aligned}
$$

In order to apply the accurate summation algorithm, knowledge of the minimal annihilating operator is crucial. If $F(k)$ is d'Alembertian expression, i.e., it has a completely factored annihilating operator

$$\big(E-r_1(k)\big)\big(E-r_2(k)\big)\dots\big(E-r_v(k)\big)$$

than the algorithm from [7] can be used to construct a minimal annihilator for $F(k)$. Observe, that the algorithm from [7] iteratively constructs a minimal completely factorable annihilating operator term by term and uses accurate summation on each iteration. Note, that the expression from the example above is d'Alembertian.

In the case when $F(k)$ is a hypergeometric term it has a first order minimal annihilating operator, and Gosper summability of the term $F(k)$ means that $G(k)$ in (83) also has a first order minimal annihilating operator.

REMARK 13. The main tools used by the algorithms mentioned above are methods of solving linear difference equations with rational function coefficients. A search for rational [4] solutions of such equations is used in Gosper and accurate summation, and a search for hypergeometric [165] solutions is used in Zeilberger's algorithm and accurate summation.

7.9. *Conclusion*

It is worth to note that computer algebra systems such as Maple contain implementations of many useful tools that can help one to solve summation problems. They also contain more general tools, such as the Maple package GFUN [185], allowing one to work with generating functions. All those tools are rapidly developing and new fast algorithms are being implemented.

We tried to establish links between those algorithms and the integral representation approach. There is another kind of link between discrete and continuous cases. It is established by abstraction of those two cases (and many more, e.g. q-difference case) in terms of Ore algebras. Ore polynomials are used to describe linear differential, difference, q-difference, etc. equations with rational function coefficients and allow one to develop universal algorithms of solving such equations. The idea of using Ore polynomials in computer algebra was proposed by M. Bronstein and M. Petkovšek in [30] and underwent intensive development by many specialists in computer algebra.

Generally, the discrete case is more difficult than the continuous case (compare for example the Ostrogradsky algorithm for the integration of rational functions to Abramov's algorithms of summation) and many algorithms for difference equations are hinted at by older well-known algorithms for differential equations. However there are examples of an opposite development, for example, Zeilberger's algorithm of creative telescoping was extended to the continuous case and is used for integration [203]. As usual in mathematics discrete and continuous approaches supplement each other and research is in this direction is useful and interesting.

8. Future developments

As academician A.I. Maltsev pointed out the representation of combinatorial relations in the algebra of analytic functions is unique. Such an approach is implemented in Sections 2 and 5 of this review, while Section 3 uses the algebra of polynomials (entire functions), or in the case of convergence [76] the algebra of analytic functions in a neighborhood of zero. At the same time, the original problem (combinatorial, enumerative, graph theoretical, etc.) is usually formulated in terms of sequences. The type and the algebra of the sequences is generally speaking arbitrary. This leads to the necessity of the development of a method of coefficients for the wide class of cases presented below.

Develop a method of coefficients in algebras of generating functions of the following types including generating functions of type $A(z) = \sum_{n=o}^{\infty} a_n \varphi_n(z)$. In particular,

1. Euler power series [16,17]: $\varphi_n(z) = \frac{z^n}{(1-q)(1-q^2)\ldots(1-q^n)}$, $n = 0, 1, 2, \ldots$.
2. Interpolation series especially series in generalized powers [110]: $\varphi_n(z) = \frac{z(z-1)\ldots(z-n+1)}{n!}$, $n = 0, 1, 2, \ldots$.
3. Dirichlet series [138,82,131].
4. Quasi-exponential series [18]: $\varphi_n(z) = \frac{z^n}{(z-1)(z-2)\ldots(z-n)}$, $n = 0, 1, 2, \ldots$.
5. Fourier series with $\{\varphi_n(z)\}_0^\infty$ a sequence of orthogonal polynomials of different types, and trigonometric series.
6. Power series over algebraic systems, especially over finite fields.
7. Asymptotics of coefficients of formal power series and pairs of invertible combinatorial relations [45,48,82,211].
8. Formal power series with noncommuting variables in formal languages of automata theory and combinatorics [184,46,186,187].
9. Möbius functions in incidence algebras of partially ordered set theory [49,21,168, 180,113].
10. Boolean functions and functions of multivalued logic.
11. Series with nonstandard φ-operations over number fields: if $a, b \in \mathbb{R}$, $a \uplus b = \varphi(\varphi^{(-1)}(a) + \varphi^{(-1)}(a))$, $a \odot b = \varphi(\varphi^{(-1)}(a) \times \varphi^{(-1)}(a))$, where $\overline{R} = \mathbb{R} \cup (\infty)$, $D_1, D_2 \subset \overline{R}$ and $\varphi : D_1 \longrightarrow D_2$ be an arbitrary one-to-one mapping.
12. Continued fractions [112].
13. Generating functions of two and more variables as listed in Problems 1–12 above.
14. Generating functions of two and more variables of "mixed" type.
15. Use of the ideas and techniques of the integral representations establishing connections between generating functions of different types. Effective example of such type is use of direct and inverse integral Mellin transforms in establishing relation between power series and Dirichlet series (see [82]).

