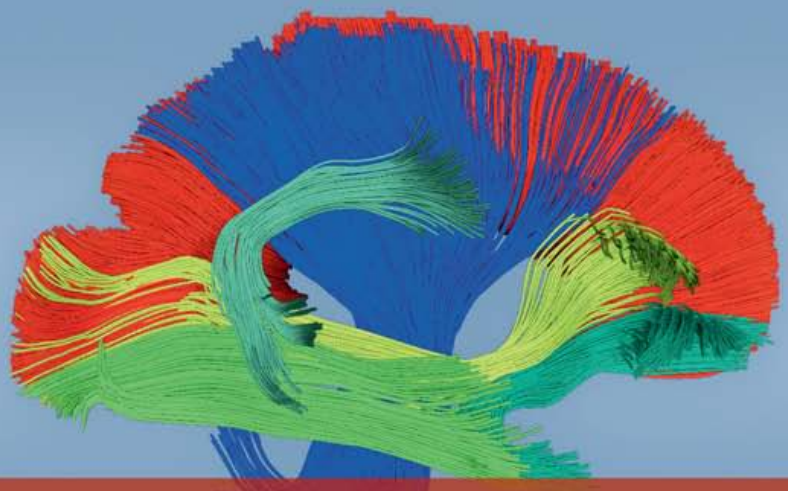


MATHEMATICS AND VISUALIZATION

David H. Laidlaw
Joachim Weickert Editors

Visualization and Processing of Tensor Fields

Advances and Perspectives

Springer

Mathematics and Visualization

Series Editors

David Laidlaw
Joachim Weickert
Editors

Visualization and Processing of Tensor Fields

Advances and Perspectives

With 131 Figures, 92 in Color and 13 Tables

Springer

David Laidlaw
Brown University
Department of Computer Science
Providence RI 02912
Box 1910
USA
dhl@cs.brown.edu

Joachim Weickert
Saarland University
Faculty of Mathematics and Computer Science
Mathematical Image Analysis Group
66041 Saarbrücken
Germany
weickert@mia.uni-saarland.de

ISBN: 978-3-540-88377-7 e-ISBN: 978-3-540-88378-4

Mathematics and Visualization ISSN 1612-3786

Library of Congress Control Number: 2008936143

Mathematics Subject Classification (2000): 68-XX, 70-XX, 15-XX, 76-XX, 92-XX

Cover design: deblik, Berlin

Printed on acid-free paper

springer.com

We dedicate this book to our families.

Preface

This book attempts to capture some of the excitement of an inspiring Dagstuhl Seminar in January 2007. The authors report on recent research results as well as opining on future directions for the analysis and visualization of tensor fields. Topics range from applications of the analysis of tensor fields to purer research into their mathematical and analytical properties. One of the goals of this seminar was to bring together researchers from along that pure-to-applied disciplinary axis with the hope of fostering new collaborations and research. This book, we hope, will continue to further that goal in a broader context.

Providence, Rhode Island, USA — *David H. Laidlaw*
Saarbrücken, Saarland, Germany — *Joachim Weickert*
August 2008

Contents

List of Contributors

Daniel C. Alexander
Centre for Medical Image Computing
Dept. Computer Science
UCL Gower Street
London
WC1E 6BT
UK
D.Alexander@cs.ucl.ac.uk

Alfred Anwander
Department of Cortical Networks
and Cognitive Functions
Max Planck Institute for Human
Cognitive and Brain Sciences
Stephanstraße 1A
04103 Leipzig Germany
anwander@cbs.mpg.de

Philip Batchelor
Division of Imaging Sciences
King's College London
The Rayne Institute 4th floor
Lambeth Wing, St Thomas Hospital
London SE1 7EH
UK
Philip.batchleor@kcl.ac.uk

Anders Brun
Centrum for Bildanalys, Box 337
751 05 Uppsala
Sweden
anders@cb.uu.se

Bernhard Burgeth
Faculty of Mathematics
and Computer Science
Saarland University, Bld. E1 1 66041
Saarbrücken, Germany
burgeth@mia.uni-saarland.de

Andrew C. Callan-Jones
Physicochimie Curie
(CNRS-UMR168)
Institut Curi, Section de Recherche
26 rue dUlm 75248, Paris
Cedex 05 France
Andrew.Callan_Jones@curie.fr

Owen T. Carmichael
University of California
Davis Center for Neuroscience
1544 Newton Court, Davis
CA 95618-4859
ocarmichael@udavis.edu

Oddvar Christiansen
Department of Mathematics
University of Bergen, Johannes
Brunsgate 12, 5008 Bergen, Oddvar
Christiansen@math.uib.no

Stephan Didas
Faculty of Mathematics and
Computer Science, Saarland
University, Building E1 1 (former)
36.1, 66041 Saarbrücken Germany
didas@mia.uni-saarland.de

Louis Feng
Institute for Data Analysis
and Visualization
Department of Computer Science
University of California, Davis
CA 95616-8562, USA
zfeng@ucdavis.edu

James C. Gee
Penn Image Computing and Science
Laboratory (PICSL),
Department of Radiology
University of Pennsylvania School
of Medicine, 3600 Market Street
Suite 370 Philadelphia
PA 19104-2644, USA
gee@mail.med.upenn.edu

Yaniv Gur
University of Tel-Aviv
School of Mathematical Sciences
Tel Aviv University, Ramat Aviv
Tel Aviv 69978, Israel
yanivg@post.tau.ac.il

Bernard Hamann
Institute for Data Analysis
and Visualization
Department of Computer Science
University of California, Davis
CA 95616-8562, USA
hamann@cs.ucdavis.edu

Mario Hlawitschka
Institute for Data Analysis
and Visualization
Department of Computer Science
University of California, Davis
CA 95616-8562, USA
hlawitschka@ucdavis.edu

Ingrid Hotz
Konrad-Zuse-Zentrum Fuer
Informationstechnik Berlin (ZIB)
Division Scientific Computing
Department Visualization and Data
Analysis, Takustraße 7
D-14195 Berlin-Dahlem, Germany
hotz@zib.de

T.J. Jankun-Kelly
Department of Computer Science
and Engineering, Mississippi State
University, Mississippi State
MS 39762, 300 Butler Hall, Corner
of Bar and George Parry Streets
Mississippi State, MS 39762, USA
tjk@acm.org

Bing Jian
Room no. E331, CISE Building
Department of Computer
and Information Science
and Engineering, University
of Florida Gainesville, FL 32611
USA
bjian@cise.ufl.edu

Kenneth Joy
Institute for Data Analysis
and Visualization
Department of Computer Science
University of California, Davis
CA 95616-8562, USA
kijoy@ucdavis.edu

Hans Knutsson
Linköping University, Department
of Biomedical Engineering
University Hospital, 581 85
Linköping, Sweden
knutte@imt.liu.se

David H. Laidlaw
Department of Computer Science
Brown University, 115 Waterman
Street, Providence, RI 02912, USA
dhl@cs.brown.edu

Johan Lie
Centre for Integrated Petroleum Research
Unifob/University of Bergen
Realfagsbyggat, Allegaten 41
5007 Bergen
Norway
Johan.lie@cipr.uib.no

Maher Moakher
Laboratory for Mathematical and Numerical Modeling in Engineering Science, National Engineering School at Tunis, ENIT-LAMSIN
B.P. 37, Tunis-Belvedere, Tunisia
maher.moakher@gmail.com

Evren Özarslan
National Institutes of Health, Bldg 13, Rm 3W16, MSC 5772, 13 South Drive Bethesda, MD 20892, USA
evren@helix.nih.gov

T.H.J.M. Peeters
Department of Biomedical Engineering, Eindhoven University of Technology, WH 2.103, 5600 MB Eindhoven, The Netherlands
T.H.J.M.peeters@tue.nl

Robert A. Pelcovits
Department of Physics, Brown University, 184 Hope Street
Providence, RI 02912, USA
pelcovits@physics.brown.edu

Björn Popilka
Department of Mathematics and Computer Science
University of Mannheim
68131 Mannheim, Germany
bpopilka@rumms.uni-mannheim.de

P.R. Rodrigues
Department of Biomedical Engineering, Eindhoven University of Technology, WH 2.103, 5600 MB Eindhoven, The Netherlands
P.R.Rodrigues@tue.nl

M.A. Rodriguez-Florido
Centre for Technology in Medicine, University of Las Palmas de Gran Canaria, Spain
marf@itccanarias.org

Juan Ruiz-Alzola
Centre for Technology in Medicine
University of Las Palmas de Gran Canaria, Spain
jruiz@itccanarias.org

Gerik Scheuermann
Universität Leipzig, Institut für Informatik, PF 100920, 04009 Leipzig, Germany
scheuermann@informatik.uni-leipzig.de

Michael Schreiner
University of Applied Science Buchs NTB, Laboratory for Industrial Mathematics, Werdenbergstrasse 4
CH-9471 Buchs, Switzerland
schreiner@ntb.ch

Thomas Schultz
MPI Informatik, Campus E1 4 66123 Saarbrücken, Germany
schultz@mpi-inf.mpg.de

Hans-Peter Seidel
MPI Informatik, Campus E1 4 66123 Saarbrücken, Germany
hpseidel@mpi-inf.mpg.de

Simon Setzer
Department of Mathematics and Computer Science
University of Mannheim
68131 Mannheim, Germany
ssetzer@kiwi.math.uni-mannheim.de

Tony J. Simon
Department of Psychiatry and Behavioral Sciences, UC Davis M.I.N.D. Institute, 2825 50th St. Sacramento, CA 95817, USA
tjsimon@ucdavis.edu

Vadim Slavin
Advanced Technology Researcher Lockheed Martin Space Systems San Francisco Bay Area San Francisco USA
vslavin@gmail.com

Nir Sochen
Department of Applied Mathematics Tel-Aviv University, Ramat-Aviv Tel-Aviv 69978, Isreal
sochen@post.tau.ac.il

Gerald Sommer
Christian-Albrechts-Universität zu Kiel, Institut für Informatik Ohlshausenstraße 40 24098 Kiel, Germany
gs@ks.informatik.uni-kiel.de

Dario Sosa-Cabrera
Center for Technology in Medicine University of Las Palmas de Gran Canaria, Spain
dario@ctm.ulpgc.es

Gabriele Steidl
Department of Mathematics and Computer Science University of Mannheim 68131 Mannheim Germany
steidl@math.uni-mannheim.de

E. Suarez-Santana
Centre for Technology in Medicine University of Las Palmas de Gran Canaria, Spain
esuarez@itccanarias.org

B.M. ter Haar Romeny
Department of Biomedical Engineering, Eindhoven University of Technology WH 2.103, 5600 MB Eindhoven, The Netherlands
B.M.terHaarRomeny@tue.nl

Baba C. Vemuri
Center for Vision, Graphics and Medical Imaging Department of CISE E324 University of Florida Gainesville, FL 32611-6120 USA
vemuri@cise.ufl.edu

Anna Vilanova
Eindhoven University of Technology Biomedical Image Analysis P.O. Box 513, 5600 MB Eindhoven The Netherlands
A.Vilanova@tue.nl

Gunther H. Weber
Computational Research Division Lawrence Berkeley National Laboratory, Berkeley CA 94720-8139 USA
ghweber@lbl.gov

Joachim Weickert
Mathematical Image Analysis Group, Faculty of Mathematics and CS, Saarland University Building E1-1, 66041 Saarbrücken Germany
weickert@mia.uni-saarland.de

Lennart Wietzke
Christian-Albrechts-Universität zu Kiel, Institut für Informatik Ohlshausenstraße 40 24098 Kiel Germany
lw@ks.informatik.uni-kiel.de

Paul A Yushkevich
Penn Image Computing and Science Laboratory, Department of Radiology
University of Pennsylvania
Philadelphia, PA 19104-4283
USA
pauly2@mail.med.upenn.edu

Di Zang
Christian-Albrechts-Universität zu Kiel, Institut für Informatik
Ohlshausenstraße 40
24098 Kiel, Germany
dz@ks.informatik.uni-kiel.de

Hui Zhang
Penn Image Computing and Science Laboratory, Department of Radiology, University of Pennsylvania, Philadelphia
PA 19104-4283, USA
huiz@seas.upenn.edu

Song Zhang
Department of Computer Science and Engineering, Bagley College of Engineering, Mississippi State University, Mississippi State
MS 39762, USA
szhang@cse.msstate.edu

Modelling, Fitting and Sampling in Diffusion MRI

Daniel C. Alexander

Centre for Medical Image Computing, Dept. Computer Science, UCL,
Gower Street, London, WC1E 6BT, UK
D.Alexander@cs.ucl.ac.uk

Summary. This chapter discusses issues of modelling the diffusion MRI signal from brain tissue, fitting models of tissue microstructure to diffusion MRI measurements and designing acquisition schemes that provide the best estimates of model parameters. We construct a simple geometric model of white-matter tissue and derive an expression that relates the model parameters to the diffusion MRI signal. The axon density and diameter are parameters of the model and we examine the accuracy and precision with which we can estimate these potentially important biomarkers. Precision and accuracy depend on the set of measurements we acquire and the method we use to fit the model parameters to the data. We investigate various strategies to optimize the experiment design, as well as various objective functions for model fitting. Experiments and results compare the different methods and provide insight into the accuracy with which we can measure axon density and diameters.

1 Introduction

Diffusion MRI measures the displacement of particles, usually water molecules, within a material over a time interval. The material microstructure controls the scatter pattern of particles within and, conversely, measurements of the particle displacement reveal information about the microstructure. The current standard diffusion MRI technique is diffusion tensor (DT) MRI [1], which provides two unique insights into material microstructure. First, DT-MRI provides quantitative measurements of the anisotropy of particle displacements and, second, it provides an estimate of the dominant orientation of particle displacements. In fibrous material, such as white matter in the brain, the dominant orientations of particle displacements reflect the dominant fibre directions. Diffusion MRI is particularly useful for brain imaging, because it reveals the orientations of white-matter fibres in each voxel of an image volume. Tractography algorithms follow these fibre-orientation estimates from point to point to determine the connectivity of the whole brain.

Diffusion tensor MRI assumes a Gaussian model of particle displacements. This simple model limits the technique's applicability. The most

well-documented departure from the Gaussian model occurs at white-matter fibre crossings in brain tissue [2–4]. This observation has motivated extensive recent work on multiple-fibre reconstruction algorithms, such as multi-tensor models [5, 6], Qball Imaging [7], PASMRI [8] and spherical deconvolution [9–11]; see [12] for a review. These techniques replace the simple Gaussian model and resolve the orientations of multiple fibre populations within single voxels, which improves the precision and accuracy of tractography algorithms [13].

Diffusion MRI reveals microstructure changes, which highlight diseased tissue (e.g. [14, 15]) and allows monitoring of development [16] and aging [17]. Connectivity mapping has provided fundamental insights in basic neuroscience [18, 19], reveals differences in disease [20] and assists prognoses after brain injury, stroke [21] or surgery [22]. The past decade has seen major research efforts develop sophisticated acquisition and processing methods for diffusion MRI, but clinical application of the technique is still relatively limited. Most clinical applications still use only simple biomarkers, such as the mean diffusivity (MD) and fractional anisotropy (FA) of the diffusion tensor. Applications of tractography and connectivity mapping are beginning to appear, in particular within neurosurgery, but are far from widespread or routine. Simple biomarkers such as MD and FA are useful as indicators of major microstructural changes, such as brain damage through stroke [23], but do not capture more subtle effects that might be earlier indicators of degenerative diseases, such as multiple sclerosis or dementia. Increases in available MRI scanner power and improved acquisition sequences now potentially allow new techniques to image more specific features of the microstructure, such as axon density, size and permeability, which also affect the diffusion MRI signal. These more direct features of the tissue microstructure are exciting prospects as biomarkers of the future.

The next section gives some background on diffusion MRI and how we use it to infer tissue microstructure. Section 3 introduces the simple model of white matter tissue and derives the model for the diffusion MRI signal. It then introduces various methods for model fitting and experiment design with reference to that simple model. Section 4 tests and compares the various techniques using the simple model and Sect. 5 concludes.

2 Background

This section covers some background on diffusion MRI. In particular, the following subsections outline the basic pulse sequence for acquiring diffusion MRI measurements and then discuss modelling techniques that relate the diffusion MRI signal to features of material microstructure.

2.1 The Measurement

The standard pulse sequence for diffusion MRI is a pulsed-gradient spin-echo (PGSE) sequence, which places equal diffusion gradient pulses on either side

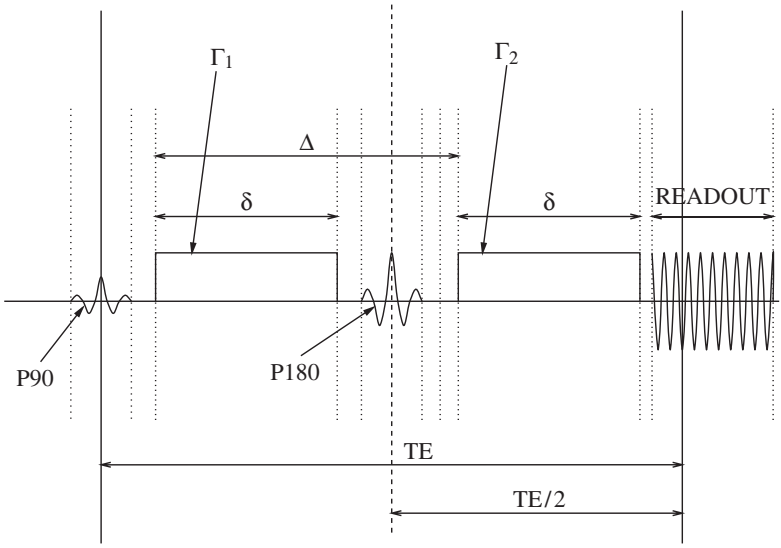


Fig. 1. The pulsed-gradient spin-echo sequence. P90 is the length of the 90° radio-frequency pulse at the start of the sequence and P180 is the length of the 180° refocussing pulse at the centre. TE is the echo time

of a 180° radio-frequency pulse at the centre of the sequence. Figure 1 shows a diagram of the pulse sequence. The length δ, strength and direction $\mathbf{G}$ and separation Δ of the gradient pulses all control the sensitivity of the signal to particle displacements. These are the three main tunable parameters of the sequence. A variety of alternative pulse sequences provide similar measurements, such as twice-refocused spin-echo [24], which reduces eddy-current distortions, STEAM [25] and steady-state [26], which allow long diffusion times and imaging of short-T_2 tissue; all have costs in signal-to-noise compared to PGSE.

If $\delta \ll \Delta$, the normalized signal $A(\mathbf{q})$ approximates the Fourier transform of the distribution of particle displacements (i.e. the scatter pattern) p at wavenumber $\mathbf{q} = \gamma\delta\mathbf{G}$, where γ is the gyromagnetic ratio; see [12, 27] for derivations. This Fourier relationship is the basis of many diffusion MRI methods. For example, in DT-MRI, p is a zero-mean Gaussian with covariance $2\mathbb{D}t$, where $\mathbb{D}$ is the 3×3 symmetric DT and $t = \Delta$ is the diffusion time. The Fourier transform gives $A(\mathbf{q}) = \exp(-t\mathbf{q}^T\mathbb{D}\mathbf{q})$, which is the standard equation for fitting the DT to measurements [1]. In practice, δ is rarely negligible. For simple scattering, such as Gaussian displacements, we can correct for finite δ by setting $t = \Delta - \delta/3$ [27]. However, p in tissue is not Gaussian and finite-δ effects are difficult to model precisely [28].

2.2 Modelling and Reconstruction

The shape of p and its evolution over time are sensitive to other parameters of tissue microstructure such as axon density, diameter distribution and membrane permeability, the density of other cells such as glial cells and the diffusivity and relaxation times (T_1 and T_2) of water within different cellular compartments. Various researchers, for example [29, 30], estimate these parameters using diffusion MRI, but their experiments require high gradient strengths and signal-to-noise ratios to separate the effects. Measurements of this type in live human subjects have broadly been considered impossible, at least with current technology.

Departures of the diffusion MRI signal from the Gaussian model are well documented in the literature, for example [30]; specifically, the signal does not decay 'mono-exponentially' as $|\mathbf{q}|$ and/or t increase. The departure is no surprise. In white matter, for example, water molecules within axon cells exhibit restricted diffusion, since the axon's myelin sheath is largely impermeable to water over typical diffusion times. Other water molecules (around 20%) are in the extracellular space and exhibit hindered diffusion. The extracellular space is normally mostly connected so that water percolates and p is approximately Gaussian [31]. Restricted intra-cellular diffusion, however, produces highly non-Gaussian p, since cell dimensions limit the maximum possible displacement. The literature contains analytic models for p within simple restricting geometries, such as spheres, cylinders and parallel planes [32–34]. These models relate simply to the diffusion MRI signal via the Fourier transform, but, as noted earlier, this fails to account for finite δ. Murday and Cotts [35] (see also [32]) derive an approximate model for the diffusion MRI signal for particles diffusing within spheres that accounts for finite pulse widths; their derivation adapts very simply to cylindrical geometries [36].