In each case it is necessary to solve the following problems:

- to give algorithmic and, if possible, integral definition (representation) for the operator L of coefficients of the generating function; write out the system rules for the operator L and to prove a completeness lemma;
- to obtain corresponding analogues of the Bürmann–Lagrange series for inverting an implicit function, if this operation is admissible in the algebra of series under consideration. To construct an analog of the matrix of type R and to give their algebraic, combinatorial and asymptotic characterizations, including classification, inversion, product and decomposition theorems.

In cases when Problems 1–15 have been solved positively, extend the uniform approach of the integral representation of sums for finding integral representation and computing combinatorial sums of these different types, including sums with q-combinatorial numbers, sums over partitions, sums over divisors, infinite sums generated by Fourier series, sums with linear constraints on the summation indices and others. These problems are of special interest and arise from applications such as, simplification of expressions in solutions of enumerative combinatorial problems, asymptotic expansion, etc. Besides calculation it is interesting to study other properties of combinatorial sums, including recurrence relations, asymptotic, upper and lower bounds, unimodality, estimates of the complexity of calculation etc. Therefore there arise a lot of new such problems for integrals and operators. The complex approach also requires the creation of computer implementations of algorithms for the investigation of combinatorial sums. Some of these problems can be solved without difficulties as they are based on classic and known results. However, many of them require further investigations.

Acknowledgements

The authors would like to thank S.A. Abramov for many fruitful discussions and I. Kotsireas for his help and useful remarks. We also would like to thank M. Hazewinkel for his patience and support of this work during many years. This support has led to the significant improvements of the structure of this paper.

References

[1] S.A. Abramov, On the summation of rational functions, USSR Comput. Math. Math. Phys. 11 (1971); transl. from: Zh. Vychisl. Math. Math. Fiz. 11 1071–1075.

[2] S.A. Abramov, The rational component of the solution of a first-order linear recurrence relation with a rational right-hand side, USSR. Comput. Maths. Math. Phys. 15 (1975), transl. from: Zh. Vychisl. Math. Math. Fiz. 15 1035–1039.

[3] S.A. Abramov, Solving difference equations of the second order with constant coefficients in the field of rational functions, USSR Comput. Math. Math. Phys. 17 (1977), transl. from: Zh. Vychisl. Math. Math. Fiz. 17.

[4] S.A. Abramov, Rational solutions of linear differential and difference equations with polynomial coefficients, Zh. Vychisl. Mat. Mat. Fiz. 29 (1989) 1611–1620.

[5] S.A. Abramov, Indefinite sums of rational functions, in: Proc. of the 1995 Internat Symposium on Symbolic and Algebraic Computation, ISSAC'1995, Montreal, Canada, ACM Press, 1995, pp. 303–308.

[6] S.A. Abramov, E.V. Zima, D'Alembertian solutions of linear inhomogeneous equations (differential, difference, and some other), in: Proc. of ISSAC'1996, ACM Press, 1996, pp. 232–240.

[7] S.A. Abramov, E.V. Zima, Minimal completely factorable annihilators, in: Proc. of ISSAC'1997, ACM Press, 1997, pp. 290–297.

[8] S.A. Abramov, M. van Hoeij, Integration of solutions of linear functional equations, Integral Transform. Spec. Funct. 8 (1999) 3–12.

[9] S.A. Abramov, M. Petkovšek, Rational normal forms and minimal decompositions of hypergeometric terms, J. Symbolic Comput. 33 (2002) 521–543.

[10] S.A. Abramov, When does Zeilberger's algorithm succeed? Adv. Appl. Math. 30 (2003) 424–441.

[11] S.A. Abramov, M. Petkovšek, Gosper's algorithm, accurate summation, and the discrete Newton–Leibniz formula, in: Proc. of ISSAC'2005, ACM Press, 2005, pp. 5–12.

[12] S.A. Abramov, On the summation of p-recursive sequences, in: Proc. of ISSAC'2006, ACM Press, 2006.

[13] S.A. Abramov, M. Petkovšek, Hypergeometric summation revisited, in: Computer Algebra 2006: Latest Advances in Symbolic Algorithms, World Scientific, 2007, pp. 1–11.

[14] L.A. Aizenberg, G.P. Egorychev, Integral representations with Szegő kernels for functions that are holomorphic in unbounded n-circular domains, Dokl. Akad. Nauk USSR 231 (1976) 265–268 (in Russian).

[15] L.A. Aizenberg, A.P. Yuzhakov, Integral Representation and Residues in Multidimensional Complex Analysis, Nauka, Novosibirsk, 1979 (in Russian).

[16] G.E. Andrews, On the foundations of combinatorial theory. IV. Finite vector space and Eulerian generating functions, Stud. Appl. Math. 49 (1970) 239–258.

[17] G.E. Andrews, Identities in combinatorics. II. A q-analog of the Lagrange inversion theorem, Proc. Amer. Math. Soc. 53 (1975) 240–245.

[18] G.V. Badaljan, Quasiexponential Series and Quasianalytic Classes of Functions, Nauka, Moscow, 1990 (in Russian).

[19] C. Banderier, M. Bousquet-Mélou, A. Denise, P. Flajolet, D. Gardy, D. Gouyou-Beuchamps, Generating functions for generating trees, Discrete Math. 246 (2002) 29–55.

[20] E. Barcucci, Lungo Del, E. Pergola, R. Pinzani, Eco: A methodology for the enumeration of combinatorial objects, J. Difference Equ. Appl. 5 (1999) 435–490.

[21] M. Barnabei, A. Brini, G. Nicoletti, Recursive matrices and umbral calculus, J. Algebra 75 (1982) 546–573.

[22] W. Basler, From Divergent Power Series to Analytic Functions. Theory and Application of Multisummable Power Series, Springer-Verlag, Berlin, 1994.

[23] W. Basler, Moment methods and formal power series, J. Math. Pures Appl. 76 (1997) 285–305.

[24] E.A. Bender, Asymptotic methods in enumeration, SIAM Rev. 16 (1974) 485–515.

[25] E.A. Bender, L.B. Richmond, A multivariate Lagrange inversion formula for asymptotic calculations, Electron. J. Combin. 5 (1998) 2–4.

[26] E.A. Bender, L.B. Richmond, Multivariate asymptotics for products of large powers with applications to Lagrange inversion, Electron. J. Combin. 6 (1999) 2–21.

[27] M.A. Borodin, Certain integral representations of functions holomorphic in doubly circular domains, Sibirsk. Mat. Zh. 10 (1969) 287–296 (in Russian).

[28] D.M. Bressoud, A matrix inverse, Proc. Amer. Math. Soc. 88 (1983) 446–448.

[29] M. Bronstein, B. Salvy, Full partial fraction decomposition of rational functions, in: Proc. of ISSAC'1993, ACM Press, 1993, pp. 157–160.

[30] M. Bronstein, M. Petkovšek, On ore rings, linear operators and factorisation, Programm. Comput. Softw. 20 (1994) 14–26.

[31] H. Cartan, Théorie élémentaire des fonctions analytiques d'une on plusieurs variables complexes, Hermann, Paris, 1961.

[32] E. Cattani, A. Dickenstein, B. Sturmfels, Computing multidimensional residues, in: Algorithms in Algebraic Geometry and Applications, Satander, 1994, in: Progr. Math., vol. 143, Birkhäuser, Basel, 1996, pp. 135–164.

[33] E. Cattani, A. Dickenstein, A global view of residues in the torus, J. Pure Appl. Algebra 117/118 (1997) 119–144.

[34] E. Cattani, A. Dickenstein, B. Sturmfels, Residues and resultants, J. Math. Sci. Univ. Tokyo 5 (1998) 119–148.

[35] A. Cayley, Note sur une formule pour la réversion des séries, J. Reine Angew. Math. 52 (1856) 276–284.

[36] B. Chalmers, Fonctions de plusieurs variables complexes. Sur le comportement, à la frontière, de la fonction-noyau de Bergman, C. R. Acad. Sci. Paris Sér. A–B 266 (1968) 1132–1134.

[37] L. Chottin, Enumeration d'arbres et formules d'inversion de séries formelles, J. Combin. Theory Ser. B 31 (1981) 23–45.

[38] F. Chyzak, B. Salvy, Non-commutative elimination in Ore algebras proves multivariate identities, J. Symbolic Comput. 26 (1998) 187–227.

[39] F. Chyzak, An extension of Zeilberger's fast algorithm to general holonomic functions, Discrete Math. 217 (2000) 115–134.

[40] P.C. Consul, Relation of modified power series distributions to Lagrangian probability distributions, Comm. Statist. Theory Methods 10 (1981) 2039–2046.

[41] P.C. Consul, L.R. Shenton, Use of Lagrange expansion for generating discrete generalized probability distributions, SIAM J. Appl. Math. 23 (1972) 239–248.

[42] P.C. Consul, L.R. Shenton, Some interesting properties of Lagrangian distributions, Comm. Statist. 2 (1973) 263–272.

[43] P.C. Consul, F. Famoye, Lagrangian Probability Distributions, Birkhäuser Boston, Boston, MA, 2006.

[44] C. Corsani, D. Merlini, R. Sprugnoli, Left-inversion of combinatorial sums, Discrete Math. 180 (1998) 107–122.

[45] M. Davis, Applied Nonstandard Analysis, Mir, Moscow, 1980 (in Russian).

[46] M.-P. Delest, G. Viennot, Algebraic languages and polynomial enumeration, Theoret. Comput. Sci. 34 (1984) 169–205.

[47] E. Deutsch, L. Shapiro, Exponential Riordan arrays, Lecture notes, Nankai Univ., 2004.

[48] R.B. Dingle, Asymptotic Expansions: Their Derivation and Interpretation, Academic Press, London, 1973.

[49] P. Doubilet, G.-C. Rota, R. Stanley, On the foundations of combinatorial theory. VI: The idea of a generating function, in: Proc. Sixth Berkeley Sympos. on Math. Statist. and Probab. (1970/71): Probability Theory, Vol. II, Univ. California Press, Berkeley, CA, 1972, pp. 267–318.

[50] M. Drmota, A bivariate asymptotic expansions of coefficients of powers of generating functions, European J. Combin. 15 (1994) 139–152.

[51] A.M. Efros, M.A. Danilevski, Operational Calculus and Contour Integrals, GNTIU, Kharkov, 1937 (in Russian).

[52] G.P. Egorychev, The ranks of the factors in the lower central series of a free solvable group, Sibirsk. Mat. Zh. 13 (1972) 708–713 (in Russian); English transl. in: Siberian Math. J. 13.

[53] G.P. Egorychev, Inversion of one-dimensional combinatorial relations, in: Some Questions on the Theory of Groups and Rings, Inst. Fiz. Sibirsk. Otdel. Akad. Nauk USSR, Krasnoyarsk, 1973, pp. 110–122 (in Russian).