Various researchers construct simple geometric models from these primitives, so that p has analytic form, to try to match diffusion MRI signals more closely. For example, Szafer et al. [37] construct a brain-tissue model consisting of non-abutting semi-permeable square cylinders. Stanisz et al.'s model [30] contains non-abutting spherical glial cells and ellipsoidal axons with semi-permeable walls embedded in a homogeneous substrate. Assaf and Basser's model [29] contains cylindrical non-abutting axons with a gamma distribution of radii. Sen and Basser [38] add thickness to cylinder walls (like myelin sheaths) and show that different parameter settings reflect the observations from brain tissue with various pathological conditions. Other models, for example [39], use similar primitives to construct models for p in red blood cells.

3 Methods

This section first develops a simple model for diffusion in white matter, which is a simplification of the CHARMED model in [29], then discusses how to fit

models of this type to diffusion MRI data. Finally, it develops novel methods for optimizing the experiment design for this kind of model.

3.1 Simple Model

We use a model for white matter with two compartments between which there is no exchange of water molecules. The model assumes parallel cylindrical axon cells with equal radii and impermeable walls embedded in a homogeneous medium. Water in the intra-cellular compartment exhibits restricted diffusion, but the intra-cellular medium is also homogeneous with the same diffusivity as the extra-cellular medium. Water in the extra-cellular compartment exhibits hindered diffusion. The normalized MRI signal is then

$$A(\mathbf{G}, \Delta, \delta) = f A_r(\mathbf{G}, \Delta, \delta) + (1 - f) A_h(\mathbf{G}, \Delta, \delta), \tag{1}$$

where $f \in [0, 1]$ is the fraction of water molecules in the intra-cellular compartment, A_r is the signal from the intra-cellular compartment alone and A_h is the signal from the extra-cellular compartment.

We model the hindered diffusion in the extra-cellular compartment as simple (Gaussian) anisotropic diffusion so that

$$A_h(\mathbf{G}, \Delta, \delta) = \exp(-t\mathbf{q}^T \mathbb{D}_h \mathbf{q}), \tag{2}$$

where the diffusion time $t = \Delta - \delta/3$, $\mathbf{q} = \gamma\delta\mathbf{G}$ is the wavenumber, $\mathbb{D}_h = (d_\parallel - d_\perp)\mathbf{n}\mathbf{n}^T + d_\perp \mathbb{I}$ is the diffusion tensor, which is cylindrically symmetric with major eigenvector $\mathbf{n}$ in the fibre direction, with corresponding eigenvalue $d_\parallel$ (the diffusivity parallel to the fibre direction) and minor eigenvalues $d_\perp$ (the apparent diffusivity perpendicular to the fibre direction), and $\mathbb{I}$ is the identity tensor.

For the restricted component, we use Van Gelderen's extension [36] of Murday's model [35] for the MR signal from water diffusing in cylinders. The model approximates the MR signal from water molecules inside a cylinder of radius R for a gradient perpendicular to the fibre:

$$\log A_{r\perp}(G_\perp, \Delta, \delta) = \tag{3}$$
$$-2\gamma^2 G_\perp^2 \sum_{m=1}^{\infty} \left(\frac{2d_\parallel \alpha_m^2 \delta - 2 + 2Y(\delta) + 2Y(\Delta) - Y(\Delta - \delta) - Y(\Delta + \delta)}{d_\parallel^2 \alpha_m^6 (R^2 \alpha_m^2 - 1)} \right),$$

where $Y(x) = \exp(-d_\parallel \alpha_m^2 x)$, α_m is the mth root of $J_1'(\alpha R) = 0$ and J_1' is the derivative of the Bessel function of the first kind, order 1. The model assumes a Gaussian distribution of spin phases at echo time, which is not true in general, but provides a close approximation for most practical purposes [40].

In general, $\mathbf{G}$ is not perpendicular to $\mathbf{n}$, but we can decompose $\mathbf{G}$ into parallel $G_\parallel = |\mathbf{G}| \cos\theta$ and perpendicular $G_\perp = |\mathbf{G}| \sin\theta$ components. The two components attenuate the signal independently. The perpendicular

component attenuates according to (3) and, assuming simple diffusion along the fibre, the parallel component according to

$$A_{r\parallel}(G_\parallel, \Delta, \delta) = \exp(-t(\gamma\delta G_\parallel)^2 d_\parallel). \tag{4}$$

Note that we assume that the diffusion coefficient is the same as that parallel to the fibre direction in the extra-cellular compartment. The intra-cellular signal accounting for both components of the gradient is

$$A_r(\mathbf{G}, \Delta, \delta) = A_{r\perp}(|\mathbf{G}| \sin\theta, \Delta, \delta) A_{r\parallel}(|\mathbf{G}| \cos\theta, \Delta, \delta). \tag{5}$$

The key differences between the simple model above and the CHARMED model are the simple model has only a single axon radius rather than a Gamma distribution; the simple model assumes cylindrical symmetry of the diffusion tensor in the extra-cellular space; and the simple model assumes the same intrinsic diffusivity of the material inside and outside the cylinders. In summary, the parameters of the model are the following:

- The volume fraction of the intra-cellular compartment, f, which relates simply to the axon density
- The apparent diffusion coefficients, $d_\parallel$ and $d_\perp$
- The fibre direction, $\mathbf{n}$
- The axon radius, R

This is an experimental model only and does not account for many of the microstructural variables in real brain tissue. Some other variables are straightforward to incorporate. We can include distributions of axon radii by integrating (5) over a prior on R, as in [41], which assumes a gamma distribution of radii. Axon wall permeability allows exchange of particles between the two compartments, which we can model with a pair of coupled differential equations [30]. White matter also contains glial cells, which we might model as a third compartment with spherical restriction, as in [30]. We can allow distributions of fibre directions by integrating over a prior on $\mathbf{n}$. We may also be able to reduce the number of parameters in the model by one, by using models of tortuosity [31] to estimate $d_\perp$ from $d_\parallel$, f and R. Other effects, such as abutting cells and loss of percolation of the extra-cellular space are difficult to incorporate in analytic models.

3.2 Fitting

The model parameters do not relate linearly to the measurements, so we have to fit them using iterative optimization. The most common objective function for such optimizations is the sum of squared errors:

$$L_G = \sum_{k=1}^{NM} \sigma_k^{-2} (A(\mathbf{G}_k, \Delta_k, \delta_k) - \tilde{A}(\mathbf{G}_k, \Delta_k, \delta_k))^2, \tag{6}$$

where $\tilde{A}$ is the measurement, NM is the total number of measurements and σ_k is the standard deviation of the kth measurement.

The sum-of-squares objective function is the log-likelihood of the data on the assumption of Gaussian noise. However, the noise on MRI measurements in not Gaussian, but Rician [42], so that

$$P(\tilde{A}) = \frac{\tilde{A}}{\sigma^2} I_0\left(\frac{A\tilde{A}}{\sigma^2}\right) \exp\left(-\frac{A^2+\tilde{A}^2}{2\sigma^2}\right), \tag{7}$$

where I_i, $i = 0, 1, 2, \cdots$, are the modified Bessel functions of the first kind; $\tilde{A}$ and A are the measurement and signal (predicted by the model), respectively, and we drop the dependence on the pulse-sequence parameters from the notation for now. This model assumes that the signal is the modulus of a complex measurement with zero-mean Gaussian noise, with variance σ^2, on each component. A feature of the Rician noise model is that, unlike the Gaussian noise model, the expectation $E(\tilde{A})$ of the measurement is not A, but rather $E(\tilde{A}) = \sqrt{A^2+\sigma^2}$. The sum of squares objective function does not account for this bias and therefore introduces systematic bias to the parameter estimates. One simple way to correct for the bias [43] is to include the constant offset σ^2 as an extra parameter in the model and fit by minimum least squares. We shall refer to this fitting procedure as *corrected least squares*.

An alternative is to derive a similar objective function from the Rician distribution [44]. The log-likelihood of a single measurement comes from taking the log of (7). Then we sum over all measurements, drop constant terms and negate to obtain the objective function to minimize for fitting with the Rician noise model:

$$L_R = \sum_{k=1}^{NM} 2\log\sigma - \log I_0\left(\frac{A_k\tilde{A}_k}{\sigma^2}\right) + \frac{A_k^2+\tilde{A}_k^2}{2\sigma^2}. \tag{8}$$

Here we also include the noise variance σ as a model parameter in this fitting scheme, but it may be fixed if known.

The three objective functions listed earlier provide three fitting procedures. The experiments in later sections minimize each using a Levenburg–Marquardt procedure. The procedure requires the first derivatives of the objective function with respect to each model parameter, which are all simple to compute.

3.3 Sampling

This section addresses experiment design, that is, the choice of measurements that give the most accurate and precise estimates of the fitted model parameters. The tunable settings of the PGSE pulse sequence are $\mathbf{G}$, Δ and δ. Each measurement in the acquisition may use a different combination. We seek the set of combinations that give the best parameter estimates.

Assaf and Basser [41] acquire 1,024 measurements: 8 repeats of each combination of Δ at 8 settings in the range $[0.02, 0.15]$ s and $|\mathbf{G}|$ at 16 settings in the range $[0, 1.2]\,\mathrm{T\,m^{-1}}$; $\delta = 0.0025$ s throughout. They use ex vivo samples with known fibre orientation $\mathbf{n}$ and set $\mathbf{G}$ perpendicular to the fibre direction for all the measurements. Stanisz [30] uses a similar acquisition.

3.3.1 In Vivo Imaging Constraints

In vivo imaging places three key constraints on the acquisition sequence. First, we must limit the number of measurements to ensure that the acquisition time is tolerable for live subjects. Here we use a limit of $NM = 120$, which modern scanners can acquire in around 30 min. Second, power and safety constraints limit the maximum gradient strength we can use to about 0.04–$0.08\,\mathrm{T\,m^{-1}}$ for human subjects. Third, in general we have no a priori knowledge on the orientation of axons, so we require an acquisition that allows estimation of the model parameters for arbitrary $\mathbf{n}$. To handle arbitrary fibre directions, we must acquire measurements with various gradient directions. In a similar way, DT-MRI commonly uses 'high-angular resolution' acquisition schemes in which $|\mathbf{G}|$, Δ and δ are the same for each measurement, but each has a unique gradient direction, with the whole set distributed evenly on a hemisphere [45].

The first class of scheme we investigate acquires the same number M of measurements in each of the N gradient directions. The M combinations of $|\mathbf{G}|$, δ and Δ are the same in each direction. We choose the N directions by electrostatic minimization [8, 45], fix them and optimize the M combinations of $|\mathbf{G}|$, δ and Δ.

3.3.2 Objective Function

The Fisher information matrix and the Cramer–Rao lower bound (CRLB) are useful tools in experiment design [46]. The CRLB provides a lower bound on the variance of a fitted model parameter that often correlates closely with the true variance.

To optimize the acquisition, we aim to minimize, with respect to the set of $|\mathbf{G}|$, δ and Δ combinations, the sum of the standard errors of each model parameter

$$\tilde{F} = \sum_{i=1}^{K} \sigma_i^2 / p_i^2, \tag{9}$$

where p_i, $i = 1, \cdots, K$, are the model parameters and σ_i is the standard deviation of p_i. We do not know σ_i^2, so we use the CRLBs in their place.

The general form of the Fisher information matrix J has (i, j)th element

$$J_{ij} = E\left(\frac{\partial^2 L}{\partial p_i \partial p_j}\right), \tag{10}$$

where L is the log-likelihood given an appropriate noise model and E denotes expectation given that noise model. For Gaussian noise,

$$J_{ij} = \sigma^{-2} \sum_{k=1}^{NM} \frac{\partial A}{\partial p_i}(p_1, \cdots, p_K; \mathbf{G}_k, \Delta_k, \delta_k) \frac{\partial A}{\partial p_j}(p_1, \cdots, p_K; \mathbf{G}_k, \Delta_k, \delta_k), \quad (11)$$

which is simple to derive from (6) [47]. The CRLB for p_i is the ith diagonal element of J^{-1}, so we replace $\tilde{F}$ by

$$F = \sum_{i=1}^{K} (J^{-1})_{ii} / p_i^2. \quad (12)$$

The function F provides the basis of an objective function that we can minimize with respect to the M combinations of $|\mathbf{G}|$, δ and Δ. However, F depends on specific choices for the p_i, which take a range of values. The full objective function therefore integrates F over prior distributions on each p_i.

Here, we assume δ-function priors on the model parameters f, $d_\parallel$, $d_\perp$ and R. In particular, throughout we set $f = 0.7$, $d_\parallel = 1.7 \times 10^{-9}\,\mathrm{s\,m^{-2}}$, $d_\perp = 0.2 \times 10^{-9}\,\mathrm{s\,m^{-2}}$, $R = 10\,\mu\mathrm{m}$. We cannot use a δ-function prior for $\mathbf{n}$, however, as we require good parameter estimates for arbitrary fibre orientation. For orientation independence, we sum F over a set of directions. To construct the set, we choose the first element at random, minimize F and find the $\mathbf{n}$ in a large set of sample directions that has the largest F with the optimized acquisition. We add that $\mathbf{n}$ to the set and iterate until the $\mathbf{n}$ with the largest F is already in the set. The process usually converges with the set containing only three or four elements.

3.3.3 Rician CRLB

We can derive an alternative CRLB from the general expression for the Fisher information matrix in (10) that uses a Rician noise model. We need to compute the expectations of the second derivatives of L_R with respect to p_i and σ. For example,

$$E\left(\frac{\partial^2 L_R}{\partial p_j \partial p_k}\right) = \int_0^\infty \frac{\partial^2 L_R}{\partial p_j \partial p_k} P(\tilde{A}) \mathrm{d}\tilde{A}. \quad (13)$$

The following expectations are straightforward to compute, although the algebra is somewhat lengthy and the details are omitted here:

$$E\left(\frac{\partial^2 L_R}{\partial p_i \partial p_j}\right) = \sum_{k=1}^{NM} \sigma^{-4} \frac{\partial A_k}{\partial p_i} \frac{\partial A_k}{\partial p_j} (Z(\sigma, A_k) - A_k^2) \quad (14)$$

$$E\left(\frac{\partial^2 L_R}{\partial p_i \partial \sigma}\right) = \sum_{k=1}^{NM} 2\sigma^{-5} A_k \frac{\partial A_k}{\partial p_i} (Z(\sigma, A_k) - \sigma^2 - A_k^2)) \quad (15)$$

$$E\left(\frac{\partial^2 L_R}{\partial \sigma^2}\right) = \sum_{k=1}^{NM} 4\sigma^{-6} (\sigma^4 - \sigma^2 A_k^2 - A_k^4 + A_k^2 Z(\sigma, A_k)), \quad (16)$$

where

$$Z(\sigma, A_k) = \int_0^\infty \tilde{A}^2 I_1^2 \left(\frac{A_k \tilde{A}}{\sigma^2}\right) I_0^{-2} \left(\frac{A_k \tilde{A}}{\sigma^2}\right) P(\tilde{A}) \mathrm{d}\tilde{A}. \tag{17}$$

The function Z does not have closed form and we have to compute it numerically. In practice, we compute a look-up table of sampled values and use linear interpolation to estimate it during optimization.

3.3.4 Optimization

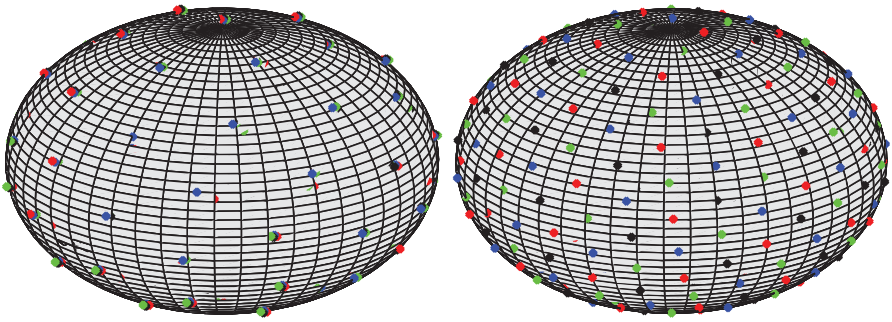
Optimizations of this type are complicated by many local minima in the objective function. We use a stochastic optimization algorithm, SOMA (Self-Organizing Migratory Algorithm) [48] with default settings, to perform the minimization. In each experiment, we repeat the optimization five times and pick the result with the smallest final value of the objective function.

During minimization, we enforce several constraints. First, $\Delta_i \geq \delta_i + \mathrm{P180}$, where P180 is the length of the 180° pulse, which we set to 0.005 s. Second, $\delta_i \geq 0$. Third, $0 \leq |\mathbf{G}_i| \leq G_{max}$, where G_{max} is the maximum available gradient strength. We also need to account for changes to the echo time, TE, required for the sequence, since T_2 effects reduce the signal more for longer TE, which increases the significance of the noise. To account for the effects of varying TE, we set σ proportional to $\exp(\max_i(\Delta_i + \delta_i)/\mathrm{T}_2)$ in the Fisher information matrix in (10) and (14). We set $\mathrm{T}_2 = 0.07\,\mathrm{s}$.

3.3.5 Subsets Acquisition Scheme

We also consider a variation on the acquisition scheme above that may give better rotational independence of the model parameter estimates. Generally, it is easier to estimate the model parameters when the $\mathbf{n}$ is approximately orthogonal to one of the N gradient directions. Distributions of N directions can leave large areas of the sphere uncovered, particularly when M is large (see Fig. 2). Improvements in orientational independence may come from spreading each group of M measurements out in the vicinity of one gradient direction.

A neat implementation of this idea exploits recent work by Cook et al. [49], who showed how to divide a set of NM evenly spaced directions into M subsets of evenly spaced directions. We use this to devise a similar acquisition scheme to that discussed earlier. The alternative divides a full set of NM gradient directions into M subsets and assigns one combination of $|\mathbf{G}|$, δ and Δ to each subset. Figure 2 compares the two acquisition schemes. Optimization is harder for these schemes as the number of sample $\mathbf{n}$ that we sum F over tends to be larger.



(a) (b)

Fig. 2. The two types of scheme. In each plot, points with the same colour are measurements acquired with the same combination of $|\mathbf{G}|$, δ and Δ. (**a**) The scheme in which all M measurements have the same direction. (**b**) The subsets scheme. In each case $M = 4$ and $N = 30$

4 Experiments and Results

This section contains some experiments and results that compare the experiment design and model fitting procedures from the previous section. All experiments use the simple model of Sect. 3.1 and we use only synthetic data throughout. Initial experiments (not shown) to compare different combinations of N and M suggest that the combinations $N = 30$, $M = 4$, $N = 24$, $M = 5$ and $N = 20$, $M = 6$ all give similar worst case performance over a large number of sample fibre directions. Other combinations give worse parameter estimates. Here we use $N = 30$, $M = 4$ throughout.

First we compare some acquisition schemes that the optimization procedure in Sect. 3.3 produces. For illustration, we set $R = 10\,\mu\text{m}$ and $G_{max} = 0.2\,\text{T}\,\text{m}^{-1}$. Table 1 shows the optimized acquisition scheme for the objective function based on the Gaussian CRLB. Table 2 shows the optimized scheme for the Rician CRLB and Table 3 shows the optimized scheme for the Rician CRLB using the subsets acquisition scheme that Sect. 3.3.5 describes.

Optimization of the Gaussian CRLB objective function suggests very high b-values that result in measurements with very low signal-to-noise. Incorporating the Rician noise model penalizes high b-values more and suggests a more intuitively reasonable set of measurements. The optimized combinations of $|\mathbf{G}|$, δ and Δ are similar for the schemes with measurements in the same directions and subsets.

Next we use simulation experiments to check the accuracy with which we can recover known parameter settings using the different optimized acquisition schemes in Tables 1–3 and the different model fitting techniques in Sect. 3.2. In all the simulation experiments, we choose settings for all the model parameters listed at the end of Sect. 3.1 and synthesize data from the model using

Table 1. Optimized combinations of $|\mathbf{G}|$, δ and Δ from the Gaussian CRLB objective function

$\lvert\mathbf{G}\rvert$ ($\mathrm{T\,m^{-1}}$)	Δ (s)	δ (s)	$\lvert\mathbf{q}\rvert$ ($10^5\,\mathrm{m^{-1}}$)	t (s)	b ($\mathrm{mm^2\,s^{-1}}$)
0.200	0.022	0.017	9.085	0.017	13,775
0.193	0.012	0.007	3.473	0.010	1,145
0.192	0.032	0.007	3.665	0.030	4,009
0.183	0.021	0.014	6.598	0.016	7,137

The table includes the more familiar quantities $|\mathbf{q}|$, $t = \Delta - \delta/3$ and $b = t|\mathbf{q}|^2$

Table 2. Optimized combinations of $|\mathbf{G}|$, δ and Δ from the Rician CRLB objective function

$\lvert\mathbf{G}\rvert$ ($\mathrm{T\,m^{-1}}$)	Δ (s)	δ (s)	$\lvert\mathbf{q}\rvert$ ($10^5\,\mathrm{m^{-1}}$)	t (s)	b ($\mathrm{mm^2\,s^{-1}}$)
0.200	0.010	0.005	2.832	0.009	684
0.180	0.031	0.005	2.591	0.029	1,956
0.166	0.024	0.013	5.609	0.019	6,111
0.160	0.023	0.013	5.753	0.018	6,085

Table 3. Optimized combinations of $|\mathbf{G}|$, δ and Δ, using the subsets acquisition scheme, from the Rician CRLB objective function

$\lvert\mathbf{G}\rvert$ ($\mathrm{T\,m^{-1}}$)	Δ (s)	δ(s)	$\lvert\mathbf{q}\rvert$ ($10^5\,\mathrm{m^{-1}}$)	t(s)	b ($\mathrm{mm^2\,s^{-1}}$)
0.200	0.028	0.010	5.108	0.025	6,544
0.200	0.010	0.005	2.873	0.009	708
0.177	0.027	0.011	5.143	0.023	6,177
0.147	0.032	0.006	2.500	0.029	1,836

the acquisition scheme we are testing. Then we add synthetic Rician noise and refit the model by minimizing one of the objective functions in Sect. 3.2. To minimize the objective functions, we use a Levenburg–Marquart procedure. The initial settings of the model parameters before optimization are perturbations of the true values by 20% Gaussian noise. We repeat each optimization ten times and pick the result with the lowest final value of the objective function. In these experiments, we do not fit for the fibre direction, but fix it to the true value to focus on the ability to recover the scalar parameters.

Figure 3 shows scatter plots of the fitted model parameters $d_\perp$, $d_\parallel$, f and R, over 100 trials with independent noise, for each fitting procedure using each of the optimized schemes in Tables 1–3. The noise level in each trial is such that the signal-to-noise ratio with no diffusion weighting is 20. The plots show very clearly the advantage of the Rician CRLB optimization for experiment design, as the recovered model parameters have much greater precision and accuracy

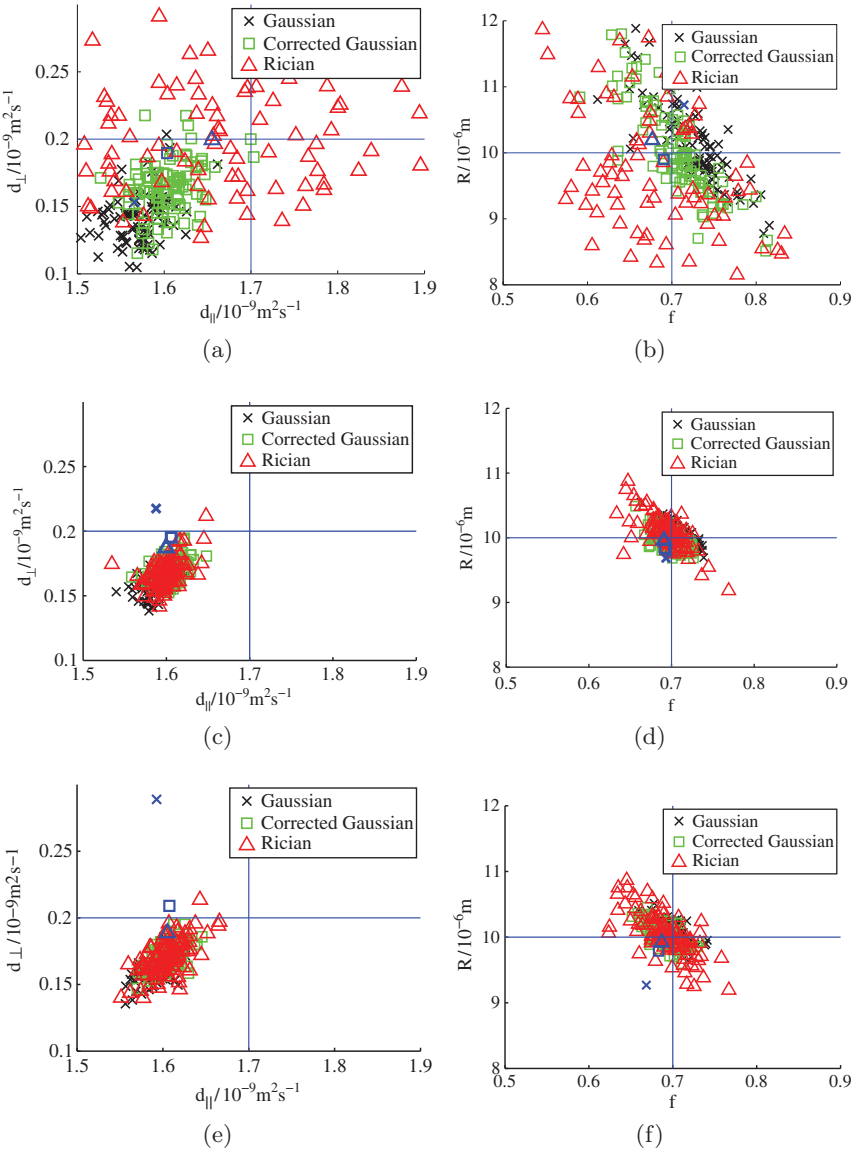


Fig. 3. Scatter plots of the fitted model parameters for each optimized acquisition scheme and objective function. The left column plots fitted $d_\parallel$ against $d_\perp$. The right column plots the fitted f against R. The first row shows the scatter for the Gaussian CRLB optimized scheme in Table 1; the second row for the Rician CRLB optimized scheme in Table 2; the third row for the Rician CRLB optimized scheme with subsets in Table 2. The blue cross hairs indicate the true parameter values. The blue markers show average values over all trials

than from the Gaussian-CRLB-optimized scheme. The subsets scheme does not show any advantage over the scheme with measurements in the same directions. In fact, in this particular experiment, it seems to give slightly worse results. In general, the least-squares fitting routines appear to give greater precision in the recovered model parameters than the Rician objective function. The results show the expected bias of the uncorrected least-squares procedure. The corrected least-squares objective function appears to produce the highest accuracy and precision overall. The fitting procedure we use occasionally produces wildly inaccurate results (off the graph), which skews the positions of the mean fitted values in Fig. 3. We also observe systematic bias in the fitted diffusivity parameters, which is a consequence of the Rician noise distribution that arises even using the correct noise model during fitting.

5 Discussion

In summary, we have discussed models that relate the diffusion MRI signal directly to parameters of the tissue microstructure that are potentially useful biomarkers. We have developed a simple model, which we use to illustrate various methods for fitting models to diffusion MRI data and designing optimal acquisition schemes. The fitting and experiment design techniques do not depend on the specific model and will work with any similar model. In particular, they may provide optimal acquisition schemes for DT-MRI, which are not yet agreed.

Simulation results are promising and suggest that measuring axon radii and density in vivo is feasible. Further work is required for practical application, however. The model we test here is very simple and needs to incorporate more effects to provide useful biomarkers. Further work will also improve the optimization for experiment design and the procedures for minimizing the fitting objective functions. For further work in this area, see [50, 51].

References

[1] Basser PJ, Matiello J and Le Bihan D (1994) MR diffusion tensor spectroscopy and imaging. *Biophysical Journal* **66**, 259–67

[2] Alexander DC, Barker GJ and Arridge SR (2002) Detection and modelling of non-Gaussian apparent diffusion coefficient profiles in human brain data. *Magnetic Resonance in Medicine* **48**, 331–40

[3] Alexander AL, Hasan KM, Lazar M, Tsuruda JS and Parker DL (2001) Analysis of partial volume effects in diffusion-tensor MRI. *Magnetic Resonance in Medicine* **45**, 770–780

[4] Frank LR (2002) Characterization of anisotropy in high angular resolution diffusion-weighted MRI. *Magnetic Resonance in Medicine* **47**, 1083–1099

[5] Parker GJM and Alexander DC (2003) Probabilistic Monte Carlo Based Mapping of Cerebral Connections Utilising Whole-Brain Crossing Fibre Information. *Proc. of 18th IPMI Conference*, LNCS 2732, Springer, Ambleside, pp. 684–695

[6] Tuch DS, Reese TG, Wiegell MR, Makris N, Belliveau JW and Wedeen VJ (2002) High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine* **48**, 577–582

[7] Tuch DS, Reese TG, Wiegell MR and Wedeen VJ (2003) Diffusion MRI of complex neural architecture. *Neuron* **40**, 885–895

[8] Jansons KM and Alexander DC (2003) Persistent Angular Structure: new insights from diffusion MRI data. *Inverse Problems* **19**, 1031–1046

[9] Alexander DC (2005) Maximum entropy spherical deconvolution for diffusion MRI, *Proc. IPMI*, LNCS 3565, Springer, Fort Collins, pp. 76–87

[10] Anderson AW (2005) Measurement of fiber orientation distributions using high angular resolution diffusion imaging. *Magnetic Resonance in Medicine* **54**, 1194–1206

[11] Tournier J-D, Calamante F, Gadian DG and Connelly A (2004) Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage* **23**, 1176–1185

[12] Alexander DC (2006) An introduction to computational diffusion MRI: the diffusion tensor and beyond. *Visualization and Image Processing of Tensor Fields,* Weichert J and Hagen H (ed.), Springer, Berlin Heidelberg New York

[13] Behrens TEJ, Johansen-Berg H, Jbabdi S, Rushworth MFS and Woolrich MW (2006) Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage* **34**, 144–155

[14] (2006) Special issue: Reviews on Neurological and Psychiatric Disease. *Neuron* **52**, 15–168

[15] Toosy T, Werring DJ, Orrell RW, Howard RS, King MD, Barker GJ, Miller DH and Thompson AJ (2003) Diffusion tensor imaging detects corticospinal tract involvement at multiple levels in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry* **74**, 1250–1257

[16] McKinstry RC, Mathur A, Miller JH, Ozcan A, Snyder AZ, Schefft GL, Almli CR, Shiran SI, Conturo TE and Neil JJ (2002) Radial organization of developing preterm human cerebral cortex revealed by non-invasive water diffusion anisotropy MRI. *Cerebral Cortex* **12**, 1237–1243

[17] Pfefferbaum A, Sullivan EV, Hedehus M, Lim KO, Adalsteinsson E and Moseley M (2000) Age-related decline in brain white matter anisotropy measured with spatially corrected echo-planar diffusion tensor imaging. *Magnetic Resonance in Medicine* **44**, 259–268

[18] Behrens TEJ, Johansen-Berg H, Woolrich MW, Smith SM, Wheeler-Kingshott CAM, Boulby PA, Barker GJ, Sillery E L, Sheehan K, Ciccarelli O, Thompson AJ, Brady JM and Matthews PM (2003)

Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature Neuroscience* **7**, 750–757

[19] Parker GJM, Luzzi S, Alexander DC, Wheeler-Kingshott CAM, Ciccarelli O and Lambon-Ralph MA (2005) Non-invasive structural mapping of two auditory-language pathways in the human brain. *NeuroImage* **24**, 656–666

[20] Jones D, Catani M, Pierpaoli C, Reeves SJ, Shergill SS and O'Sullivan M (2005) Age effects on diffusion tensor magnetic resonance imaging tractography measures of frontal cortex connections in schizophrenia. *Human Brain Mapping* **27**, 230–238

[21] Newton JM, Ward NS, Parker GJM, Deichmann R, Alexander DC, Friston KJ and Frackowiak RSJ (2006) Non-invasive mapping of corticofugal fibres from multiple motor areas – relevance to stroke recovery. *Brain* **127**, 1844–1858

[22] Powell HWR, Parker GJM, Alexander DC, Symms MR, Boulby PA, Barker GJ, Koepp MJ and Duncan JS (2005) MR-Tractography predicts visual field defects following temporal lobe resection. *Neurology* **64**, 596–599

[23] Sotak CH (2002) The role of diffusion tensor imaging in the evaluation of ischemic brain injury – a review. *NMR in Biomedicine* **15**, 561–569

[24] Reese TG, Heid O, Weisskoff RM and Wedeen VJ (2003) Reduction of eddy-current-induced distortion using a twice-refocussed spin echo. *Magnetic Resonance in Medicine* **49**, 177–182

[25] Merboldt KD, Hanicke W and Frahm J (1991) Diffusion imaging using stimulated echoes. *Magnetic Resonance in Medicine* **19**, 233–239

[26] Miller KL, Hargreaves BA, Gold GE and Pauly JM (2004) Steady-state diffusion-weighted imaging of in vivo knee cartilage. *Magnetic Resonance in Medicine* **51**, 394–398

[27] Callaghan PT (1991) *Principles of Magnetic Resonance Microscopy*, Oxford Science Publications, Oxford, UK

[28] Mitra PP and Halperin BI (1995) Effects of finite gradient-pulse widths in pulsed-field-gradient diffusion measurements. *Journal of Magnetic Resonance* **113**, 94–101

[29] Assaf Y and Basser PJ (2005) Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *NeuroImage* **30**, 1100–1111

[30] Stanisz GJ, Szafer A, Wright GA and Henkelman M (1997) An analytical model of restricted diffusion in bovine optic nerve. *Magnetic Resonance in Medicine* **37**, 103–111

[31] Nicholson C (2001) Diffusion and related transport mechanisms in brain tissue. *Reports on Progress in Physics* **64**, 815–884

[32] Neuman CH (1974) Spin echo of spins diffusing in a bounded medium. *Journal of Chemical Physics* **60**, 4508–4511

[33] Callaghan PT (1995) Pulsed-gradient spin-echo NMR for planar, cylindrical and spherical pores under conditions of wall relaxation. *Journal of Magnetic Resonance* **113**, 53–59
[34] Soderman O and Jonsson B (1995) Restricted diffusion in cylindrical geometry. *Journal of Magnetic Resonance A* **117**, 94–97
[35] Murday JS and Cotts RM (1968) Self-diffusion coefficient of liquid lithium. *Journal of Chemical Physics* **48**, 4938–4945
[36] Van Gelderen P, DesPres D, van Zijl PCM and Moonen CTW (1994) Evaluation of Restricted DIffusion in Cylinders. Phosphocreatine in rabbit leg muscle. *Journal of Magnetic Resonance Series B* **103**, 255–260
[37] Szafer A, Zhong J and Gore JC (1995) Theoretical model for water diffusion in tissues. *Magnetic Resonance in Medicine* **33**, 697–712
[38] Sen PN and Basser PJ (2005) A model for diffusion in white matter. *Biophysical Journal* **89**, 2927–2938
[39] Latour LL, Svoboda K, Mitra PP and Sotak CH (1994) Time-dependent diffusion of water in a biological model system. *Proceedings of the National Academy of Sciences* **91**, 1229–1233
[40] Balinov B, Jonsson B, Linse P, Soderman O (1993) The NMR self-diffusion method applied to restricted diffusion. Simulation of echo attenuation from molecules in spheres and between planes. *Journal of Magnetic Resonance Series A* **104**, 17–25
[41] Assaf Y, Freidlin RZ and Basser PJ (2005) The measurement of the axon diameter distribution in white matter using diffusion MR methods. *Proceedings of ISMRM*, Miami, ISMRM, 842
[42] Henkelman RM (1985) Measurement of signal intensities in the presence of noise in MR images. *Medical Physics* **12**, 232–233
[43] Jones DK and Basser PJ (2004) Squashing peanuts and smashing pumpkins: How noise distorts diffusion-weighted MR data. *Magnetic Resonance in Medicine* **52**, 979–993
[44] Sijbers J, den Dekker AJ, Scheunders P and van Dyck D (1998) Maximum-likelihood estimation of Rician distribution parameters. *IEEE Transactions on Medical Imaging* **17**, 357–361
[45] Jones DK, Horsfield MA and Simmons A (1999) Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine* **42**, 515–525
[46] Pukelsheim F (1993) *Optimal Design of Experiments* , John Wiley, New York
[47] Lindgren BW (1993) *Statistical Theory*, Chapman and Hall, New York
[48] Zelinka I (2004) SOMA – self organizing migrating algorithm, *New Optimization Techniques in Engineering*, Babu BV and Onwubolu G (ed.), Springer, Berlin Heidelberg New York
[49] Cook PA, Boulby PA, Symms MR and Alexander DC (2007) Optimal acquisition orders of diffusion-weighted MRI measurements. *Journal of Magnetic Resonance Imaging* **25**, 1051–1058

[50] Alexander DC (2007) Axon radius measurements in vivo from diffusion MRI: a feasibility study, *Proc. ICCV Workshop on MMBIA*, Rio de Janiero, IEEE

[51] Alexander DC (2008) A general framework for experiment design in diffusion MRI and its application to measuring direct tissue-microstructure features. *Magnetic Resonance in Medicine* **in Press**,

Tensors, Polynomials and Models for Directional Data

P.G. Batchelor

Imaging Sciences, King's College, London
Philip.batchleor@kcl.ac.uk

Summary. In this chapter we describe tensorial tools for the study of directional data. We first make explicit different tensor representations. We then use this to compare models, as one of the tensor representations is polynomial. Thus a multivariate Taylor expansion uses Higher Order Tensors, and models will use an expansion in higher order polynomials. We show provide formulas for different model-to-model conversions.

1 Introduction

We describe tensorial tools for the study of directional data. By this we mean data sampled on spheres, or spheres with antipodal points identified (projective space). Tensors play a role because we use *models*, that is continuous functions on the space of directions. Such functions are approximated by polynomials of varying orders. This defines a natural decomposition in homogeneous terms, which leads to the connection between multivariate polynomials and symmetric tensors.

The text is structured in opposite order: we first describe tensors (Sect. 2), especially symmetric ones, and their connection with polynomials. This seemed to be of interest to some of the audience in the Dagstuhl meeting, thus hopefully the same is true for readers of this chapter. We then specialise to directional signals, in a more specific section (Sect. 3).

2 Tensors, Symmetric Tensors and Multivariate Polynomials

Several options are available to write tensors, for example as multi-dimensional arrays of coefficients or in a coordinate free form. The essential characteristic is, however, the same; tensors are multi-linear objects. This explains the coefficients, they are coefficients of linear combination in a specific basis, but it also

explains how the tensors change under linear changes of coordinates. Let V be a linear space. The dimension of V is denoted by d. Multi-linear functions of p variables on V form a vector space of dimension d^p. When $p = 1$, this space is the *dual* space V^*. Let $\mathbb{A}, \mathbb{B} \in V^*$, thus $\mathbb{A}, \mathbb{B}$ are linear maps on V.

Definition 1. *The* tensor product $\mathbb{A} \otimes \mathbb{B}$ *of* $\mathbb{A}$ *with* $\mathbb{B}$ *is the multi-linear map from* $V^2 \to \mathcal{K}$ *(bilinear map) defined by*

$$\mathbb{A} \otimes \mathbb{B} : V^2 \to \mathcal{K}; \quad (\mathbb{A} \otimes \mathbb{B})(\mathbf{v}_1, \mathbf{v}_2) = \mathbb{A}(\mathbf{v}_1)\mathbb{B}(\mathbf{v}_2) \tag{1}$$

for any $\mathbf{v}_1, \mathbf{v}_2 \in V$. *Here* $\mathcal{K}$ *is a number field, either* $\mathbb{R}$ *or* $\mathbb{C}$. *The definition can be extended to vector-valued tensor product and spaces, below we use an inner product to perform such identifications.* V *is a vector space over* $\mathbb{K}$, *which means it is isomorphic to* $\mathbb{R}^d$ *or* $\mathbb{C}^d$.

The space of bilinear maps is then spanned by tensor products. The definition extends to higher orders of p.

The dimension of the dual space is also d, thus V and V^* are isomorphic. When an inner product $\langle \mathbf{v}, \mathbf{w} \rangle$ is defined on V, we have a canonical (coordinate independent) isomorphism: any vector $\mathbf{v} \in V$ defines a linear form in V^* via $\langle \mathbf{v}, \cdot \rangle$. For the standard space $V = \mathbb{R}^d$, where we write vectors as column arrays, this amounts to taking the transpose row vector. Furthermore, the dual of the dual is V. Generally, we can define tensor products of p vectors and q forms, which we can identify to tensors of order $r = p + q$ by using the canonical isomorphism. In summary, we can talk about *the* tensor space of pth order tensors, and think of tensors as multi-linear maps over V of order p.

The inner product extends naturally to the tensor product (see also M. Moakher's text). In other words, we can talk about the norm of a tensor, or the distance between two tensors. To avoid the cumbersome indicial notations, we introduce *multi-indices* $\mathbf{i} = (i_1, \ldots, i_p)$. Then the inner product can be written as a normal dot product, or using the Einstein convention: $\langle \mathbb{A}, \mathbb{B} \rangle = \mathbf{a_i} \cdot \mathbf{b_i}$. Normal dot product refers to the sum over corresponding components. The right hand-side here uses a vectorisation of $\mathbb{A}, \mathbb{B}$ to $\mathbf{a}, \mathbf{b}$.

2.1 Tensor Functions

We can define different functions from $\mathcal{T}^p(V)$ to $\mathcal{T}^q(V)$, the space of tensors of order p to the space of tensors of order q.