[54] G.P. Egorychev, Inversion of Combinatorial Relations, in: Combin. Anal., vol. 3, Krasnoyarsk State University, Krasnoyarsk, 1974, pp. 10–14 (in Russian).

[55] G.P. Egorychev, A.P. Yuzhakov, On the determination of generating functions and combinatorial sums with the help of multidimensional residues, Sibirsk. Mat. Zh. 15 (1974) 1049–1060, (in Russian); English transl. in: Siberian Math. J. 15.

[56] G.P. Egorychev, The method of coefficients (multidimensional case), in: Combin. Asymptot. Anal., vol. 1, Krasnoyarsk State Univ., Krasnoyarsk, 1975, pp. 151–161 (in Russian).

[57] G.P. Egorychev, Computation of Szegő and Bergman kernels for certain n-circular domains with the help of multidimensional residues, Izv. Vysh. Uchebn. Zaved. Mat. 152 (1975) 101–103 (in Russian); English transl. in: Soviet Math. J. 19.

[58] G.P. Egorychev, Computation of certain combinatorial sums with linear constraints on the summation indices, Sibirsk. Mat. Zh. 16 (1975) 856–862 (in Russian); English transl. in: Siberian Math. J. 16.

[59] G.P. Egorychev, Integral Representation and the Computation of Combinatorial Sums, Novosibirsk, Nauka, 1977 (in Russian); English transl. in: Transl. Math. Monogr., vol. 59, Amer. Math. Soc., 1984, 2nd ed., 1989.

[60] G.P. Egorychev, Polynomial identity for the permanent, Mat. Zametki 26 (1979) 961–964 (in Russian).

[61] G.P. Egorychev, New formulas for the permanent, Dokl. Akad. Nauk USSR 254 (1980) 784–787 (in Russian).

[62] G.P. Egorychev, V.M. Levchuk, Enumerative problems for Lie type groups and algebras, Dokl. Akad. Nauk 330 (1993) 464–467 (in Russian).

[63] G.P. Egorychev, V.M. Levchuk, Enumeration of characteristic subgroups of unipotent Lie-type groups, in: Algebra, Proc. of the 3rd Internat. Conf. on Algebra, Krasnoyarsk, August 23–28, 1993, de Gruyter, Berlin, 1996, pp. 49–62.

[64] G.P. Egorychev, Algorithms of integral representation of combinatorial sums and their applications. Formal power series and algebraic combinatorics, in: Proc. 12th International Conference FPSAC'2000, Moscow, Russia, June 2000, 2000, pp. 15–29.

[65] G.P. Egorychev, V.M. Levchuk, Enumeration in the Chevalley algebras, SIGSAM Bull. 35 (2001) 20–34.

[66] G.P. Egorychev, Integral representation and combinatorial identities, Technical report of SCG, University of Waterloo, Waterloo, Canada, August 2001.

[67] G.P. Egorychev, E.V. Zima, On integral representation and algorithmic approaches to the evaluation of combinatorial sums, Technical report CS-2002-02, School of Computer Science, University of Waterloo, Waterloo, Canada, January 2002.

[68] G.P. Egorychev, Solution of the Margenstein–Matiyasevich's question in $3x + 1$ problem. Krasnoyarsk State Technical Univ., Krasnoyarsk, 2004, Preprint, ISBN 5-7636-0632-9 (in Russian).

[69] G.P. Egorychev, E.V. Zima, The characteristic function in $3x + 1$ problem, in: Issues of International School-Seminar Synthesis and Complexity of Management Systems, Math. Inst. of Siberian Branch of Russian Acad. Sci., Novosibirsk, 2004, pp. 34–40 (in Russian).

[70] G.P. Egorychev, E.V. Zima, Decomposition and group theoretic characterization of pairs of inverse relations of Riordan type, Acta Appl. Math. 85 (2005) 93–109.

[71] G.P. Egorychev, F. Kuzucuoglu, V.M. Levchuk, Some enumerative questions for algebras and modules over rings with strongly maximal ideals, 2005, in press.

[72] G.P. Egorychev, V.A. Stepanenko, Combinatorial identity on the Jacobian conjecture, Acta Appl. Math. 85 (2005) 111–120.

[73] Egorychev G.P., Theoretic-functional model of $3x + 1$ problem (continued), in: Mat. Systems, vol. 4, Krasnoyarsk State Agriculture Univ., Krasnoyarsk, 2005, pp. 18–24 (in Russian).

[74] G.P. Egorychev, Margenstein–Matiyasevich question and the characteristic function of height stopping in Collatz conjecture, 2006, in press.

[75] G.P. Egorychev, E.V. Zima, Simple formulae for the number of quadrics and symmetric forms of modules over local rings, Comm. Algebra (2008), in press.

[76] M.A. Evgrafov, Series and integral representations, in: Sovrem. Probl. Mat., Fund. Naprav., vol. 13, VINITI, Moscow, 1986, pp. 5–92 (in Russian).

[77] A. van den Essen, Polynomial Automorphisms and the Jacobian Conjecture, Progr. Math., vol. 190, Birkhäuser Verlag, Basel, 2000.

[78] M.V. Fedorjuk, Integrals and Series, Nauka, Moscow, 1977 (in Russian).

[79] W. Feller, An Introduction to Probability Theory and Its Applications, vol. 1, 2nd ed., Wiley, New York, 1957.

[80] P. Flajolet, A.M. Odlyzko, Limit distributions for coefficients of iterates of polynomials with applications to combinatorial enumerations, Math. Proc. Cambridge Philos. Soc. 96 (1984) 237–253.

[81] P. Flajolet, B. Salvy, Euler sums and contour integral representations, Experiment. Math. 7 (1998) 15–35.

[82] P. Flajolet, R. Sedgewick, Analytic combinatorics, 2007, in press, http://algo.inria.fr/flajolet/publications/books.html.

[83] D. Gardy, Some results on the asymptotic behaviour of coefficients of large powers of functions, Discrete Math. 139 (1995) 189–217.

[84] J. von zur Gathen, J. Gerhard, Modern Computer Algebra, 2nd ed., Cambridge Univ. Press, Cambridge, 2003.

[85] J. Gerhard, M. Giesbrecht, A. Storjohann, E. Zima, Shiftless decomposition and polynomial-time rational summation, in: Proc. of the 2003 Internat. Symposium on Symbolic and Algebraic Computation, ISSAC'2003, ACM Press, Philadelphia, PA, 2003, pp. 119–126.

[86] I.M. Gessel, A noncommutative generalization and q-analog of the Lagrange inversion formula, Trans. Amer. Math. Soc. 257 (1980) 455–482.

[87] I.M. Gessel, A factorization for formal Laurent series and lattice path enumeration, J. Combin. Theory Ser. A 28 (1980) 321–337.

[88] I.M. Gessel, D. Stanton, Another family of q-Lagrange inversion formulas, Rocky Mountain J. Math. 16 (1986) 373–384.

[89] I.M. Gessel, B.E. Sagan, Y.N. Yeh, Enumeration of trees by inversions, J. Graph Theory 19 (1995) 435–459.

[90] I.J. Good, Generalizations to several variables of Lagrange's expansion, with applications to stochastic processes, Proc. Cambridge Philos. Soc. 56 (1960) 367–380.

[91] I.J. Good, The generalization of Lagrange's expansion and the enumeration of trees, Proc. Cambridge Philos Soc. 61 (1965) 499–517; correction in: Proc. Cambridge Philos Soc. 64 (1968) 160.

[92] Yu.M. Gorchakov, G.P. Egorychev, The ranks of the factors in the lower central series of a free polynilpotent group, Dokl. Akad. Nauk USSR 204 (1972) 12–14 (in Russian).

[93] R.W. Gosper, Decision procedure for indefinite hypergeometric summation, Proc. Natl. Acad. Sci. USA 75 (1977) 40–42.

[94] H.W. Gould, Combinatorial identities. A standardized set of tables listing 500 binomial coefficient summations, Morgantown, WV, 1972.

[95] H.W. Gould, L.C. Hsu, Some new inverse series relations, Duke Math. J. 40 (1973) 885–891.

[96] I.P. Goulden, D.M. Jackson, Combinatorial Enumeration, Wiley, New York, 1983.

[97] I.P. Goulden, D.M. Kulkarni, Multivariate Lagrange inversion, Gessel–Viennot cancellation and the matrix tree theorem, J. Combin. Theory Ser. A 80 (1997) 295–308.

[98] R.L. Graham, D.E. Knuth, O. Patashnik, Concrete Mathematics. A Foundation for Computer Science, Addison–Wesley, Reading, MA, 1989.

[99] D.H. Green, D.E. Knuth, Mathematics for the Analysis of Algorithms, Birkhäuser Boston, Boston, MA, 1982.

[100] M. Hall Jr., The Theory of Groups, Macmillan, New York, 1959.

[101] P.R. Halmos, Measure Theory, Van Nostrand, Princeton, NJ, 1950.

[102] G.H. Hardy, Divergent Series, Clarendon Press, Oxford, 1949.

[103] P. Henrici, Applied and Computational Complex Analysis, Wiley, New York, 1991.

[104] M. van Hoeij, Rational solutions of linear difference equations, in: Proc. of ISSAC'1998, ACM Press, 1998, pp. 120–123.

[105] M. van Hoeij, Factoring polynomials and the knapsack problem, J. Number Theory 95 (2002) 167–189.

[106] M. van Hoeij, A conjecture in the problem of rational definite summation, abstract, arXiv.org. math.CO/0210158, 2002.

[107] I.-Ch. Huang, Reversion of power series by residues, Comm. Algebra 26 (1998) 803–812.

[108] I.-Ch. Huang, Inverse relations and Schauder bases, J. Combin. Theory Ser. A 97 (2002) 203–224.

[109] A. Hurvitz, R. Courant, Vorlesungen über allegemeine Funktionentheorie und elliptische Funktionen, 3rd ed., Springer-Verlag, Berlin, 1929.

[110] I.I. Ibragimov, Interpolation Methods and Some Applications, Nauka, Moscow, 1971 (in Russian).

[111] E. Jabotinski, Analytic iterations, Trans. Amer. Math. Soc. 108 (1963) 457–477.

[112] W.B. Jones, W.J. Thron, Continued Fractions. Analytic Theory and Applications, Mir, Moscow, 1985 (in Russian).

[113] S.A. Joni, Lagrange inversion in higher dimensions and umbral operators, Linear Multilinear Algebra 6 (1978) 111–121.

[114] A. Joval, Une théorie combinatoire des séries formelles, Adv. Math. 42 (1981) 1–82.