Definition 2. *Define the contraction* $C_{ij} : \mathcal{T}^p(V) \to \mathcal{T}^{p-2}(V)$ *and transposition, symmetry and antisymmetry operators* $\tau_{ij}, \sigma_{ij}, \alpha_{ij} : \mathcal{T}^p(V) \to \mathcal{T}^p(V)$ *by*

$$(C_{ij}\mathbb{A})[\mathbf{v}_1, \ldots, \mathbf{v}_{p-2}] := \sum_k \mathbb{A}[\mathbf{v}_1, \ldots, \mathbf{e}_k, \ldots, \mathbf{e}_k, \ldots, \mathbf{v}_p] \tag{2}$$

$$(\tau_{ij}\mathbb{A})[\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_p] := \mathbb{A}[\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_p] \tag{3}$$

$$(\sigma_{ij}\mathbb{A}) := \frac{1}{2}(\mathbb{A} + \tau_{ij}\mathbb{A}); \quad (\alpha_{ij}\mathbb{A}) := \frac{1}{2}(\mathbb{A} - \tau_{ij}\mathbb{A}); \tag{4}$$

The image tensors from σ_{ij}, α_{ij} are *symmetric* or *antisymmetric* in entries ij and the following applies mostly to symmetric tensors. In the same way, we can construct tensors with specific symmetries in indices by summing over all corresponding permutations. The number of transpositions required to express a permutation is unique modulo 2. $\mathbb{A}$ is fully symmetric if it is invariant under any permutation of the p arguments. For symmetric tensors, the standard component notation is not optimal in the sense that components are redundant. For a multi-index $\mathbf{i}$, we write $\hat{\mathbf{i}}$ for the re-ordered multi-index $i_1 \leq \ldots \leq i_p$, and given a vector $\mathbf{x}$, we write $\mathbf{x}_{\hat{\mathbf{i}}}$ for the product $x_{i_1} \cdots x_{i_p}$ of its components. Formally, we have a many-to-one map h such that $h(\mathbf{i}) = \hat{\mathbf{i}}$. The ordered multi-indices define a unique indexing of symmetric tensors, it is an example of monomial ordering [7].

2.1.0.1 Notations

We try to use here notations as consistent as possible with the ones in the chapter by M. Moakher. As the focus here is on tensors of any order, we use $\mathbb{A}, \mathbb{B}$ for abstract tensors, of any order. Vectors, which can be identified with tensors of order 1, are lower case bold $\mathbf{v}, \mathbf{a}, \mathbf{b}$. The number of components should be clear from the context, thus $\mathbf{v} \in V$ would have d components, where $\mathbf{a}, \mathbf{b}$, the vectors of tensor components would have d^p components. Their exact ordering would depend on the bijection ψ between the space of tensors and $\mathbb{R}^{d^p}$. See chapter by M. Moakher in this book for more details on the bijection ψ. The letter $\mathbf{x}$ denotes an independent variable, here d-dimensional vector in V.

For a symmetric tensor $C_{ij}(\mathbb{A}) = C_{i'j'}(\mathbb{A})$ for all $i \neq j, i' \neq j'$. In other words we have a unique contraction:

Definition 3. *The* Trace *is the contraction operator from* $\mathcal{S}^p(V) \to \mathcal{S}^{p-2}(V)$ *(e.g. last two indices in practice).*

Note that this definition differs from the one by Moakher in this book and the 'generalised trace' defined by Özarslan [14]. Our definition produces tensors of lower order and a scalar only for tensors of order 2, whereas theirs produces scalars, but the definitions are consistent for order 2 tensors.

The inner product of the space of tensors induces an inner product on the space of symmetric tensors. Using the multi-index notation $\mathbf{a}_\mathbf{i}\mathbf{b}_\mathbf{i} = \sqrt{n_{\hat{\mathbf{i}}}}\mathbf{a}_{\hat{\mathbf{i}}}\sqrt{n_{\hat{\mathbf{i}}}}\mathbf{b}_{\hat{\mathbf{i}}}$. Here $n_{\hat{\mathbf{i}}}$ is the number of equivalent multi-indices, that is $n_{\hat{\mathbf{i}}} = \text{card}\{h^{-1}(\hat{\mathbf{i}})\}$.

We can define a homogeneous function $q_\mathbb{A} : V \to \mathbb{K}$ by $q_\mathbb{A}(\mathbf{x}) := \mathbb{A}[\mathbf{x}, \ldots, \mathbf{x}]$. Obviously, $q_\mathbb{A}$ is blind to permutations of the arguments in $\mathbb{A}$, thus we can restrict to fully symmetric polynomials, and by multi-linearity $q_\mathbb{A}(\mathbf{x}) = A_{i_1 \ldots i_p} x_{i_1} \cdots x_{i_p} = \mathbf{a}_\mathbf{i}\mathbf{x}_\mathbf{i}$, where the last tensor product in $\mathbf{x}$ has p terms. This is a multivariate polynomial of order p, and equivalently a multivariate polynomial q of order p defines a unique fully symmetric tensor A_q. In other words, the set of homogeneous polynomials of order p $\mathcal{P}^p$ is isomorphic

to the set of fully symmetric tensors of order p, $\mathcal{P}^p \simeq \mathcal{S}^p$. As $i_n \in \{1, \ldots, d\}$, some indices are repeated, and we can rewrite $x_{i_1} \cdots x_{i_p} = x_1^{p_1} \cdots x_d^{p_d}$, where p_i is the number of times that i appears in the multi-index $\mathbf{i}$. Note that $p_i \in \{0, \ldots p\}$ is 0-based. This provides an alternative notation for symmetric tensor components, where we would write A indexed by the multi-indices of p_i. For a vector $\mathbf{x}$, we define the generalised power $\mathbf{x}^{\mathbf{p}} = x_1^{p_1} \cdots x_d^{p_d}$ for such partition $\mathbf{p}$, $p_1 + \cdots + p_d = p$.

Example 1. Let $q(\mathbf{x}) = x^2y^2 + x^2z^2 + y^2z^2$. The standard component notation for this tensor could be summarised here by $A_{1122} = A_{1133} = A_{2233} = 1$, with all other components being 0, in general we would need to write a four-dimensional array of components. To be explicit, the component A_{1122} multiplies the monomial $xxyy$, A_{1133} multiplies $xxzz$. We see that this notation does not simplify monomials. Alternatively, we could write $A_{\tilde{2}\tilde{0}\tilde{2}} = A_{\tilde{0}\tilde{2}\tilde{2}} = A_{\tilde{2}\tilde{2}\tilde{0}} = 1$, where the tilde means the components are in the form $(p_1, \ldots, p_d)$. Specifically, $A_{\tilde{2}\tilde{0}\tilde{2}}$ multiplies $x^2y^0z^2$.

In summary, to represent a fully symmetric tensor of order p we have the following options:

Polynomial	Ordered multi-index of length p	Partition of p
degree p	$\texttt{length}(\hat{\mathbf{i}})$	$\texttt{sum}(\mathbf{p})$
$\dim(V)$	$\max(\{\hat{\mathbf{i}}\})$	$\texttt{length}(\mathbf{p})$

Note that rank has another meaning related to linear independence, thus we use the word 'order'.

Given a vector $\mathbf{v}$, we know it defines a tensor of order 1 by inner product $(\mathbf{v} \cdot \mathbf{x})$. Its tensor components are the vector components. Consider the pth power $(\mathbf{v} \cdot \mathbf{x})^p$. By the multinomial formula

$$(\mathbf{v} \cdot \mathbf{x})^p = \sum_{p_1 + \cdots + p_d = p} \binom{p}{p_1, \ldots, p_d} v_1^{p_1} x_1^{p_1} \cdots v_d^{p_d} x_d^{p_d} = \sum_{\mathbf{p}} \mathbf{x}^{\mathbf{p}} c_{\mathbf{p}}(\mathbf{v}).$$

This defines a homogeneous polynomial of order p, and we have an explicit formula for the corresponding *symmetric* tensor components.

This defines a function of $\mathbf{x}$, and the counting terms $n_{\mathbf{p}} = n_{p_1, \ldots p_d} = \binom{p}{p_1, \ldots, p_d} = \binom{p}{\mathbf{p}}$ were put together with the other vector components. These counting terms come from the projection from the space of all tensors to the space of symmetric tensors: A_{1212} and A_{1122} are different in the space of all tensors, but identical in the space of symmetric tensors. If we replace $\mathbf{v}$ by $\mathbf{x}$, we get a homogeneous polynomial of degree $2p$, and by the same formula

$$(\mathbf{x} \cdot \mathbf{x})^p = \sum_{\mathbf{p}} \sqrt{n_{\mathbf{p}}} \mathbf{x}^{\mathbf{p}} \sqrt{n_{\mathbf{p}}} \mathbf{x}^{\mathbf{p}}. \tag{5}$$

We write $\sqrt{n_{\mathbf{p}}}\sqrt{n_{\mathbf{p}}}$ instead of $n_{\mathbf{p}}$ to underline the symmetry in the right hand term. This is yet another explanation for the $\sqrt{2}$ required to define an orthonormal basis on the space of symmetric tensors. This expression should be compared with the ones in the chapter by M. Moakher.

2.1.0.2 Differential Operators

The *Laplace* operator $\Delta := \nabla^2$ reduces polynomial order by 2: $\mathcal{P}^p \to \mathcal{P}^{p-2}$. By multi-linearity $q_{\mathbb{A}}(\mathbf{x} + t\mathbf{e}) = q_{\mathbb{A}}(\mathbf{x}) + t\sum_l A[\mathbf{x}, \ldots, \mathbf{x}, \mathbf{e}_l] + O(t^2)$. Thus, $\nabla q_{\mathbb{A}}(\mathbf{x}) \cdot \mathbf{e} = \frac{d}{dt}|_{t=0} q_A(\mathbf{x}+t\mathbf{e}) = \sum_l \mathbb{A}[\mathbf{x}, \ldots, \mathbf{x}, \mathbf{e}_l]$. The divergence of a vector field of the form $(f(\mathbf{x}))\mathbf{e}$ is $\nabla f(\mathbf{x}) \cdot \mathbf{e}$, thus we can reuse the previous formula:

$$\nabla \cdot \nabla q_{\mathbb{A}}(\mathbf{x}) = \sum_n \sum_l \mathbb{A}(\mathbf{x}, \ldots, \mathbf{x}, \mathbf{e}_l, \mathbf{e}_n)\mathbf{e}_l \cdot \mathbf{e}_n = \sum_l \mathbb{A}(\underbrace{\mathbf{x}, \ldots, \mathbf{x}}_{p-2 \text{ terms}}, \mathbf{e}_l, \mathbf{e}_l)$$

because the basis is orthonormal. The last term is a trace.

$$\nabla q_{\mathbb{A}}(\mathbf{x}) = \sum_l \mathbb{A}[\underbrace{\mathbf{x}, \ldots, \mathbf{x}}_{p-1 \text{ terms}}, \mathbf{e}_l]\mathbf{e}_l; \quad \Delta q_{\mathbb{A}}(\mathbf{x}) = \operatorname{tr}(\mathbb{A})\, \underbrace{[\mathbf{x}, \ldots \mathbf{x}]}_{p-2 \text{ terms}}. \tag{6}$$

In other words, the algebra of homogeneous polynomials and the symmetric tensors have a one-to-one correspondence, which preserves some important operators.

Note that if $\mathbb{A}$ and $\mathbb{B}$ are the symmetric tensor products $\mathbb{A} = \mathbf{e}_{i_1} \cdots \mathbf{e}_{i_p}$ and $\mathbb{B} = \mathbf{e}_{j_1} \cdots \mathbf{e}_{j_q}$, $q_{\mathbb{A}}(\mathbf{x})q_{\mathbb{B}}(\mathbf{x}) = x_{i_1} \cdots x_{i_p} x_{j_1} \cdots x_{j_q} = q_{\mathbf{e}_{i_1} \cdots \mathbf{e}_{i_p} \cdot \mathbf{e}_{j_1} \cdots \mathbf{e}_{j_q}}(\mathbf{x}) = q_{\mathbb{A} \cdot \mathbb{B}}(\mathbf{x})$ and $q_{\alpha\mathbb{A}+\beta\mathbb{B}}(\mathbf{x}) = \alpha q_{\mathbb{A}}(\mathbf{x}) + \beta q_{\mathbb{B}}(\mathbf{x})$; thus the symmetric tensor product corresponds to the polynomial product.

2.1.0.3 Integral Operators

To extend further the table of equivalences, we would like to find a polynomial expression for the tensor inner product. The restriction of the L^2-inner product on the sphere to homogeneous polynomials induces a bilinear form $\int_{S^{d-1}} q_{\mathbb{A}} q_{\mathbb{B}}$. As $q_{\mathbb{A}}(\mathbf{x})q_{\mathbb{B}}(\mathbf{x}) = q_{\mathbb{A}\mathbb{B}}(\mathbf{x})$, we need to be able to integrate a polynomial $q_{\mathbb{C}}(\mathbf{x})$ on the unit sphere. Note that we can write $q_{\mathbb{C}}(\mathbf{x}) = \langle \mathbb{C}, \mathbb{X} \rangle$, where $\mathbb{X}$ is the tensor whose components are monomials $\mathbf{x}_{\hat{\mathbf{i}}} = \mathbf{x}^{\mathbf{P}}$. By integration over the sphere, we get

$$\int_{S^{d-1}} q_{\mathbb{C}}\, dS = A(d)\langle \mathbb{C}, \mathbb{M} \rangle \tag{7}$$

$$\langle q_{\mathbb{A}}, q_{\mathbb{B}} \rangle_{L^2(S^{d-1})} = A(d)\langle \mathbb{A}, \mathbb{M}\mathbb{B} \rangle_{\mathcal{T}}, \tag{8}$$

where $A(d)$ is the surface area of the unit sphere in dimension d. In the second row, the *tensor of moments* $\mathbb{M}$ is interpreted as a *linear mapping* on the set of tensors (See M. Moakher's chapter). Let $p_1 + \cdots + p_d = p$ be a partition of p. If any p_i is odd, the integral is zero by symmetry. Otherwise, we get the $(p_1, \ldots, p_d)$ moment

Proposition 1. $\mathbb{M}_{p_1,\ldots,p_d} = \mathbb{M}_{\mathbf{p}} = \frac{\Gamma((p_1+2)/2) \cdots \Gamma((p_d+2)/2)}{\Gamma((p_1+\cdots+p_d+2)/2)} =: \frac{\Gamma(\mathbf{p}/2+1)}{\Gamma(p/2+1)}$,

where we have extended the Γ-function to partitions in the obvious way. This can be computed by direct integration over a sphere (see [1, 3]). Alternatively,

because moments for a uniform distribution on a sphere cannot depend on the coordinate systems, the components must be *invariant* under the effect of the group of rotations [5, 6]: $\mathbb{M}^R_{p_1,\ldots,p_2} = \mathbb{M}_{p_1,\ldots,p_2}$, where the superscript R describes the components in the coordinate system rotated by $R \in SO(d)$. This introduces the very important concepts of invariance and covariance [8, 12].

Definition 4. *The homogeneous polynomial $q(\mathbf{x})$ is invariant under the operation of the group G if $q(g\mathbf{x}) = q(\mathbf{x})$ for all $g \in G$.*

Example 2. The elementary symmetric polynomials $\sigma_k(\mathbf{x})$ are

$$\begin{aligned}
\sigma_0(\mathbf{x}) &= 1 \\
\sigma_1(\mathbf{x}) &= x_1 + \cdots + x_d \\
\sigma_2(\mathbf{x}) &= x_1x_2 + x_2x_3 \cdots + x_{d-1}x_d \\
&\vdots \\
\sigma_d(\mathbf{x}) &= x_1x_2 \cdots x_d
\end{aligned}$$

these are the coefficients in $\Pi_{l=1}^d(x_l + t) = \sum_{k=0}^d \sigma_{d-k}(\mathbf{x})t^k$.

The previous example would be just that, were it not for the following fundamental theorem (e.g. [12], Theorem 4.23):

Theorem 1. *For any symmetric polynomial $q_{\mathbb{A}}(\mathbf{x})$ there is a polynomial P such that*

$$q_{\mathbb{A}}(\mathbf{x}) = P(\sigma_1(\mathbf{x}), \cdots, \sigma_d(\mathbf{x})). \tag{9}$$

This theorem is interesting in that it tells us that any symmetric tensor can be written as some simple function of fundamental ones.

Example 3. A simple illustration uses again the Euclidean norm: $\sigma_1(\mathbf{x})^2 = \sum_{i=1}^d x_i^2 + \sum_{i\neq j} x_ix_j = |\mathbf{x}|^2 + 2\sigma_2(\mathbf{x})$; thus

$$|\mathbf{x}|^2 = \sigma_1(\mathbf{x})^2 - 2\sigma_2(\mathbf{x}).$$

In tensor component terms, $A_{\sigma_1} = [1; \ldots; 1] =: \mathbf{1}^t$, thus $A_{\sigma_1} \otimes A_{\sigma_1} = \mathbf{1}\mathbf{1}^t$. This relation expresses that the identity matrix is the matrix of ones minus the matrix of ones in the diagonal, showing once again that the multiple points of view can shed a different light on relationships.

2.2 Spherical Harmonics

The space of all polynomials splits into subspaces of homogeneous polynomials of degree p: $\mathcal{P} = \oplus_p \mathcal{P}_p$. The Laplace operator in spherical coordinates is $\Delta_{\mathbb{R}^d} = \frac{1}{r^2}\partial_r(r^2\partial_r) + \frac{1}{r^2}\Delta_{S^{d-1}}$. Thus, $\Delta_{\mathbb{R}^d}|_{\mathcal{P}_p} = \frac{1}{r^2}\left(p(p+1) + \Delta_{S^{d-1}}\right)$. It follows that homogeneous harmonic polynomials are eigenfunctions of the

Laplace–Beltrami operator on the sphere. These functions are called *spherical harmonics*, and this space is denoted $\mathcal{H}_p$. It is then natural to subdivide further $\mathcal{P}_p$: $\mathcal{P}_p = \mathcal{H}_p \oplus \mathcal{H}_p^{\perp}$. Harmonic polynomials correspond to traceless tensors. This is a simple form of the important *Hodge-de Rahm* decomposition.

Example 4. The second order homogeneous polynomial $\mathbf{x}^t A\mathbf{x}$ splits into

$$\mathbb{A}_0|\mathbf{x}^2| + \mathbf{x}^t(\mathbb{A} - \mathbb{A}_0 I)\mathbf{x}.$$

The second term is the spherical harmonic of order 2. Here $\mathbb{A}_0$ is the 0th order term $(\text{trace}(\mathbb{A})/d)$ corresponding to the constant function.

2.2.1 Anisotropies

In general, we know that the Trace is a symmetric tensor of order $p-2$, thus we can write any symmetric tensor as $\mathbb{A} = \text{tr}(()\mathbb{A}) \otimes_S I + (\mathbb{A} - \text{tr}(()\mathbb{A}) \otimes_S I)$. The second term is then defined as the *pure order p* traceless part, denoted $\mathbb{A}_p$. The process can be repeated, for example

$$\mathbb{A} = \mathbb{A}_0 + \mathbb{A}_2 + \cdots \mathbb{A}_p \tag{10}$$

In this way we have defined a decomposition of the tensors without introducing explicit spherical harmonic functions! By analogy with the example, we define a lth order *anisotropy* $\text{FA}_l(\mathbb{A}) := \|\mathbb{A}_l\|/\|\mathbb{A}\|$.

2.2.2 Non-Standard Metrics

Polynomials are a compact way to represent fully symmetric tensors and thus to manipulate them. As an example, in the previous Dagstuhl book, the Kullback–Leibler divergence was introduced as a function on distributions. This function can be generalised to the general case. Given a symmetric tensor $\mathbb{A}$, we have defined the associated polynomial $q_{\mathbb{A}}(x)$. The concept of positive definite tensors is easily defined from the polynomial version: a second order tensor $\mathbb{D}$ is positive-definite if and only if $q_{\mathbb{D}}(x) > 0$. We extend this definition to tensors of any even order, and call such tensors for short *positive*. We use the notation $\mathbb{D}$ for positive tensors because of the motivation from Diffusion Tensor MRI.

Definition 5. *Given a semi-positive-definite tensor $\mathbb{D}$, we define the function $g_{\mathbb{D}}(\mathbf{x}) = e^{-q_{\mathbb{D}}(\mathbf{x})}/\int_{\mathbb{R}^d} e^{-q_{\mathbb{D}}(\mathbf{x})} =: e^{-q_{\mathbb{D}}(\mathbf{x})}/c_{\mathbb{D}}$. This makes $g_{\mathbb{D}}$ into a probability distribution.*

In [11] the Kullback–Leibler divergence and the symmetrised Kullback–Leibler divergence are defined as

$$\text{KL}(\mathbb{D}_1, \mathbb{D}_2) = \int g_{\mathbb{D}_1} \log(g_{\mathbb{D}_1}/g_{\mathbb{D}_2})$$

$$\text{KL}_s(\mathbb{D}_1, \mathbb{D}_2) = \frac{1}{2}(\text{KL}(\mathbb{D}_1, \mathbb{D}_2) + \text{KL}(\mathbb{D}_2, \mathbb{D}_1)).$$

We can extend these measures without modification to positive definite tensors of any order.

3 Signals and Directional Signals

Directions are represented by lines in space, or pairs of antipodal points $\pm\mathbf{v}$ on a sphere. The space of directions is the real *projective* space $P\mathbb{R}^{d-1}$, a $d-1$-dimensional space.

Definition 6. *A* directional signal $s(\pm\mathbf{u})$ *is a function on* $P\mathbb{R}^{d-1}$.

The interest of tensors is that the usual Taylor expansion provides an approximation of signals in terms of homogeneous symmetric polynomials. We try to keep the discussion general for other potential applications, but our main motivation for directional signals was from diffusion tensor MRI (DT-MRI). In this field, by varying magnetic field gradient directions, we acquire signals depending on this direction. Assuming a specific model, it is then possible to reconstruct parameters from the measurements. The standard model for DT-MRI is based on the Stejskal–Tanner equation, where it is hypothesised that the signal depends on direction $\mathbf{e}$ as $s(\mathbf{e}) = s_0 e^{-b\mathbf{e}^t\mathbf{d}\mathbf{e}}$, where b is a known weighting factor dependent on diffusion times and gradient strength, and s_0 a non-weighted image. The unknown is then the second order tensor $\mathbb{D}$.

By restricting to directions, we ignore the norms of sampling positions. We can re-introduce by working on spheres of different radii. In Diffusion MRI, this amounts to varying the b-value. From this point of view, we are working in spherical coordinates. Homogeneous polynomials are particularly suited to this.

3.1 Directional Sampling

In a practical situation, we often have to assume an underlying model. To be specific, let us define models:

Definition 7. *A* model *is a parametrization of directional signals* $s(\mathbf{e};\theta)$.

Here we used the standard notation for parameters in parameter estimation theory. We restrict ourselves to parameters in a tensor space. We could call such models generically *tensorial* models, in which case we write $s(\mathbf{e};\mathbb{A}_1,\ldots,\mathbb{A}_n)$, where $\mathbb{A}_l$ is a tensor of any order. The class of models is restricted even further by the condition that $s(\mathbf{e};\mathbb{D}_1,...,\mathbb{D}_n)$ is a linear combination of signals of the form $g_{\mathbb{D}_l}$. Thus we assume from now on that the tensors are positive.

Definition 8. *We define models of this form as HARDI models, in connection with High Angular Resolution Diffusion Imaging. Thus a HARDI model takes the form*

$$s(\mathbf{e};\mathbb{D}_1,\ldots,\mathbb{D}_n) = \sum f_l g_{\mathbb{D}_l}(\mathbf{e}).$$

The choice of model is the choice of function s, and by sampling in the space of directions, we get a parameter estimation $\hat{\mathbb{D}}$, which ideally is as independent as possible on the choice of sampling directions $\mathbf{e}_1, \ldots, \mathbf{e}_N$, although for a finite set of sampling directions it is unavoidable that it still does depend on it: $\hat{\mathbb{D}} = \hat{\mathbb{D}}(\mathbf{e}_1, \ldots, \mathbf{e}_N)$.

It is common to expand (the log of) directional signals in terms of even order spherical harmonics Y_l

$$\log s(\mathbf{e}; \mathbb{D}_1, \ldots, \mathbb{D}_n) = \sum_{l=0}^{p} s_l(\mathbb{D}_1, \ldots, \mathbb{D}_n) Y_l(\mathbf{e}).$$

Note that a Least-Squares fit assumes a normal distribution of errors. This is *not* preserved by the logarithm, this is discussed in the DT-MRI literature, and in a chapter of this book. Here, we are interested in the functional (model) dependencies, thus from the point of view here, the logarithm is just a bijective data transformation. We write $\tilde{\mathbf{s}}$ for the array of logarithms of samples of the data at $\mathbf{e}_1, \ldots, \mathbf{e}_N$.

3.2 HARDI Methods

Among the multitude of methods in HARDI, we mention Spherical Deconvolution [17], FORECAST [2], Q-Ball [9, 10, 18], and the Diffusion Orientation Transform [13]. Most have in common a two-step approach based on spherical harmonics:

- Step 1: Solve $\tilde{\mathbf{s}} = Y\mathbf{c}$, to get the array $[c_k]_k = [c_{l_k,m_k}]_k$.
- Step 2: Apply some operation on $\mathbf{c}$ to extract the desired information.

They will be considered in more details in the following comparisons.

3.3 Linear to Linear Models

Some of the models are linear, in the sense that the relationship between the measured data $\tilde{\mathbf{s}}$ and the generic model parameters $\boldsymbol{\theta}$ is given by a matrix equation $\tilde{\mathbf{s}} = \mathcal{M}\boldsymbol{\theta}$. Suppose now that we have two potential linear models described by matrices $\mathcal{M}_1$ and $\mathcal{M}_2$. Assuming that model 1 describes the true underlying process but that we use model 2, we can directly get the relationship between the estimated parameters, because they must explain the same data:

$$\mathcal{M}_1\boldsymbol{\theta}_1 = s = \mathcal{M}_2\boldsymbol{\theta}_2$$

from which follows that $\boldsymbol{\theta}_2 = \mathcal{M}_2^+\mathcal{M}_1\boldsymbol{\theta}_1 =: \mathcal{M}_{21}\boldsymbol{\theta}_1$, where the last equality is a definition. In the middle equality we used the *pseudo-inverse* $\mathcal{M}_2^+$. This means

$$\mathcal{M}_{21} = (\mathcal{M}_2^t\mathcal{M}_2)^{-1}(\mathcal{M}_2^t\mathcal{M}_1);$$

we call this matrix the *model-to-model* matrix. We assume that the number of parameters in model 2 is larger or equal than in model 1.

3.3.1 Polynomial to Polynomial, i.e. Tensor to Tensor

The standard diffusion model leads, after the logarithm operation, to a second order polynomial model. The model matrix ('direction encoding') contains in its rows all second order monomials sampled on the chosen gradient directions $\mathbf{e}_n$. Generally, we write E_p for the matrix containing all monomials of order p, in a specific lexicographic order. We note that $E_p^t E_q$ contains sums of monomials of order $p+q$, over the directions. We thus have by definition $E_p^t E_q/(p+q) =: M_{p+q}$, the matrix of order $p+q$ moments. It follows that the model-to-model matrix of polynomial models of different orders is a 'ratio' of matrices of moments. Its components are rational functions.

Let us compute the model-to-model matrix when we fit an order 2 polynomial function to what is truly an order 4 data. To simplify, we work in dimension $d = 2$, with conventions

$$E_2 = [x_i^2, y_i^2, 2x_i y_i]; \quad E_4 = [x_i^4, y_i^4, 4x_i^3 y_i, 4x_i y_i^3, 6x_i^2 y_i^2].$$

Note that these matrices, thus the model-to-model matrix, depend on the direction encoding. We assume that directions are uniformly distributed, and take the limit when the number of directions goes to infinity. This describes an ideal sampling, and in this way we get a model-to-model description, which is independent of the choice of directions. Let D_{xx}, D_{yy}, D_{xy} represent the components of the order 2 tensor, and $D_{xxxx}, D_{yyyy}, D_{xxxy}, D_{yyyx}, D_{xxyy}$ the components of the order 4 tensor. We get

$$\begin{bmatrix} D_{xx} \\ D_{yy} \\ D_{xy} \end{bmatrix} = \frac{1}{8} \begin{bmatrix} 7 & -1 & 0 & 0 & 6 \\ -1 & 7 & 0 & 0 & 6 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} D_{xxxx} \\ D_{yyyy} \\ D_{xxxy} \\ D_{yyyx} \\ D_{xxyy} \end{bmatrix},$$

thus $D_{ii} = \frac{1}{8}\left(7D_{iiii} - D_{jjjj} + 6D_{iijj}\right)$ and $D_{ij} = \frac{1}{8}(D_{iiij} + D_{jjji})$ for $i \neq j$. A similarly compact result can be obtained in three dimensions:

$$D_{ii} = \frac{3}{35}\left(9D_{iiii} - D_{jjjj} - D_{kkkk}\right) + \frac{3}{35}\left(8D_{iijj} + 8D_{iikk} - 2D_{jjkk}\right) \quad (11)$$

$$D_{ij} = \frac{30}{35}\left(D_{iiij} + D_{jjji} + D_{ijkk}\right), \quad (12)$$

where i, j, k run through x, y, z and are all different. This tells us what second order symmetric tensor is estimated if the true underlying process corresponds to a fourth order symmetric tensor. These results are consistent with Moakher's results (see chapter by M. Moakher in this book, and Özarslan [14]).

The aim of HARDI methods is to resolve multiple principal directions in a voxel. What the methods have in common is the input data, but may vary

slightly on how they express their output. For a full comparison, a common output is thus required. We have at least three possibilities:

- The output is the function $p(\mathbf{r})$, the probability displacement function. As it is a general function on 3D space, it would be expressed as linear combination of basis functions.
- The output is the function $odf(\mathbf{e})$, the orientation distribution function. It is a function on the space of orientation, and would be expressed as a linear combination of spherical harmonics.
- The output is a finite set of orientations $\mathbf{e}_i$.

As mentioned earlier, most methods contain two steps, the first one, to express the data as a linear combination of spherical harmonics, is common to Q-ball, FORECAST and Spherical Deconvolutions. The second step is to compute the ODF from $\mathbf{c}$. Here the methods differ, but again most of them are linear, in that they express the ODF as a linear combination of spherical harmonics.

3.4 Linear to Non-Linear Models

3.4.1 Specific HARDI Models

3.4.1.1 Spherical Deconvolution

Here the array $\mathbf{c}$ is used to extract the parameters of a response function. The signal is then a linear superposition of these responses, thus the methods are in theory related to multi-exponential fitting, although in practice they are likely to be more robust than the notoriously unstable direct non-linear fitting. In any case, the model describes the effect of the response function as a spherical convolution, $\mathbf{c} = \mathbf{R}\mathbf{o}_S$, thus $\boldsymbol{\theta}_S = \mathbf{O}_S\mathbf{c} = (\mathbf{R}^t\mathbf{R})^{-1}\mathbf{R}^t\mathbf{c}$ (Eq. (3) in [17]).

3.4.1.2 FORECAST

Here again, the coefficients c_{lm} are just scaled, thus $\mathbf{c} = \mathbf{O}_F\boldsymbol{\theta}_F$ (Eq. (15) in [2]). This time, though, the scaling factor depends on a specific assumption about the shape of the underlying tensors. Thus strictly speaking, the method is not linear, as $\mathbf{O}_F$ depends on the ratio of longitudinal to perpendicular diffusivity, thus on the anisotropy. Equations (21) and (27) in [2] give directly the connection with the q-ball ODF.

3.4.1.3 Higher Order Tensors (HOT)

Özarslan et al. also point out the relationship between spherical harmonics based methods and higher order tensors [13]. As they have introduced a more recent method (see next paragraph), we will not dwell on this here.

3.4.1.4 Diffusion Orientation Transform (DOT)

The diffusion orientation transform (DOT) assumes the same directional response function, an exponential of a second order tensor. The Fourier transform expressing the PDF as a function of the signal is computed in spherical coordinates; once the radial part has been integrated, the resulting term is expressed in terms of spherical harmonics [15]. They do not compute the ODF, but the PDF directly. Thus to express explicitly the connection, we have to integrate Equations (14) or Equations (9) of [15] over r.

A generic function $m(\boldsymbol{\theta})$ of the parameters defines a non-linear model. The estimate of parameters from a linear model would then be formally

$$\boldsymbol{\theta}_{\text{linear}} = M^{+} m(\boldsymbol{\theta}_{\text{nonlinear}}).$$

For a general non-linear to non-linear situation, we do not have access to the pseudo-inverse. A direct linearisation of the left-hand side gives

$$m_1(\boldsymbol{\theta}_1) = m_2(\boldsymbol{\theta}_2) \Rightarrow \boldsymbol{\theta}_1 = (Dm_1)^{+}(0) m_2(\boldsymbol{\theta}_2),$$

which could be applied in particular to the non-linear to linear situation. Here we made the assumption that $m_1(0) = 0$. This could be extended to better non-linear least squares solution by using expansions to higher orders, but it is only of formal interest for abstract non-linear functions. In HARDI, the typical non-linear model is multi-exponential, and we are able to give a much more precise characterization of the relationship between linear models, that is tensors of *any* order, and the multi-exponential.

3.5 Non-Linear to Linear Models

A standard DT-MRI signal is described as a single exponential $s(b) = s_0 e^{-b\mathbf{e}^t D \mathbf{e}}$. Although this affects noise distribution, it is useful to work with the logarithm of this signal, as this is then simply a quadratic form, and linear methods can be used. The constant term s_0 is irrelevant here, and absorbed into s. In practice, however, the signal is often a *mixture* $s(b) = f_1 e^{-b\mathbf{e}^t D_1 \mathbf{e}} + f_2 e^{-b\mathbf{e}^t D_2 \mathbf{e}}$, where $f_1 + f_2 = 1$. Here we give a description of the logarithm of such a multi-exponential in terms of tensors. The parameter is the value b, which controls the strength of the diffusion. The result in words is, to first order, the logarithm of the signal is the same as what one would get with a single exponential, but where the tensor is the weighted average $f_1 D_1 + f_2 D_2$. Furthermore, *all* higher order tensorial terms are just the difference of the tensor to the corresponding tensorial power, with coefficients that are explicitly simple polynomials in f_1, f_2. Note that in some sense, this means that all the information is contained in the first and second order term. Here we are trying to be more specific than [4, 16].

Theorem 2. *When* $f_1 e^{-b\xi_1} \neq f_2 e^{-b\xi_2}$,

$$\log(f_1 e^{-b\xi_1} + f_2 e^{-b\xi_2}) = -b(f_1\xi_1 + f_2\xi_2) + \sum_{k\geq 2} c_k(f_1, f_2) b^k (\xi_1 - \xi_2)^k, \quad (13)$$

where c_k *is an explicit polynomial in* f_1, f_2.

The table of some of the first c_k *is*

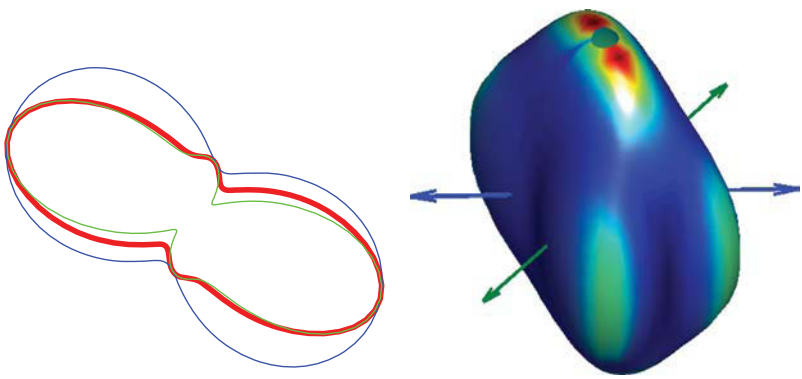
$$\begin{aligned}
c_2 &= f_1 f_2 \\
c_3 &= f_1 f_2 (f_1 - f_2) \\
c_4 &= f_1 f_2 (f_1^2 - 4 f_1 f_2 + f_2^2) \\
c_5 &= f_1 f_2 (f_1 - f_2)(f_1^2 - 10 f_1 f_2 + f_2^2) \\
c_6 &= f_1 f_2 (f_1^4 - 26 f_2 f_1^3 + 66 f_1^2 f_2^2 - 26 f_1 f_2^3 + f_2^4) \\
c_7 &= f_1 f_2 (f_1 - f_2)(f_1^4 - 56 f_1 f_2^3 + 246 f_1^2 f_2^2 - 56 f_2 f_1^3 + f_2^4) \\
&\vdots
\end{aligned}$$

In other words, for a bi-exponential, the second order tensor estimated from a multi-exponential is the weighted average of the tensors (Fig. 1):

$$\text{2nd order} \sim f_1 D_1 + f_2 D_2 \quad (14)$$

$$\text{4th order} \sim f_1 f_2 (D_1 - D_2) \cdot (D_1 - D_2), \quad (15)$$

where the dot represents the symmetric tensor product.



(a) (b)

Fig. 1. (**a**) Multi-exponential 'peanut' contours with order 2 and order 4 term as in Theorem 2; (**b**) A modified 'peanut shape from a multi-exponential crossing of pure directions (i.e. a bi-exponential where the second-order tensors have eigenvalues in ratio [1:0:0]. Here, instead of representing $s(\mathbf{e})$ as the radius in direction $\mathbf{e}$, we use $1/\sqrt{\log(s(\mathbf{e}))}$. The advantage of this representation is that the shape for a single second-order tensor is an ellipsoid

We can attempt to use a Taylor expansion. Then we need an explicit expression of a logarithmic derivative. Another option is using formal power series algebra. Suppose $u(x) = \sum_k a_k x^k$ is some real analytic function. In the following, we use the convention that a sum symbol $\sum_k = \sum_{k=0}^{\infty}$. To make the argument clearer, we use some lemmas. We want to expand the logarithm of a sum of exponentials. The sum of exponentials can be written as a power series, and the logarithm expanded will then be a sum of sums. The first lemma essentially describes how we write such sums of sums. The second lemma uses the first one to write the logarithm of a power series as a power series. The third lemma gives a closed form formula for some of the explicit components.

Lemma 1. *The pth power of u is also real analytic, with same radius of convergence, and*

$$u(x)^p = (\sum_k a_k x^k)^p = \sum_l a_l^{(p)} x^l,$$

where $a_l^{(p)} = \sum_{\{k_i\}} a_{k_1} \cdots a_{k_p}$.

Here the k_is are taken over all partitions of l in p integers. The number of such partitions is the multinomial coefficient $\binom{l}{k_1 \cdots k_p}$. This lemma follows directly from applying the multinomial formula.

The next lemma is analytically a bit trickier. We know the power expansion of $\log(1+x)$ for $|x| < 1$. We would like to generalise this by replacing x by an analytic function:

Lemma 2. *Suppose* $|u(x)| < 1$ *for all x inside the convergence radius, and the series* $b_l = \sum_{p \geq 1} \frac{1}{p}(-1)^{p-1} a_l^{(p)}$ *are absolutely convergent.*
Then $\log(1 + u(x)) = \sum_l b_l x^l$.

Proof. Because $|u(x)| < 1$, we can write $\log(1+u(x)) = \sum_p \frac{1}{p}(-1)^{p-1}(u(x))^p$. The previous lemma provides an expression for $u(x)^p$. We need to reorder the sums. This is possible thanks to the assumption of absolute convergence, thus $\sum_p \frac{1}{p} \sum_l (-1)^{p-1} a_l^{(p)} x^l = \sum_l x^l \sum_p \frac{1}{p}(-1)^{p-1} a_l^{(p)}$.

Lemma 3. *For* $|r| < 1$*:* $\sum_p p^l r^p = (r\frac{d}{dr})^{(l)} \frac{1}{1-r}$.

Proof. $r\frac{d}{dr}\frac{1}{1-r} = \sum_p p r^p$ thus $r\frac{d}{dr}(r\frac{d}{dr}\frac{1}{1-r}) = \sum_p p^2 r^p$. Iterating the derivatives l times gives the result.

Proof (Theorem). Assume that $f_1 e^{-b\xi_1} > f_2 e^{-b\xi_2}$. Then $\log(f_1 e^{-b\xi_1} + f_2 e^{-b\xi_2}) = \log(f_1 e^{-b\xi_1}) + \log(1 + f_2/f_1 e^{-b(\xi_2 - \xi_1)})$. Apply lemma 2 to $u(b) = f_2/f_1 e^{-b(\xi_2-\xi_1)})$, then

$$
\begin{aligned}
b_l &= -\sum_p \sum_{\{k_i\}} \frac{1}{p} \left(-\frac{f_2}{f_1}\right)^p \frac{1}{k_1! \cdots k_p!} \\
&= -\frac{1}{l!} \sum_p \frac{1}{p} \left(-\frac{f_2}{f_1}\right)^p \sum_{\{k_i\}} \binom{l}{k_1 \cdots k_p} \\
&= -\frac{1}{l!} \sum_p \left(-\frac{f_2}{f_1}\right)^p \frac{1}{p} (1 + \cdots + 1)^l = -\frac{1}{l!} \sum_p (-\frac{f_2}{f_1})^p p^{l-1}.
\end{aligned}
$$

Thus

$$
\log(s(b)) = \log f_1 - b\xi_1 - \log(1/(1 + f_2/f_1)) + b(\xi_1 - \xi_2)\frac{f_2/f_1}{1 + f_2/f_1} + \cdots
$$

4 Summary and Conclusions

We have summarised the connection between symmetrics tensors and polynomials, and explicited the correspondence of different formulas. For example, the trace of a tensor corresponds to integrals of polynomials over the sphere, and the same holds for inner products of tensors. Furthermore, traceless polynomials correspond to spherical harmonics, giving a third representation.

We then show how this correspondence can be used to compare different models of directional signals. The example comes from DT-MRI, where we can write conversion formula between polynomial models of different order. We demonstrate this for tensors of order 4 represented as tensors of order 2.

Finally, we establish which tensors represent the log of a sum of exponentials. Such models are also common in DT-MRI. Once again, the formulas turn out to be fairly compact, and we can conclude that if we use a Stejskal–Tanner model where the truth is a sum of exponential terms, we would measure the weighted average of the individual tensors, weighted by volume fractions. If we fit higher order terms, we would measure tensor products of the differences.

In conclusion, although many models are used to describe directional signals, it is possible to describe what we would measure assuming a specific model.

Acknowledgments

I thank the reviewers and M. Moakher for pointing and correcting numerous mistakes in the first version.

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York, ninth dover printing, tenth gpo printing edition, 1964.