[115] J. Kaucký, Combinatorial Identities, Veda, Bratislava, 1975 (in Slovak).

[116] J.G. Kemeni, Matrix representation for combinatorics, J. Combin. Theory Ser. A 36 (1984) 279–306.

[117] O.H. Keller, Ganze Cremona-Transformationen, Monatsh. Math. Phys. 47 (1939) 299–306.

[118] D.E. Knuth, The Art of Computer Programming, vol. I, Addison–Wesley, New York, 1968.

[119] W. Koepf, Power series in computer algebra, J. Symbolic Comput. 13 (1992) 581–603.

[120] Kourovka Notebook, Unsolved Problems in the Theory of Groups, 2nd ed., Inst. Mat. Sibirsk. Otdel. Akad. Nauk USSR, Novosibirsk, 1967 (in Russian).

[121] Ch. Krattenthaler, A new q-Lagrange formula and some applications, Proc. Amer. Math. Soc. 90 (1984) 338–344.

[122] Ch. Krattenthaler, Operator methods and Lagrange inversion: A unified approach to Lagrange formulas, Trans. Amer. Math. Soc. 305 (1988) 431–465.

[123] Ch. Krattenthaler, A new matrix inverse, Proc. Amer. Math. Soc. 124 (1996) 47–59.

[124] Ch. Krattenthaler, M. Schlosser, A new matrix inverse with applications to multiple q-series, Discrete Math. 204 (1999) 249–279.

[125] V.A. Kudrjavtzev, Summation of Power Numbers of the Positive Integers and Bernoulli Numbers, L. NKTP, Moscow, 1936 (in Russian).

[126] F. Kuzucuoğlu, V.M. Levchuk, Ideals of some matrix rings, Comm. Algebra 28 (2000) 3503–3513.

[127] G. Labelle, Une nouvelle démonstration combinatoire des formules d'inversion de Lagrange, Adv. Math. 42 (1981) 217–247.

[128] J.C. Lafon, Summation in finite terms, in: B. Buchberger, G. Collins, R. Loos (Eds.), Computer Algebra, Springer-Verlag, Berlin, 1983, pp. 71–77.

[129] J.E. Lagarias, The $3x + 1$ problem and its generalizations, in: J. Borwein, et al. (Eds.), Organic Mathematics, Proc. Workshop Simon Fraser Univ., Barnaby, Canada, December 12–14, 1995, Amer. Math. Soc., Providence, RI, 1997, pp. 305–334.

[130] J.W. Layman, C.L. Prather, Generalized Bell numbers and zeros of successive derivatives of an entire function, J. Math. Anal. Appl. 96 (1983) 42–51.

[131] A.F. Leont'ev, The Sequences of Exponential Polynomials, Nauka, Moscow, 1980 (in Russian).

[132] V.M. Levchuk, Symmetric forms and quadrics of projective spaces over local rings, in: Proc. of the Conference Antalya Algebra Days, Bilgi Univ., Istanbul, 2004, p. 34.

[133] E. Lucas, Characteristic Functions, Nauka, Moscow, 1979 (in Russian).

[134] V.N. Lyamin, B.I. Selivanov, Enumeration problems for simple hypergraphs, in: Combinatorial Analysis, vol. 3, Krasnoyarsk State Univ., Krasnoyarsk, 1974, pp. 10–14 (in Russian).

[135] P.A. MacMahon, Combinatory Analysis, vol. I, Cambridge Univ. Press, 1915;
P.A. MacMahon, Combinatory Analysis, vol. II, Cambridge Univ. Press, 1916.

[136] Y.-K. Man, On computing closed forms for indefinite summation, J. Symbolic Comput. 16 (1993) 355–376.

[137] Y.-K. Man, F.J. Wright, Fast polynomial dispersion computation and its application to indefinite summation, in: Proc. of ISSAC'1994, ACM Press, 1994, pp. 175–180.

[138] S. Mandelbrojt, Dirichlet Series. Principles and Methods, Mir, Moscow, 1973 (in Russian).

[139] T. Mansour, Combinatorial identities and inverse binomial coefficients, Adv. Appl. Math. 28 (2002) 196–202.

[140] M. Margenstern, Y. Matiyasevich, A binomial representation of the $3x + 1$ problem, Acta Arith. 91 (1999) 367–378.

[141] O.I. Marichev, Handbook of Integral Transformation of Higher Transcendental Functions. Theory and Algorithmic Tables, Wiley/Harwood, New York, 1984.

[142] L.F. Matusevich, Rational summation of rational functions, Beiträge Algebra Geometry 41 (2000) 531–536.

[143] E.B. McBride, Obtaining Generating Functions, Springer-Verlag, Berlin, 1971.

[144] A. Meir, J.W. Moon, The asymptotic behaviour of coefficients of powers of certain generating functions, European J. Combin. 11 (1990) 581–587.

[145] D. Merlini, D.G. Rogers, R. Sprugnoli, M.C. Verri, On some alternative characterizations of Riordan groups, Canad. J. Math. 49 (1997) 301–320.

[146] D. Merlini, M.C. Verri, Generating trees and proper Riordan arrays, Discrete Math. 218 (2000) 167–183.

[147] D. Merlini, R. Sprugnoli, M.C. Verri, The method of coefficients, Amer. Math. Monthly 114 (2007) 40–57.