[2] A. Anderson. Measurement of fiber orientation distributions using high angular resolution diffusion imaging. *Magn. Reson. in Med.*, 54:1194–1206, 2005.
[3] J. A. Baker. Integration over spheres and divergence theorems for balls. *The American Mathematical Monthly*, 104(1):36–47, 1997.
[4] P. J. Basser. Relationships between diffusion tensor and q-space MRI. *Magn. Reson. in Med.*, 47:392–397, 2002.
[5] P. G. Batchelor, D. Atkinson, D. L. G. Hill, F. Calamante, and A. Connelly. Anisotropic noise propagation in diffusion tensor MRI sampling schemes. *Magn. Reson. in Med.*, 49:1143–1151, 2003.
[6] G. E. P. Box and J. S. Hunter. Multi-factor experimental designs for exploring response surfaces. *The Annals of Mathematical Statistics*, 28(1): 195–241, 1957.
[7] D. A. Cox, J. Little, and D. O'Shea. *Using Algebraic Geometry*. Springer, 2004.
[8] Y. I. Dimitrienko. *Tensor Analysis and Nonlinear Tensor Functions*. Kluwer Academic Publishers, 2002.
[9] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche. Apparent diffusion coefficients from high angular resolution diffusion imaging: Estimation and applications. *Magn. Reson. in Med.*, 56:395–410, 2006.
[10] C. P. Hess, P. Mukherjee, E. T. Eric T. Han, D. Xu, and D. B. Vigneron. Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. *Magn. Reson. in Med.*, 56:104–117, 2006.
[11] M. Moakher and P. G. Batchelor. *Symmetric Positive Definite Matrices: From Geometry to Applications and Visualization*, volume Visualization and Processing of Tensor Fields. of *Mathematics+Visualization*. Springer, 2006.
[12] P. J. Olver. *Classical Invariant Theory*, volume 44 of *Students Texts*. London Mathematical Society, 1999.
[13] E. Özarslan and T. H. Mareci. Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. *Magn. Reson. in Med.*, 50:955–965, 2004.
[14] E. Özarslan, B. C. Vemuri, and T. H. Mareci. Generalized scalar measures for diffusion MRI using trace, variance, and entropy. *Magn. Reson. in Med.*, 53:866–876, 2005.
[15] E. Özarslan, T. M. Shepherd, and B. C. Vemuri, S. J. Blackband, and T. H. Mareci. Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT), *NeuroImage*, 31:1086–1103, 2006.
[16] J. Stepišnik. Validity limits of Gaussian approximation in cumulants expansion for diffusion attenuation of spin echo. *Physica B*, 270:110–117, 1999.

[17] J.-D. Tournier, F. Calamante, D. G. Gadian, and A. Connelly. Direct estimation of the fibre orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, 23:1176–1185, 2004.

[18] D. S. Tuch. Q-ball imaging. *Magn. Reson. in Med.*, 52:1358–1372, 2004.

A Mixture of Wisharts (MOW) Model for Multifiber Reconstruction

Bing Jian[1], Baba C. Vemuri[1], and Evren Özarslan[2]

[1] Department of Computer and Information Science and Engineering, University of Florida, P.O. Box 116120, Gainesville, FL 32611, USA
{bjian,vemuri@}@cise.ufl.edu

[2] Section on Tissue Biophysics and Biomimetics, LIMB, NICHD, National Institutes of Health, Bethesda, MD 20892, USA
evren@helix.nih.gov

Summary. Diffusion-weighted magnetic resonance imaging (DW-MRI) is a non-invasive imaging technique that allows neural tissue architecture to be probed at a microscopic scale in vivo. By producing quantitative data on the motion of water molecules that naturally occurs in brain tissues as part of the physical diffusion process, DW-MRI can be used to map the fiber paths in the brain white matter. Estimating the local complex architecture in the presence of intra-voxel heterogeneity caused by multiple fiber pathways is of great importance and requires sophisticated modeling techniques.

In this chapter, we present a review of our recently introduced novel mathematical model [27] and an accompanying unified computational framework for spherical deconvolution to perform multi-fiber reconstruction [28]. In our model, the diffusion-weighted MR signal attenuation is characterized by the Laplace transform of a continuous mixture of diffusion tensors. When the mixing distribution is a Wishart distribution, the Laplace transform yields a closed form expression. Additionally, we show that the traditional diffusion tensor model is a limiting case of this continuous mixture of tensors model. We then formulate the reconstruction problem in a unified deconvolution framework that facilitates investigation of several deconvolution schemes and achieves stable, sparse and accurate solutions. Finally, we present results of testing this theoretical model and the accompanying unified computational framework, on synthetic data and real rat brain data. The comparisons with several competing methods empirically suggest that the proposed model yields efficient and accurate solutions in the presence of intra-voxel orientational heterogeneity.

1 Introduction

By providing quantitative measures that depict various characteristics of diffusional motion of water molecules, diffusion-weighted magnetic resonance imaging (DW-MRI) allows neural tissue architecture to be probed at a microscopic scale, noninvasively and in vivo. The information extracted from DW-MRI data can be used to estimate the orientation of fibers in the central

nervous system, which can be further exploited for neuronal connectivity inference. Since the neuronal connectivity and tissue microstructure may be altered by various physiological and pathological processes, DW-MRI has also become a valuable imaging tool in studies of brain development and diseases [1].

Diffusion tensor MRI (DT-MRI or DTI), introduced in [2], provides a relatively simple way of quantifying diffusion anisotropy as well as extracting fiber directions locally from multidirectional diffusion MRI data. For recent review articles on DTI, see [3–5]. Although promising results have been achieved using DTI to study regions of the brain and spinal cord with significant white-matter coherence and to map anatomical connections in the central nervous system [6–9], this model has two significant weaknesses: it can reveal only one fiber orientation in each voxel and it fails in regions containing intra-voxel orientational heterogeneity (IVOH) [10–13].

This limitation of the diffusion tensor model has stimulated exploration of both more demanding image acquisition strategies and more sophisticated reconstruction methods. Methods of representing the apparent diffusivity profiles include spherical harmonics expansion [10] and the equivalent higher order tensor model [14]. Based on the observation that the peaks of the diffusivity profile do not necessarily yield the orientations of the distinct fiber populations, a number of model-independent approaches have been developed, which attempt to transform the multidirectional diffusion-weighted MR signals into a function that describes the probability of water molecule displacement. The Q-ball imaging (QBI) method approximates the radial integral of the displacement probability distribution function (PDF) by the spherical Funk–Radon transform [15]. More recent studies have analyzed QBI's Funk–Radon transform using a spherical harmonic basis [16–18]. Diffusion spectrum imaging (DSI) can be used to measure the microscopic diffusion function directly based on the Fourier relation between the diffusion signal and the diffusion function, but its time-intensive q-space sampling burden makes this use impractical [19]. The diffusion orientation transform (DOT) transforms the diffusivity profiles into probability profiles by explicitly expressing the Fourier relation in spherical coordinates and evaluating the radial part of the integral analytically [20].

Aiming to describe the non-monoexponential MR signal decay, a variety of multicompartmental models or multiple-fiber population models using a finite mixture of Gaussians have been formulated [13, 21–23]. Assuming a distribution, rather than a discrete number, of fiber orientations, the spherical deconvolution method [24] can be treated as a continuous extension of the finite discrete mixture model. Compared to the multicompartmental models, the spherical deconvolution framework offers two significant advantages. First, it avoids specifying the number of underlying fiber populations, which is an intrinsic model selection problem in the multiple-compartment models. Second, the spherical deconvolution can often be reformulated as a linear system and then efficiently solved, whereas the multicompartment models usually involve nonlinear optimization, which is more computationally expensive. Prompted by these merits of the spherical deconvolution framework, recently a number

of variants of spherical deconvolution approaches [16, 25, 26] have been reported with different choices of basis functions, deconvolution kernels, and regularization schemes.

In this chapter, we present a review of our recently introduced novel statistical model [27] and an accompanying unified computational framework for spherical deconvolution to perform multifiber reconstruction [28]. The novel statistical model is a generalization of the diffusion tensor model [2] and additionally has the power to resolve multiple fiber orientations within a voxel. First, it is postulated that each voxel is associated with an underlying probability distribution defined on the space of diffusion tensors (the manifold of 3×3 positive-definite matrices). Conceptually, one can view this as a natural extension of the multiple-compartment models [13, 23]. Additionally, this continuous mixture model can express the MR signal attenuation as a Laplace transform defined for matrix-variate functions. For a class of parameterized distributions, namely, the Wishart distributions, the Laplace transform can be evaluated in closed form, leading to a Rigaut-type function, which has been used in previously published literature [29] to explain the MR signal decay phenomenologically. A deconvolution framework for the multi-fiber reconstruction problem was seamlessly formulated from this continuous mixture model, where the deconvolution kernel is the Laplace transform kernel and the basis functions are the Wishart distributions [28]. This model was then put into a unified deconvolution framework to further investigate several deconvolution schemes designed to achieve stable, sparse, and accurate solutions. Finally, we present results of testing this theoretical model and the accompanying unified computational framework on synthetic data and real rat brain data. The comparisons with several competing methods empirically suggest that the proposed model combined with a nonnegative least squares (NNLS) deconvolution method yields efficient and accurate solutions for the multifiber reconstruction problem in the presence of intra-voxel orientational heterogeneity.

2 A Mixture of Wisharts Statistical Model

The key assumption in our continuous mixture model is that at each voxel there is an underlying probability measure defined on the manifold of $n \times n$ symmetric positive-definite (SPD) matrices, denoted by $\mathcal{P}_n$ (by default $n = 3$). Let $f(\mathsf{D})$ be the density function of the assumed probability measure with respect to some carrier measure $\mathrm{d}\mathsf{D}$ on $\mathcal{P}_n$, then the diffusion weighted MR signal $S(\boldsymbol{q})$ can be expressed as

$$S(\boldsymbol{q})/S_0 = \int_{\mathcal{P}_n} f(\mathsf{D}) \, \exp(-b\boldsymbol{g}^T \mathsf{D} \boldsymbol{g}) \, \mathrm{d}\mathsf{D}, \tag{1}$$

where S_0 is the signal in the absence of a diffusion weighting gradient, $\boldsymbol{q}$ encodes the magnitude (G) and direction ($\boldsymbol{g}$) of the diffusion sensitizing

gradients, and b is the diffusion weighting factor depending on the strength as well as the effective time of diffusion. Note that (1) implies a continuous form of mixture model, with $f(\mathsf{D})$ being a mixing density over the components in the mixture. Clearly, this model simplifies to the diffusion tensor model when the underlying probability measure is the Dirac measure.

Because $b\boldsymbol{g}^T\mathsf{D}\boldsymbol{g} = \text{trace}(\mathsf{BD})$, where $\mathsf{B} = b\boldsymbol{g}\boldsymbol{g}^T$, (1) can be expressed as the Laplace transform (in the matrix-valued variable case):

$$S(\boldsymbol{q})/S_0 = \int_{\mathcal{P}_n} \exp(-\text{trace}(\mathsf{BD}))\, f(\mathsf{D})\mathrm{d}\mathsf{D} = (\mathscr{L}_f)(\mathsf{B}), \tag{2}$$

where $\mathscr{L}_f$ denotes the Laplace transform of a function f, which takes its argument as symmetric positive definite matrices from $\mathcal{P}_n$. (For the formal definition of the Laplace transform on $\mathcal{P}_n$, see [30]).

Based on this model, an inverse problem naturally arises: recovering a distribution defined on $\mathcal{P}_n$ that best explains the observed diffusion signal $S(\boldsymbol{q})$. Unfortunately, this problem is ill-posed and, in general is intractable unless some prior knowledge about the probability function is used. Note that in DT-MRI, the diffusion tensor can be interpreted as the concentration matrix (inverse of the covariance matrix) of the Gaussian distribution in $\boldsymbol{q}$-space, and is usually estimated by solving a least squares problem, which is equivalent to maximum likelihood estimate when assuming the Gaussian noise-model on the log-transformed signals. Instead, our approach views the diffusion tensor as a random variable (matrix) belonging to some known distribution family, which allows us to model the uncertainty in the diffusion tensor estimation. In multivariate statistics, it is common practice to impose a Wishart distribution (see definition below) as a prior on the concentration matrix. As one of the most important probability distribution families for nonnegative-definite matrix-valued random variables ("random matrices"), the Wishart distribution [31] is most typically used when describing the covariance matrix of multinormal samples in multivariate statistics [32]. Moreover, in the case of a Wishart distribution, a closed form expression for the Laplace transform exists and leads to a Rigaut-type asymptotic fractal law for the MR signal decay behavior, which has been observed in the past (see explanation below). These factors motivate us to use the Wishart distribution as the mixing distribution $f(\mathsf{D})$.

In the following, we first briefly introduce the definition of a Wishart distribution as well as its relevant properties, then we analytically derive that when the mixing distribution in the proposed continuous mixture tensor model is a Wishart distribution, the Laplace transform leads to a Rigaut-type asymptotic fractal law for the MR signal decay.

Definition 1. *[30] For $\Sigma \in \mathcal{P}_n$ and for p in $\left(\frac{n-1}{2}, \infty\right)$, the Wishart distribution $\gamma_{p,\Sigma}$ with scale parameter Σ and shape parameter p is defined as*[1]

[1] Note that the correspondence between this definition and the conventional Wishart distribution $W_n(p, \Sigma)$ is given simply by $\gamma_{p/2,2\Sigma} = W_n(p, \Sigma)$.

$$\mathrm{d}\gamma_{p,\Sigma}(\mathsf{Y}) = \Gamma_n(p)^{-1}\,|\mathsf{Y}|^{p-(n+1)/2}\,|\Sigma|^{-p}\,e^{-\mathrm{trace}(\Sigma^{-1}\mathsf{Y})}\,\mathrm{d}\mathsf{Y}, \tag{3}$$

where Γ_n is the multivariate gamma function and $|\cdot|$ is the matrix determinant.

The Wishart distribution $\gamma_{p,\Sigma}$ is known to have the closed-form Laplace transform

$$\int e^{-\mathrm{trace}(\Theta\mathsf{Y})}\,\mathrm{d}\gamma_{p,\Sigma}(\mathsf{Y}) = (1+\mathrm{trace}(\Theta\Sigma))^{-p}, \tag{4}$$

where $\Theta + \Sigma^{-1} \in \mathcal{P}_n$. Consider the family of Wishart distributions $\gamma_{p,\Sigma}$ with a fixed expected value $\hat{\mathsf{D}} = p\Sigma$. When f in (2) is taken as a density function of this distribution family, the expression (2) takes the form

$$S(\boldsymbol{q}) = S_0\,(1+(b\,\boldsymbol{g}^T\hat{\mathsf{D}}\boldsymbol{g})/p)^{-p}\,. \tag{5}$$

This is a familiar Rigaut-type[2] asymptotic fractal expression [33], implying a signal decay characterized by a power-law, which is the expected asymptotic behavior for the MR signal attenuation in porous media [34]. Note that, although this form of a signal attenuation curve had been phenomenologically fitted to the diffusion-weighted MR data before [29], to the best of our knowledge, the proposed Wishart distribution model is the first rigorous derivation of the Rigaut-type expression that was used to explain the MR signal behavior as a function of b-value. Therefore, this derivation could be useful in understanding the apparent fractal-like behavior of the neural tissue in diffusion-weighted MR experiments. Also note when p tends to infinity, we have $S(\boldsymbol{q}) \to S_0\,\exp(-b\boldsymbol{g}^T\hat{\mathsf{D}}\boldsymbol{g})$, which implies that the mono-exponential model can be viewed as a limiting case ($p \to \infty$) of our model.

The density of a simple Wishart distribution as a function of diffusion tensor reaches one single diffusion maximum at its expected value; therefore, a single Wishart model can not resolve the intra-voxel orientational heterogeneity. Just as a discrete mixture of tensors model can be adapted, so can a discrete mixture of Wisharts model where the mixing distribution in (2) is given as $\mathrm{d}F = \sum_{i=1}^{N} w_i \mathrm{d}\gamma_{p_i,\Sigma_i}$. To make the problem tractable, the following simplifying assumptions are made. First, in this model the set of (p_i, Σ_i) is treated as a discretization of a submanifold of $\mathcal{P}_n$. Note that the number of components in the mixture, N, only depends on the resolution of this discretization and should not be interpreted as the expected number of fiber bundles. It is further assumed that all the p_i take the same value, $p_i = p = 2$, which is a reasonable assumption based on the analogy between (5) and the Debye–Porod law of diffraction [34] in 3D space. Since the fibers have an approximate axial symmetry, it also makes sense to assume that the two smaller eigenvalues of diffusion tensors are equal. In practice, the eigenvalues of $\mathsf{D}_i = p\Sigma_i$ are fixed to

[2] The phrase "Rigaut-type" is used to distinguish this function from Rigaut's own formula [33] function. Although slightly different, the Rigaut-type function shares many of the desirable properties of Rigaut's own function such as concavity and the asymptotic linearity in the log–log plots.

specified values $(\lambda_1, \lambda_2, \lambda_3) = (1.5, 0.4, 0.4)\mu^2\,\mathrm{m}^{-1}\mathrm{s}^{-1}$ consistent with the values commonly observed in white-matter tracts [13]. Because of this rotational symmetry, the discretization of $\mathcal{P}_n$ forming the mixture of Wisharts is reduced to a spherical tessellation. Accordingly, the prominent eigenvector of each Σ_i can be taken from the unit vectors uniformly distributed on the unit sphere. Because of the antipodal symmetry, the sampling is actually performed on the projective plane, that is, only half of a normal spherical tessellation is used. All the above assumptions leave us with the weights $\boldsymbol{w} = (w_i)$ as the unknowns to be estimated. For K measurements with $\boldsymbol{q}_j$, the signal model equation

$$S(\boldsymbol{q}) = S_0 \sum_{i=1}^{N} w_i (1 + \mathrm{trace}(\mathsf{B}\Sigma_i))^{-p} \tag{6}$$

leads to a linear system $\boldsymbol{A}\boldsymbol{w} = \boldsymbol{s}$, where $\boldsymbol{s} = (S(\boldsymbol{q})/S_0)$ contains the normalized measurements, $\boldsymbol{A}$ is the matrix with $\boldsymbol{A}_{ji} = (1 + \mathrm{trace}(\mathsf{B}_j\Sigma_i))^{-p}$, and $\boldsymbol{w} = (w_i)$ is the weight vector to be estimated.

Interestingly but not surprisingly, our method can be cast into a unified convolution framework that incorporates many existing reconstruction methods. Expressing the signal as the convolution of a probability density function and a kernel function, this unified convolution framework can be stated as follows:

$$S(\boldsymbol{q})/S_0 = \int_{\mathcal{M}} R(\boldsymbol{q}, x) f(x) \mathrm{d}x \tag{7}$$

In (7), the integration is over a manifold $\mathcal{M}$ whose individual points x contain the useful information like orientation and anisotropy. The convolution kernel $R(\boldsymbol{q}, x) : \mathbb{R}^3 \times \mathcal{M} \mapsto \mathbb{R}$ models how the signal receives response from a single fiber. To handle the intra-voxel orientational heterogeneity, the volume fractions represented by a continuous function $f(x) : \mathcal{M} \mapsto \mathbb{R}$ models the distribution of fiber bundles. Hence, the deconvolution problem is to estimate $f(x)$, given the specified $R(\boldsymbol{q}, x)$ and measurements $S(\boldsymbol{q})/S_0$. In practice, $f(x)$ is usually expressed as a linear combination of N basis functions: $f(x) = \sum w_j f_j(x)$. The choice of convolution kernels and basis functions often depend on the underlying manifold $\mathcal{M}$. A simple example is to let $\mathcal{M}$ be the unit sphere (or more precisely, the projective plane due to the antipodal symmetry), which leads to the spherical deconvolution problem [24, 25]. Several other approaches start from the manifold of diffusion tensors, but again reduce to the spherical deconvolution problem since only rotationally symmetric tensors are considered. [16, 23, 24].

Following [16, 23, 24], we choose the standard diffusion tensor kernel in our model. However, it is the Wishart basis function that distinguishes our method from these related methods. It is worth noting that the Wishart basis reduces to the Dirac function on $\mathcal{P}_n$ when $p = \infty$ and thus leads to the tensor basis function method [23] as well as to the FORECAST method [16], which both estimate fiber orientations using the continuous axially symmetric tensors and hence resemble our method very closely.

3 Stable, Sparse, and Positive Deconvolution

The general convolution problem (7) can be posed as the following linear system:

$$\boldsymbol{A}\boldsymbol{w} = \boldsymbol{s} + \eta, \tag{8}$$

where $\boldsymbol{s}$ contains K measurements $S(\boldsymbol{q})/S_0$, the $K \times N$ matrix $\boldsymbol{A} = \{A_{ij}\}$ is given by $A_{ij} = \int_{\mathcal{M}} R(\boldsymbol{q}_i, x) f_j(x) \mathrm{d}x$, and η represents a certain noise model. Note that the integral to compute the entries of $\boldsymbol{A}$ may have an analytical solution as in our model and others that use the tensor kernel [16, 24, 35], or may need to be numerically approximated as in [25]. In either case, once the response kernel $R(\boldsymbol{q}, x)$ and the basis function are specified, the matrix $\boldsymbol{A}$ can be fully computed (or approximated) and only $\boldsymbol{w}$, a column vector containing K unknown coefficients, remains to be estimated.

Under the assumption that the measurement errors η are independent and follow an identical normal distribution, the maximum likelihood estimate of $\boldsymbol{w}$ naturally leads to the L_2 norm as a goodness measure and is equivalent to the corresponding least squares (LSQ) problem that minimizes the residual sum of squares

$$(P_1) \qquad \min \|\boldsymbol{A}\boldsymbol{w} - \boldsymbol{s}\|^2. \tag{9}$$

The solution in the least squares sense is given by $\boldsymbol{w} = \boldsymbol{A}^+\boldsymbol{s}$, where $\boldsymbol{A}^+ = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}$ is the pseudoinverse of $\boldsymbol{A}$. Direct methods are adequate here since in our application the size of the linear system in (9) is not so large that it requires iterative methods. Moreover, since the matrix $\boldsymbol{A}$ is independent of the spatial location, the pseudoinverse is only computed once, and hence the computational burden is light. Despite its simplicity and efficiency, a direct solution to the linear system is highly susceptible to noise when the matrix $\boldsymbol{A}$ is ill-conditioned, which is usually the case in our application as illustrated in Fig. 1.

Many methods aiming at reliable multifiber reconstruction in the presence of noise have been employed, including low-pass filtering [24], the maximum entropy principle [25], and Tikhonov regularization [26]. In the Tikhonov regularization framework, the problem in (9) is reformulated as

$$(P_2) \qquad \min \|\boldsymbol{A}\boldsymbol{w} - \boldsymbol{s}\|^2 + \alpha\|\mathsf{T}\boldsymbol{w}\|^2, \tag{10}$$

where α is a regularization parameter and T is a regularization operator. To penalize the magnitude of the estimates, (P_2) in (10) with T being the identity operator, I, is often used and yields the solution: $\boldsymbol{w} = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T + \alpha\mathsf{I})^{-1}\boldsymbol{s}$. Recently, a damped singular value decomposition was used to regularize the fiber orientation distribution [36], where the damping factor α is determined by minimizing the generalized cross validation (GCV) criterion, which provides a simple and objective method for choosing the regularization parameter, though the real optimal solution is never guaranteed due to the theoretical limits of the GCV method [37].

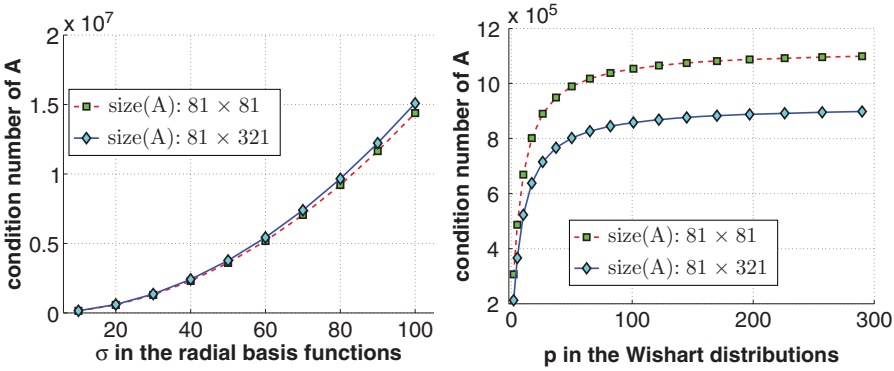


Fig. 1. The linear system constructed in (8) is often extremely ill-conditioned with very high condition number. The plot on the left shows the profile of condition numbers when the $\boldsymbol{A}$ matrices are constructed from the radial basis function and the tensor kernel model as in [25]. The plot on the right shows the case with a standard diffusion tensor kernel weighted by a mixture of Wisharts. In both scenarios there are $K = 81$ diffusion gradient directions. Two tessellation schemes of different resolution levels ($N = 81$ and $N = 321$) are considered for each model

In the normal clinical setting, the number of diffusion MR image acquisitions, K, is rarely greater than 100. On the other hand, a high resolution tessellation with $N > 100$ is usually preferred for an accurate reconstruction. This situation yields a "fat" rectangular system matrix $\boldsymbol{A}$ in (8) and leads to an under-determined system of equations with infinitely many solutions in the least squares sense. A direct solver may produce a solution with many negative-valued components, which is not physically meaningful. Another issue related to this configuration is the sparsity constraint. Since the number of significant spikes in $\boldsymbol{w}$ is indicative of the number of maxima in the displacement probability surfaces, $\boldsymbol{w}$ is expected to have a sparse support.

A recent strand of research has established a number of interesting results, both theoretical and experimental, on stable recovery of sparse overcomplete representations in the presence of noise (see [38] and references therein). The concept of *mutual coherence* (defined in [39]) of the dictionary, $\boldsymbol{A}$, plays an important role in these results. It is shown in [39] that when the system $\boldsymbol{A}$ has a property of *mutual incoherence* and that when the ideal noiseless signal $\boldsymbol{s}$ has a sufficiently sparse representation with respect to $\boldsymbol{A}$, minimizing the L_1 norm of the solution ($\|\boldsymbol{w}\|_1 = \sum_i |w_i|$) often recovers the sparsest solution and is locally stable, that is, under the addition of small amounts of noise, the result has an error that is at worst proportional to the input noise level. Unfortunately, it turns out that our systems consistently have very high mutual coherence values, in the range of 0.95–1, which makes these nice theoretical bounds inapplicable to our problem. To investigate the performance

of L_1 minimization-based methods in the context of our problem, we experimentally tested the $L1$-MAGIC package,[3] a collection of MATLAB routines for solving the convex optimization problems central to sparse signal recovery [40]. Among the several problems implemented in the $L1$-MAGIC package, we were particularly interested in these two problems: (1) Min-L_1 with equality constraints:

$$(P_3) \quad \min \|\boldsymbol{w}\|_1 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{w} = \boldsymbol{s} \tag{11}$$

and (2) Min-L_1 with quadratic constraints:

$$(P_4) \quad \min \|\boldsymbol{w}\|_1 \quad \text{subject to} \quad \|\boldsymbol{A}\boldsymbol{w} - \boldsymbol{s}\|_2 \leq \epsilon \tag{12}$$

where ϵ is a user specified parameter. Both problems (P_3) and (P_4) find the vector with the smallest L_1 norm that best explains the observation $\boldsymbol{s}$. (P_3) can be recast as a linear-programming (LP) problem, while (P_4) can be recast as a second-order cone programming (SOCP) problem (see [38, 41] and references therein for details). We will report the results of implementation of these methods for the sake of comparison in the next section.

However, the nonnegativity constraints are not explicitly enforced in the above L_1 minimization methods (P_3) and (P_4). To reduce the ambiguity of the infinitely many solutions, we constrain the solution to the nonnegative space while minimizing the least-squares criterion:

$$(P_5) \quad \min \|\boldsymbol{A}\boldsymbol{w} - \boldsymbol{s}\|_2^2 \quad \text{subject to} \quad \boldsymbol{w} \geq 0. \tag{13}$$

This nonnegative least squares (NNLS) minimization can be formulated as a quadratic programming problem that finds the minimum point of a concave quadratic function in a linearly bounded convex feasible hyperspace. Lawson and Hanson's algorithm [42, Chap. 23] is one of the most widely used algorithms for NNLS. In this classic algorithm, the linear inequality constraints are treated by an *active set strategy*, which tends to find the sparse solution if there exists such a solution, even though the sparsity constraint is not explicitly imposed. The number of iterations to reach the full convergence, as expected, depends on the amount of noise in the measurements. However, a fairly satisfactory solution can often be achieved well before the full convergence. Additionally, unlike other iterative methods mentioned earlier, this algorithm requires no arbitrary cutoff parameter and the output is not sensitive to the initial guess. There are two practical matters in solving the NNLS problem. The first is the size of the system matrix and the second is the number of right-hand sides. In the context of our reconstruction problem, the size of the matrix is relatively small, at most hundreds by hundreds, which makes the active set method still a reasonable choice over some other algorithms for large scale problems [43]. However, in real reconstruction of volume data, we do have to solve the NNLS problem voxelwise. Unlike the unconstrained least squares and regularized least squares where the pseudoinverse or the damped

[3] http://www.l1-magic.org

inverse can be computed only once and reused for multiple right-hand sides, the active set method usually has to solve different sequences of subproblems in its inner loop. To alleviate this problem, a recent variant of NNLS algorithm [44] is able to avoid the many unnecessary recomputations by rearranging the calculations in the standard active set method on the basis of combinatorial reasoning. This so-called fast combinatorial NNLS (FC-NNLS) has been tested in our experiments and proved much faster than the standard NNLS algorithm in the real MR volume data.

Once $\boldsymbol{w}$ is estimated using one of the deconvolution methods, the signal decay $S(\boldsymbol{q})/S_0$ is expressed as the Laplace transform of the mixing Wisharts distribution in the continuous diffusion tensor model(1), that is, $S(\boldsymbol{q})/S_0) = \int_{\mathcal{P}_n} e^{-\boldsymbol{q}^T \mathsf{D} \boldsymbol{q} t} \mathrm{d}F(\mathsf{D})$, where $F(\mathsf{D}) = \sum_{i=1}^{N} w_i \mathrm{d}\gamma_{p_i,\Sigma_i}$. Then the Fourier transform relation can be applied to compute the displacement probabilities:

$$\begin{aligned} P(\boldsymbol{r}) &= \int (S(\boldsymbol{q})/S_0)\, \exp(-i\boldsymbol{q}\cdot\boldsymbol{r})\, \mathrm{d}\boldsymbol{q} \\ &= \int_{R^3}\int_{\mathcal{P}_n} e^{-\boldsymbol{q}^T \mathsf{D} \boldsymbol{q} t} \mathrm{d}F(\mathsf{D}) e^{-i\boldsymbol{q}\cdot\boldsymbol{r}} \mathrm{d}\boldsymbol{q} \approx \sum_{i=1}^{N} \frac{w_i}{\sqrt{(4\pi t)^3 |\hat{\mathsf{D}}_i|}} \exp(\frac{-\boldsymbol{r}^T \hat{\mathsf{D}}_i^{-1} \boldsymbol{r}}{4t}), \end{aligned} \tag{14}$$

where $\boldsymbol{r}$ is the displacement vector and $\hat{\mathsf{D}}_i = p\Sigma_i$ are the expected values of γ_{p,Σ_i}. The $P(\boldsymbol{r})$ function defined earlier describes the probability for water molecules to move a fixed distance and has been employed in the DOT method [20]. Note that the end result here is approximated as a mixture of oriented Gaussians. Thanks to the nice analytic properties of Gaussian functions, many of the quantities produced by other methods including the radial integral of $P(\boldsymbol{r})$ in QBI [15] and the integral of $P(\boldsymbol{r})\boldsymbol{r}^2$ in DSI [19] are analytically computable using our technique. This fact provides an important opportunity to understand these quantities and evaluate their performances in resolving complicated structures.

4 Experimental Results

4.1 Numerical Simulations

Simulations were performed on the synthetic data generated from the cylindrical boundary restricted diffusion model [45], which is a more realistic HARDI simulation of the diffusion MR signals than the multiexponential model. Figure 2 illustrates the three fiber geometries used in our simulations. The same dataset was used in [20].

In our first experiment,the noiseless 1-fiber simulated HARDI data was passed to the five deconvolution methods described in Sect. 3. Figure 3 shows

Fig. 2. HARDI simulations of 1-, 2-, and 3-fibers ($b = 1,500\,\mathrm{s\,mm^{-2}}$) visualized in QBI ODF surfaces using [16, (21)]. Orientation configurations: azimuthal angles, $\phi_1 = 30°,\ \phi_2 = \{20°, 100°\},\ \phi_3 = \{20°, 75°, 135°\}$; polar angles are all 90°

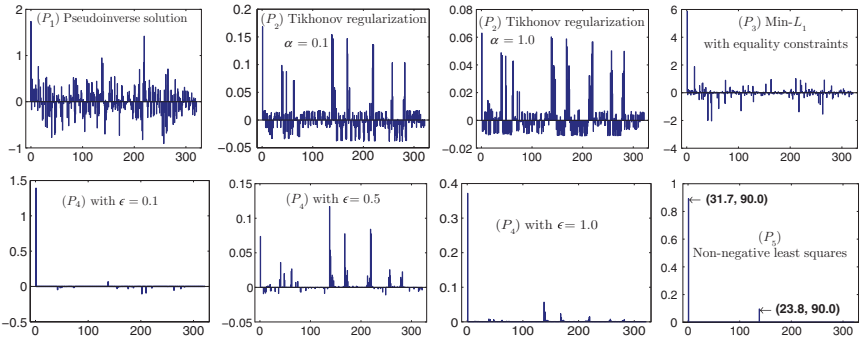


Fig. 3. Deconvolution-based computation of $\boldsymbol{w}$ in the 1-fiber HARDI simulation. The matrix $\boldsymbol{A}$ is of size 81×321 and is constructed from the Wishart model with $p = 2$. The Min-L_1 algorithms are solved using the package developed in [41]. The NNLS solver is the MATLAB built-in *lsqnonneg* function based on Lawson and Hanson's algorithm [42]

the results of $\boldsymbol{w}$ obtained from these methods and is able to give a qualitative impression of these methods. It is clear that the least squares solution to (P_1) contains a large portion of negative weights and has a relatively large magnitude. A zeroth-order Tikhonov regularization (P_2) is able to reduce the magnitude significantly but does not help achieve the sparsity and nonnegativity. By minimizing the L_1-norm with equality constraints, (P_3) yields a relatively sparse solution, but the magnitude and the negative values are not well controlled. The result produced by (P_4) has better sparsity and nonnegativity. The best result, however, is obtained by solving (P_5) using NNLS. Among the 321 components, only two are significant spikes, both of which lie in the neighborhood of the true fiber orientations $(30°, 90°)$. It is important to note that (1) the true fiber orientations do not necessarily occur at the maxima of the discrete $\boldsymbol{w}$ vector, and (2) although all of these different results for $\boldsymbol{w}$ actually lead to a very good approximation of the true displacement probability function $P(\boldsymbol{r})$ after taking the Fourier transforms, a sparse

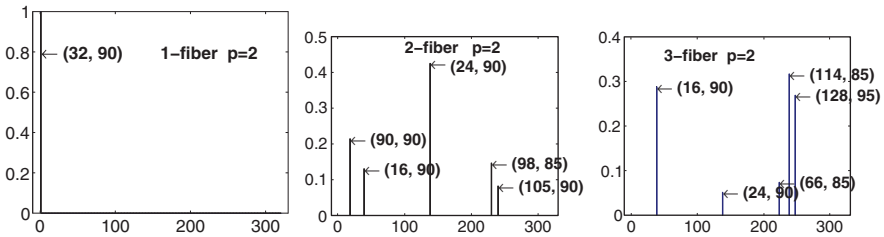


Fig. 4. Deconvolution-based computation of $\boldsymbol{w}$ using NNLS on simulated data in the presence of noise. Gaussian noise with $\sigma = 0.06$ was added to the real and imaginary parts of the signal, which is a common approximation to the Rician noise model. The matrix $\boldsymbol{A}$ is constructed by using the Wishart model with $p = 2$ and a tessellation of size $N = 321$. From left to right are the 1-fiber, 2-fiber, and 3-fiber simulations, respectively. The spikes in each resulting $\boldsymbol{w}$ are marked with the corresponding azimuthal and polar angles in degrees. In the ground truth orientation configurations, azimuthal angles are $\phi_1 = 30^\circ$, $\phi_2 = \{20^\circ, 100^\circ\}$, and $\phi_3 = \{20^\circ, 75^\circ, 135^\circ\}$; polar angles are all 90°

positive representation of $\boldsymbol{w}$ obviously offers a great advantage in setting the initial guess in the optimization procedure used to find the fiber orientations by estimating the extrema of $P(\boldsymbol{r})$. Considering the additional computational overhead for solving (P_3), (P_4), and (P_5) due to the iterative optimization, (P_5) is slightly slower than (P_3) (LP), but significantly faster than (P_4) (SOCP). Figure 4 further shows the results of using NNLS on the noisy simulated data. Clearly, NNLS is still able to produce quite accurate solutions, which also are sparse representations.

In the second experiment on the simulated data, a quantitative comparison was performed among the proposed mixture of Wisharts (MOW) model and the two model-free methods, namely, the Q-ball ODF[15] and the DOT [20]. All the resulting $P(\boldsymbol{r})$ surfaces were represented by spherical harmonic coefficients up to order $l = 6$. The Q-ball ODF was computed using the formula in [16, (21)]. First, to gain a global assessment of these methods in terms of stability, we calculated the similarity between each noisy $P(\boldsymbol{r})$ and the corresponding noiseless $P(\boldsymbol{r})$ using the angular correlation coefficient formula given in [16, (71)]. The angular correlation ranges from 0 to 1, where 1 implies identical probability profiles. Then, we estimated the fiber orientations of each system by finding the maxima of the probability surfaces with a Quasi-Newton numerical optimization algorithm and computed the deviation angles between the estimated and the true fiber orientations. Figure 5 shows the mean and standard deviation of the angular correlation coefficients, and error angles, respectively, for the 2-fiber simulation. Note that among the three methods examined, only MOW results in small error angles and high correlation coefficients in the presence of relatively large noise. This trend also holds for the 1-fiber and the 3-fiber simulations. This can be explained by noting that NNLS is able to locate the sparse spikes quite accurately even in the presence of a lot of noise.

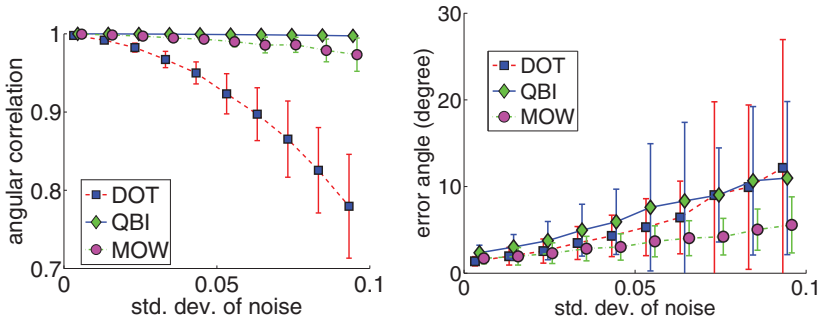


Fig. 5. The plots on the left and on the right show the statistics of angular correlation coefficients and error angles for the 2-fiber simulation, respectively. The displayed values for error angles are averaged over the 2-fiber orientations

4.2 Real Data Experiments

One of the DW-MRI datasets was obtained from a perfusion-fixed excised rat optic chiasm at 14.1 Telsa using a Bruker Avance imaging system (Bruker NMR Instruments, Billerica, MA) with a diffusion-weighted spin echo pulse sequence. Because the rat optic chiasm has a distinct myelinated structure with both parallel and decussating (crossing) optic nerve fibers, it is well suited for the validation of the fiber reconstruction results. In this dataset, there are 46 images acquired from 46 gradient directions at a b-value of 1,250 s mm^{-2}. One additional single image was acquired at $b \approx 0$ s mm^{-2}. Echo time and repetition time were 23 ms and 0.5 s respectively; Δ and δ values were set to 12.4 and 1.2 ms respectively; bandwidth was set to 35 kHz; signal average was 10; a matrix size of $128 \times 128 \times 5$ was used with a resolution of $33.6 \times 33.6 \times 200\,\mu\text{m}^{-3}$. The optic chiasm images were signal averaged to a $67.2 \times 67.2 \times 200\,\mu\text{m}^{-3}$ resolution before the subsequent processing.

The displacement probability functions for the optic chiasm image were generated using four different methods, namely (a) QBI-ODF, (b) DOT, (c) MOW+Tikhonov regularization, and (d) MOW+NNLS. The results on a region of interest are shown in Fig. 6. For each method, the corresponding S_0 image is also shown in the upper left corner of each panel as a reference. The fast combinatorial NNLS method [44] is used here as the NNLS solver. The computation time for this region of interest containing 1,024 voxels is less than 0.5 s for all four methods on an Intel Core Duo 2.16 GHz CPU, while the standard NNLS takes about 8 s. As seen in the figure, the fiber-crossings in the optic chiasm region cannot been identified by using the QBI-ODF method. Note that both the DOT method and the MOW method with two different schemes are able to demonstrate the distinct fiber orientations in the central region of the optic chiasm, where ipsilateral myelinated axons from the two optic nerves cross and form the contralateral optic tracts. However, it is evident from the figure that compared to all other solutions, the MOW technique in

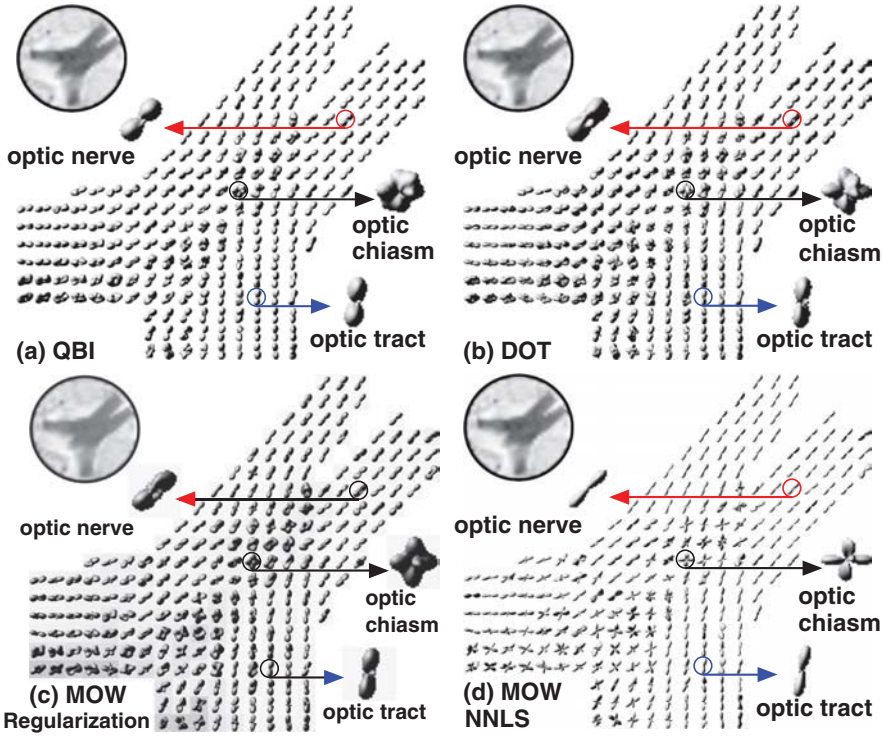


Fig. 6. Probability surfaces computed from a rat optic chiasm image using (**a**) QBI-ODF, (**b**) DOT, (**c**) MOW + Tikhonov regularization, and (**d**) MOW + NNLS. Note the decussation of myelinated axons from the two optic nerves at the center of the optic chiasm

conjunction with the NNLS scheme yields significantly sharper displacement probability surfaces. This is particularly borne out in the optic chiasm, in the center of each panel. The probability surfaces in the QBI and DOT models are blurred, in part, because both yield a corrupted $P(\boldsymbol{r})$ rather than the actual displacement probability surfaces.The corrupting factor for the QBI is a zeroth order Bessel function, for the DOT method it is a function that does not have an analytic form. This corruption affects the accuracy of the reconstructed fiber orientations as evidenced in the simulated data case where the ground truth was known. Note that validating the precise angle of the fiber crossing in this real data set is nontrivial as a histology stack must be created and then fiber directions estimated from this stack must be validated against those obtained from the DW-MRI data.