[148] St.C. Milne, G. Bhatnagar, A characterization of inverse relations, Discrete Math. 193 (1998) 235–245.

[149] R. Milson, Spectral residues of second-order differential equations: a new method for summation identities and inversion formulas, Stud. Appl. Math. 107 (2001) 337–366.

[150] H. Minc, Permanents, Encyclopedia Math. Appl., vol. 6, Addison-Wesley, Reading, MA, 1978.

[151] M. Mishima, M. Jimbo, A series of identities for the coefficients of inverse matrices on a Hamming scheme, Discrete Math. 156 (1996) 285–290.

[152] D.S. Mitrinović, J.D. Kekrić, Cauchy Method of Residues. Theory and Applications, Vols. I, II, Math. Appl., vol. 259, Kluwer Academic Press, Dordrecth, 1984.

[153] M. Monagan, K. Geddes, K. Heal, G. Labahn, S. Vorkoetter, J. McCarron, P. DeMarco, Maple 8 Introductory Programming Guide, Waterloo Maple, Waterloo, 2002.

[154] M. Nahay, Linear Differential Resolvents, Dissertation Rutgers State Univ., New Brunswick, NJ, May 2000, pp. 157–160.

[155] E. Netto, Lehrbuch der Combinatorik, 2nd ed., Teubner, Leipzig, 1927, reprint Chelsea, New York, 1958.

[156] A.M. Odlyzko, Asymptotic enumeration methods, in: Handbook of Combinatorics, vols. 1, 2, Elsevier, Amsterdam, 1995, pp. 1063–1229.

[157] The Otto Dunkel Memorial Problem Book, New York, 1957; Russian transl.: Selected Problems from "American Mathematical Monthly", Mir, Moscow.

[158] P.S. Pankov, Demonstrative Computations on Computers, ILIM, Frunze, 1978 (in Russian).

[159] P. Paule, Computer Algebra Algorithmen für q-Reihen und kombinatorische Identitaten, RISC Linz, No. 90-02.0, 1990, 1–25.

[160] P. Paule, Greatest factorial factorization and symbolic summation, J. Symbolic Comput. 20 (1995) 235–268.

[161] P. Peart, W.-J. Woan, A divisibility property for a subgroup of Riordan matrices, Discrete Appl. Math. 98 (2000) 255–263.

[162] R. Pemantle, M.C. Wilson, Asymptotics of multivariate sequences. I. Smooth points of the singular variety, J. Combin. Theory Ser. A 97 (2002) 129–161.

[163] J.K. Percus, A note on extension of the Lagrange inversion formula, Comm. Pure Appl. Math. 17 (1964) 137–146.

[164] J.K. Percus, Combinatorial Methods, Springer-Verlag, Berlin, 1971.

[165] M. Petkovšek, Hypergeometric solutions of linear recurrences with polynomial coefficients, J. Symbolic Comput. 14 (1992) 243–264.

[166] M. Petkovšek, H.S. Wilf, D. Zeilberger, $A = B$, A K Peters, Wellesley, MA, 1996.

[167] V.M. Petrogradsky, Growth of finitely generated polynilpotent Lie algebras and groups, generalized partitions, and functions analytic in the unit circle, Internat. J. Algebra Comput. 9 (1999) 179–212.

[168] R. Pirastu, On combinatorial identities: symbolic summation and umbral calculus, PhD thesis, RISC Linz, July 1996.

[169] M.L. Platonov, Combinatorial Numbers of a Class of Mapping and Their Applications, Nauka, Moscow, 1979 (in Russian).

[170] G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, Acta Math. 68 (1937) 145–254.

[171] A.P. Prudnikov, Yu.A. Brychkov, O.M. Marichev, Integrals and Rings. Special Functions, vol. 1, Wiley, New York, 1988.

[172] G.N. Pyhteev, The Precise Methods of Cauchy-type Integrals, Nauka, Novosibirsk, 1980 (in Russian).

[173] G.N. Pyhteev, The Approximate Methods of Calculation of Cauchy-type Integrals, Nauka, Novosibirsk, 1982 (in Russian).

[174] J. Riordan, An Introduction to Combinatorial Analysis, Wiley/Hall, New York, 1958.

[175] J. Riordan, Inverse relations and combinatorial identities, Amer. Math. Monthly 71 (1964) 485–498.

[176] J. Riordan, Combinatorial Identities, Wiley, New York, 1968.

[177] S. Roman, The algebra of formal series, Adv. Math. 31 (1979) 309–329.

[178] S. Roman, The algebra of formal series. II. Sheffer sequences, J. Math. Anal. Appl. 74 (1980) 120–143.

[179] S. Roman, The algebra of formal series. III. Several variables, J. Approx. Theory 26 (1979) 340–381.

[180] S. Roman, The Umbral Calculus, Academic Press, New York, 1984.

[181] K.A. Rybnikov, Introduction to Combinatorial Analysis, Moscow State Univ., Moscow, 1972 (in Russian).

[182] G.-C. Rota, On the foundations of combinatorial theory. I. Theory of Möbius functions, Z. Wahrsch. Verw. Geb. 2 (1964) 340–368.

[183] R.A. Sack, Generalization of Lagrange's expansion for functions of several implicitly defined variables, SIAM J. Appl. Math. 13 (1965) 913–926.

[184] A. Salomaa, M. Soittola, Automata-theoretic Aspects of Formal Power Series, Springer-Verlag, New York, 1978.

[185] B. Salvy, P. Zimmermann, Maple package for the manipulation of generating and holonomic functions in one variable, rapport de récherche No 143, INRIA, 1992, 1–22.

[186] M.P. Schützenberger, Certain elementary families of automata, in: Proc. Sympos. Math. Theory of Automata, Polytech. Inst. of Brooklyn Press, New York, 1962, pp. 139–153.

[187] M.P. Schützenberger, On context-free languages and push-down automata, Information and Control 6 (1963) 246–264.

[188] I.J. Schwatt, An Introduction to Operations with Series, Univ. Pennsylvania Press, Philadelphia, PA, 1924.

[189] B.V. Shabat, An Introduction to the Complex Analysis, Nauka, Moscow, 1969 (in Russian).

[190] L.W. Shapiro, S. Getu, W.-J. Woan, L.C. Woodson, The Riordan group, Discrete Appl. Math. 34 (1991) 229–239.

[191] J. Sheehan, An identity, Amer. Math. Monthly 77 (1970) 168.

[192] A.L. Shmel'kin, Free polynilpotent groups, Izv. Akad. Nauk USSR Ser. Mat. 28 (1964) 91–122 (in Russian).

[193] N.J.A. Sloane, S. Plouffe, The Encyclopedia of Integer Sequences, Academic Press, San Diego, CA, 1995.

[194] V.G. Sokolov, An analogue of the Witt's formula for free solvable groups, Algebra i Logika 8 (1969) 367–372 (in Russian).

[195] R. Sprugnoli, Riordan arrays and combinatorial sums, Discrete Math. 132 (1994) 267–290.

[196] R. Sprugnoli, Riordan arrays and the Abel–Gould identity, Discrete Math. 142 (1995) 213–233.

[197] R.P. Stanley, Enumerative Combinatorics, vol. 2, Cambridge Univ. Press, Cambridge, 1997.

[198] O.A. Starikova, Enumeration of the quadrics and the symmetric forms of modules over local rings, PhD thesis, Krasnoyarsk State Technical Univ., Krasnoyarsk, 2004.

[199] V.A. Stepanenko, The inversion formula of Cayley–Sylvester–Sack and the Jacobian conjecture, in: Proc. of the Internat. Conference Algebra and Its Applications, Krasnoyarsk State University, Krasnoyarsk, 2002, pp. 114–115 (in Russian).

[200] V.A. Stepanenko, On the solution of a system of n algebraic equations with n variables with the help of hypergeometric functions, Vestnik KrasGU, 2005, pp. 35–48 (in Russian).

[201] I.E. Strazdin, On the theoretic-group method of obtaining of combinatorial numbers, in: Voprosy Kibernetiki 16, Trudy II Vses. Seminara po Kombinatornoi Matematike 1, 1975, pp. 103–110 (in Russian).

[202] J.J. Sylvester, On the change of system of indefinite variables, Q. J. Pure Appl. Math. 1 (1857) 42–56 and 126-134.

[203] A. Tefera, MultInt, a MAPLE package for multiple integration by the WZ method, J. Symbolic Comput. 34 (2002) 329–353.

[204] A.K. Tsikh, Multidimensional Residues and Their Applications, Amer. Math. Soc., Providence, RI, 1992.

[205] W.T. Tutte, On elementary calculus and the Good formula, J. Combin. Theory Ser. B 18 (1975) 97–137.

[206] V.A. Ufnarovsky, Combinatorial and asymptotic methods in algebra, in: Sovr. Probl. Mat. Fund. Naprav., vol. 57, VINITI, 1990, pp. 5–178 (in Russian).

[207] N.Ya. Vilenkin, Combinatorics, Academic Press, New York, 1971.

[208] O.V. Viskov, Inversion of power series and Lagrange formulas, Soviet. Math. Dokl. 22 (1980) 330–342.

[209] H.S. Wilf, The "Snake-Oil"-method for proving combinatorial identities, in: Surveys in Combinatorics, in: London Math. Soc. Lecture Note Ser., vol. 141, Cambridge Univ. Press, Cambridge, 1989, pp. 208–217.

[210] H.S. Wilf, Generatingfunctionology, 2nd ed, Academic Press, Boston, MA, 1994.

[211] M.C. Wilson, Asymptotics for generalized Riordan arrays, in: 2005 Internat. Conf. on Analysis of Algorithms, in: Discrete Math. Theor. Comput. Sci. Proc., AD, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2005, pp. 323–333 (electronic).

[212] G.J. Wirsching, The Dynamic System Generated by the $3n+1$ Function, Lecture Notes in Math., vol. 1681, Springer-Verlag, Berlin, 1998.

[213] K.-W. Yu, On the value distribution of $\varphi(z)[f(z)]^{n-1}f^{(k)}(z)$, J. Inequal. Pure Appl. Math. 3 (1) (2002) 1–5, article 8.

[214] A.P. Yuzhakov, Principles of the Theory of Multidimensional Residues, Krasnoyarsk State Univ., Krasnoyarsk, 1975 (in Russian).

[215] D. Zeilberger, The method of creative telescoping, J. Symbolic Comput. 11 (1991) 195–204.

[216] B.S. Zinov'ev, On reproducing kernels for multiply circular domains of holomorphy, Sibirsk. Mat. Zh. 15 (1974) 35–48 (in Russian).

This page intentionally left blank

Subject Index

This page intentionally left blank