5 Conclusions

A novel mathematical model for the reconstruction of the diffusion MR signals that we recently presented in literature was reviewed in detail along with implementation results. In this model, the diffusion MR signal is related to probability distribution for diffusion tensors through a Laplace transform. An interesting discovery here is that the closed form expression for the Laplace transform of Wishart distributions leads to a Rigaut-type asymptotic fractal law for the MR signal decay behavior, which has been observed experimentally in the past [29]. Moreover, the traditional diffusion tensor model can be explained as a limiting case of our continuous mixture of Wisharts model. Further, a spherical deconvolution method was developed for resolving multiple fiber orientations using the mixture of Wisharts (MOW) model. Additionally, some important computational issues in solving the deconvolution problem that were not addressed adequately in previous studies are described in detail here. Several deconvolution schemes for achieving stable, sparse, and accurate solutions were investigated. Experimental results on both simulations and real data were reported. The comparisons with two other competing methods empirically suggest that our model combined with a nonnegative least squares deconvolution method yields efficient and accurate solutions for the multifiber reconstruction problem in the presence of intra-voxel orientational heterogeneity.

Acknowledgments

The authors thank Liz Salak for editing the manuscript. This research was supported by the National Institutes of Health (NIH) Grants EB007082, NS42075, NS36992, and RR16105, the Intramural Research Program of the National Institute of Child Health and Human Development (NICHD), and the National High Magnetic Field Laboratory (NHMFL), Tallahassee. All MRI data were obtained at the Advanced Magnetic Resonance Imaging and Spectroscopy (AMRIS) facility in the McKnight Brain Institute at the University of Florida.

References

[1] LeBihan, D. Looking into the functional architecture of the brain with diffusion MRI. Nat. Rev. Neurosci. **4**(6) (2003) 469–480
[2] Basser, P.J., Mattiello, J., Bihan, D.L. MR diffusion tensor spectroscopy and imaging. Biophys. J. **66** (1994) 259–267
[3] LeBihan, D., van Zijl, P.C.M. From the diffusion coefficient to the diffusion tensor. NMR Biomed. **15**(7/8) (2002) 431–434

[4] Parker, G.J.M. Analysis of MR diffusion weighted images. Br. J. Radiol. **77** (2004) S176–S185
[5] Kingsley, P.B. Introduction to diffusion tensor imaging mathematics: Part III. Tensor calculation, noise, simulations, and optimization. Concepts Magn. Reson. Part A **28A** (2006) 155–179
[6] Conturo, T.E., Lori, N.F., Cull, T.S., Akbudak, E., Snyder, A.Z., Shimony, J.S., McKinstry, R.C., Burton, H., Raichle, M.E. Tracking neuronal fiber pathways in the living human brain. Proc Natl Acad Sci **96** (1999) 10422–10427
[7] Mori, S., Crain, B.J., Chacko, V.P., van Zijl, P.C.M. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Ann Neurol **45** (1999) 265–269
[8] Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A. In vivo fiber tractography using DT-MRI data. Magn. Reson. Med. **44**(4) (2000) 625–632
[9] Mori, S., van Zijl, P.C.M. Fiber tracking: principles and strategies – a technical review. NMR Biomed. **15**(7-8) (2002) 468–480
[10] Frank, L. Characterization of anisotropy in high angular resolution diffusion weighted MRI. Magn. Reson. Med. **47**(6) (2002) 1083–1099
[11] Alexander, D.C., Barker, G.J., Arridge, S.R. Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. Magn. Reson. Med. **48**(2) (2002) 331–340
[12] von dem Hagen, E., Henkelman, R. Orientational diffusion reflects fiber structure within a voxel. Magn. Reson. Med. **48**(3) (2002) 454–459
[13] Tuch, D.S., Reese, T.G., Wiegell, M.R., Makris, N., Belliveau, J.W., Wedeen, V.J. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magn. Reson. Med. **48**(4) (2002) 577–582
[14] Özarslan, E., Mareci, T.H. Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. Magn. Reson. Med. **50**(5) (2003) 955–965
[15] Tuch, D.S. Q-ball imaging. Magn. Reson. Med. **52**(6) (2004) 1358–1372
[16] Anderson, A.W. Measurement of fiber orientation distributions using high angular resolution diffusion imaging. Magn. Reson. Med. **54**(5) (2005) 1194–1206
[17] Hess, C.P., Mukherjee, P., Han, E.T., Xu, D., Vigneron, D.B. Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. Magn. Reson. Med. **56**(1) (2006) 104–117
[18] Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R. Regularized, fast and robust analytical q-ball imaging. Magn. Reson. Med. **58** (2007) 497–510
[19] Wedeen, V.J., Hagmann, P., Tseng, W.Y.I., Reese, T.G., Weisskoff, R.M. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. Magn. Reson. Med. **54**(6) (2005) 1377–1386

[20] Özarslan, E., Shepherd, T.M., Vemuri, B.C., Blackband, S.J., Mareci, T.H. Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT). NeuroImage **31** (2006) 1086–1103
[21] Behrens, T., Woolrich, M., Jenkinson, M., Johansen-Berg, H., Nunes, R., Clare, S., Matthews, P., Brady, J., Smith, S. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn. Reson. Med. **50**(2) (2003) 1077–1088
[22] Hosey, T., William, G., Ansorge, R. Inference of multiple fiber orientations in high angular resolution diffusion imaging. Magn. Reson. Med. **54** (2005) 1480–1489
[23] Ramirez-Manzanares, A., Rivera, M., Vemuri, B.C., Carney, P.R., Mareci, T.H. Diffusion basis functions decomposition for estimating white matter intravoxel fiber geometry. IEEE Trans. Med. Imaging **26**(8) (2007) 1091–1102
[24] Tournier, J.D., Calamante, F., Gadian, D.G., Connelly, A. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. NeuroImage **23**(3) (2004) 1176–1185
[25] Alexander, D.C. Maximum entropy spherical deconvolution for diffusion MRI. In: IPMI. (2005) 76–87
[26] Tournier, J.D., Calamente, F., Connelly, A. Improved characterisation of crossing fibres: spherical deconvolution combined with Tikhonov regularization. In: ISMRM. (2006)
[27] Jian, B., Vemuri, B.C., Özarslan, E., Carney, P.R., Mareci, T.H. A novel tensor distribution model for the diffusion weighted MR signal. NeuroImage **37**(1) (2007) 164–176
[28] Jian, B., Vemuri, B.C. A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI. IEEE Trans. Med. Imaging **26**(11) (2007) 1464–1471
[29] Köpf, M., Metzler, R., Haferkamp, O., Nonnenmacher, T.F. NMR studies of anomalous diffusion in biological tissues: Experimental observation of Lévy stable processes. In: Losa, G.A., Merlini, D., Nonnenmacher, T.F., Weibel, E.R., eds. Fractals in Biology and Medicine. Volume 2. Birkhäuser, Basel (1998) pp. 354–364
[30] Mathai, A.M. Jacobians of matrix transformations and functions of matrix argument. World Scientific, Singapore (1997)
[31] Wishart, J. The generalized product moment distribution in samples from a normal multivariate population. Biometrika **20** (1928) 32–52
[32] Murihead, R.J. Aspects of multivariate statistical theory. Wiley, New York (1982)
[33] Rigaut, J.P. An empirical formulation relating boundary lengths to resolution in specimens showing 'non-ideally fractal' dimensions. J. Microsc. **133** (1984) 41–54
[34] Sen, P.N., Hürlimann, M.D., de Swiet, T.M. Debye-Porod law of diffraction for diffusion in porous media. Phys. Rev. B **51**(1) (1995) 601–604

[35] Ramirez-Manzanares, A., Rivera, M., Vemuri, B.C., Mareci, T.H. Basis functions for estimating intra-voxel structure in DW-MRI. In: Proc. IEEE Medical Imaging Conference, Rome, Italy (2004) pp. 4207–4211
[36] Sakaie, K.E., Lowe, M.J. An objective method for regularization of fiber orientation distribution derived from diffusion-weighed MRI. NeuroImage **34** (2007) 169–176
[37] Wahba, G. Spline models for observational data. SIAM, Philadelphia, PA (1990)
[38] Candès, E.J., Romberg, J.K., Tao, T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Info. Theory **52**(2) (2006) 489–509
[39] Donoho, D.L., Elad, M., Temlyakov, V.N. Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans. Info. Theory **52**(1) (2006) 6–18
[40] Candes, E.J., Romberg, J., Tao, T. Stable signal recovery from incomplete and inaccurate measurements. Commun. Pure Appl. Math. **59**(8) (2006) 1207–1223
[41] Candés, E., Romberg, J. l_1-MAGIC (http://www.l1-magic.org) (2006)
[42] Lawson, C., Hanson, R.J. Solving least squares problems. SIAM, Philadelphia, PA (1995)
[43] Chen, D., Plemmons, R.J. Nonnegativity constraints in numerical analysis. In: Bultheel, A., Cools, R., eds. The Proceedings of the Symposium on the Birth of Numerical Analysis, World Scientific Press, Singapore (2007)
[44] Benthem, M.H.V., Keenan, M.R. Fast algorithm for the solution of large-scale non-negativity-constrained least squares problems. J. Chemometrics **18**(10) (2004) 441–450
[45] Söderman, O., Jönsson, B. Restricted diffusion in cylindirical geometry. J. Magn. Reson. B **117**(1) (1995) 94–97

The Algebra of Fourth-Order Tensors with Application to Diffusion MRI

Maher Moakher

Laboratory for Mathematical and Numerical Modeling in Engineering Science, National Engineering School at Tunis, B.P. 37, 1002 Tunis-Belvédère, Tunisia
maher.moakher@gmail.com

Summary. In this chapter we give different descriptions of fourth-order tensors. We show that under certain symmetries it is possible to describe a fourth-order tensor in three-dimensional space by a second-order tensor in a six-dimensional space. Such a representation makes the manipulation of fourth-order tensors similar to that of the more familiar second-order tensors. We discuss the algebra of the space of fourth-order symmetric tensors and describe different metrics on this space. Special emphasis is placed on totally symmetric tensors and on orientation tensors. Applications to high angular resolution diffusion imaging are discussed.

1 Introduction

The word "tensor" derives from the Latin word "tensus" the past participle of "tendĕre," that is, to stretch. It has been used in anatomy as early as 1704 to denote muscle that stretches. In mathematics, it was first introduced in 1846 by William Rowan Hamilton in the context of quaternion algebra. The tensor of a quaternion is the generalization of the modulus of a complex number. However, the tensor in Hamilton sense did not survive. The current meaning of a tensor is known to be due to Woldemar Voigt, who used the term *tensortripel* in crystal elasticity (1899). Nevertheless, John Willard Gibbs used, as early as 1884, the notion of *right tensor* as a generalized vector to describe strain in an elastic body. The foundations of tensor calculus, then known as *absolute differential calculus*, have been developed much earlier in the studies of Gregorio Ricci-Curbastro and Tullio Levi-Civita. Their work is a continuation of the pioneering works of Karl Friedrich Gauß (1827), Bernhard Riemann (1861), and Elwin B. Christoffel (1869). Further historical accounts can be found, for example, in [10, 12, 16]. Tensor calculus and analysis are now used in many fields such as continuum mechanics, electricity, magnetism, differential geometry, general relativity, multilinear algebra, and recently diffusion tensor imaging [21].

There are several ways to describe tensors. In an abstract setting, tensors can be defined as multilinear forms, or as multilinear maps. With respect to

a chosen coordinate system,[1] tensors can also be viewed as multidimensional arrays. We emphasize that these arrays must transform according to specific rules under a change of coordinates. Tensors are often considered as a generalization of vectors, whereby scalars are zeroth-order tensors and vectors are first-order tensors.

The algebra of second-order tensors is fairly standard and is used in various branches of continuum physics such as continuum mechanics, electromagnetism. In contrast, the use of fourth-order tensors has been mainly restricted to elasticity tensors. Unsurprisingly, most of the notions and notations are borrowed from elasticity literature. Recently, with the introduction of high angular resolution diffusion imaging (HARDI), fourth-order tensors have gained some interests within the diffusion imaging community [4, 7, 15]. The aim of this chapter is twofolds. First it is intended to give the basic algebra of fourth-order tensors. Second, we show that it is possible to express this algebra in terms of the more familiar algebra of second-order tensors but in higher spatial dimension.

Let V be a real vector space of dimension n. While in practice the underlying vector space V has dimension $n = 2$ or $n = 3$, the results presented in this chapter hold for arbitrary n. For elasticity and fourth-order diffusion tensors, we give details for the case $n = 3$. Let $\mathfrak{B} := \{\boldsymbol{e}_i\}_{i=1,\dots,n}$ be an orthonormal basis of V. Then any vector $\boldsymbol{u} \in V$ can be represented as $\boldsymbol{u} = u_i \boldsymbol{e}_i$ and the scalars u_i are called the components of $\boldsymbol{u}$ with respect to the basis $\mathfrak{B}$. On V we define the inner product

$$\boldsymbol{u} \cdot \boldsymbol{v} := \boldsymbol{u}^T \boldsymbol{v} = u_i v_i. \tag{1}$$

Here and throughout, the Einstein convention on repeated indices is employed. We shall use both direct (coordinate-free) and indicial notations [9].

Although we are concerned primarily with fourth-order tensors in the n-dimensional space V, calculation and presentation are sometimes better performed using second-order tensors in higher dimensions. We shall adopt the following scheme of notation. Boldface lower case Latin (e.g., $\boldsymbol{a}$), boldface upper case Latin (e.g., $\boldsymbol{A}$), and open face upper case Latin (e.g., $\mathbb{A}$) indicate n-dimensional tensors of order 1, 2, and 4, respectively. Boldface lower case Latin with a superposed $\tilde{}$ (e.g., $\tilde{\boldsymbol{a}}$) and boldface upper case Latin with a superposed $\tilde{}$ (e.g., $\tilde{\boldsymbol{A}}$) indicate n^2-dimensional vectors and second-order tensors, respectively. Similarly, boldface lower case Latin with a superposed $\hat{}$ (e.g., $\hat{\boldsymbol{a}}$) and boldface upper case Latin with a superposed $\hat{}$ (e.g., $\hat{\boldsymbol{A}}$) indicate $\frac{1}{2}n(n+1)$-dimensional vectors and second-order tensors, respectively.

The remainder of this chapter is organized as follows. The algebra of n-dimensional second-order tensors and their representations as vectors in higher dimensions is reviewed in Sect. 2. In Sect. 3 we give the algebra of

[1] In this chapter we shall consider Cartesian coordinate system and orthonormal basis only. Accordingly, we consider only Cartesian tensors and we shall not dwell into covariant and contravariant tensors, components, etc.

fourth-order tensors. Section 4 is devoted to a presentation of the different symmetries of fourth-order tensors. Section 5 gives the details of six-dimensional second-order tensor representation of three-dimensional fourth-order tensors. In Sect. 6 we discuss various types of material symmetries of three-dimensional fourth-order tensors. Orientation tensors are presented in Sect. 7. Finally, in Sect. 8 we discuss fourth-order diffusion tensors.

2 Second-Order Tensors

2.1 Second-Order Tensors as Linear Maps

A second-order tensor $\boldsymbol{A}$ is a linear map from V to V, which assigns to each vector $\boldsymbol{u} \in V$ the vector $\boldsymbol{Au}$ (see, e.g., [9]):

$$\begin{aligned} \boldsymbol{A} : V &\to V \\ \boldsymbol{u} &\mapsto \boldsymbol{Au}. \end{aligned}$$

Let $\boldsymbol{e}_i \otimes \boldsymbol{e}_j$, $i, j = 1, \ldots, n$, denote the second-order tensor such that for all $\boldsymbol{u} \in V$, we have $(\boldsymbol{e}_i \otimes \boldsymbol{e}_j)\boldsymbol{u} = (\boldsymbol{u} \cdot \boldsymbol{e}_j)\boldsymbol{e}_i$. Then any second-order tensor $\boldsymbol{A}$ has the representation

$$\boldsymbol{A} = A_{ij}\boldsymbol{e}_i \otimes \boldsymbol{e}_j, \tag{2}$$

where $A_{ij} = \boldsymbol{e}_i \cdot \boldsymbol{Ae}_j$. The coefficients A_{ij} are called the components of $\boldsymbol{A}$ with respect to the basis $\mathfrak{B}$. It follows that the set of all second-order tensors, that is, the set of all linear maps from V to V, is a linear space of dimension n^2 and is denoted by $\mathrm{Lin}(V)$. The product $\boldsymbol{AB}$ of two second-order tensors is defined by composition

$$(\boldsymbol{AB})\boldsymbol{u} = \boldsymbol{A}(\boldsymbol{Bu}), \quad \forall\, \boldsymbol{u} \in V.$$

In components we have $(\boldsymbol{AB})_{ij} = A_{ik}B_{kj}$.

The transpose of $\boldsymbol{A} \in \mathrm{Lin}(V)$ is the second-order tensor denoted $\boldsymbol{A}^T$ and defined by

$$\boldsymbol{u} \cdot \boldsymbol{A}^T\boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{Au} \quad \forall\, \boldsymbol{u}, \boldsymbol{v} \in V. \tag{3}$$

On $\mathrm{Lin}(V)$ we define the inner product

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle := \mathrm{tr}(\boldsymbol{A}^T\boldsymbol{B}) = A_{ij}B_{ij}, \tag{4}$$

and the induced metric

$$\mathrm{d}_E(\boldsymbol{A}, \boldsymbol{B}) := \|\boldsymbol{A} - \boldsymbol{B}\| = [\mathrm{tr}((\boldsymbol{A} - \boldsymbol{B})^T(\boldsymbol{A} - \boldsymbol{B}))]^{1/2}. \tag{5}$$

With respect to the inner product (4), the set $\{\boldsymbol{e}_i \otimes \boldsymbol{e}_j\}_{1 \le i,j \le n}$ forms an orthonormal basis of $\mathrm{Lin}(V)$.

2.2 Second-Order Tensors as Vectors

We consider an invertible map ψ that assigns to each pair of indices (i,j), $1 \leq i,j \leq n$, a single index $\psi(i,j)$ that ranges from 1 to n^2:

$$\begin{aligned} \psi : [\![1:n]\!] \times [\![1:n]\!] &\to [\![1:n^2]\!] \\ (i,j) &\mapsto \psi(i,j). \end{aligned} \tag{6}$$

There are of course $n^2!$ such maps. For $n=3$, an example of such maps is

$$\begin{array}{c|ccccccccc} (i,j) & (1,1) & (2,2) & (3,3) & (2,3) & (1,3) & (1,2) & (3,2) & (3,1) & (2,1) \\ \hline \psi(i,j) & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array}. \tag{7}$$

For $1 \leq i,j \leq n$, we set $\tilde{\boldsymbol{e}}_{\psi(i,j)} = \boldsymbol{e}_i \otimes \boldsymbol{e}_j$. Then $\{\tilde{\boldsymbol{e}}_\alpha\}_{1\leq\alpha\leq n^2}$ is an orthonormal basis of $\tilde{V} \equiv \mathrm{Lin}(V)$. Therefore, any $\boldsymbol{A} \in \mathrm{Lin}(V)$ can be viewed as an n-dimensional second-order tensor $\boldsymbol{A} = A_{ij}\boldsymbol{e}_i \otimes \boldsymbol{e}_j$, or, as an n^2-dimensional vector

$$\boldsymbol{A} \leftrightarrow \tilde{\boldsymbol{a}} = \tilde{a}_\alpha \tilde{\boldsymbol{e}}_\alpha, \qquad \text{where } \tilde{a}_{\psi(i,j)} = A_{ij}, \quad 1 \leq i,j \leq n.$$

We have $\mathrm{tr}(\boldsymbol{A}^T\boldsymbol{B}) = \langle \boldsymbol{A}, \boldsymbol{B}\rangle = \tilde{\boldsymbol{a}} \cdot \tilde{\boldsymbol{b}}$ and $\|\boldsymbol{A}\| = \|\tilde{\boldsymbol{a}}\|$. Therefore, the two-way map $\boldsymbol{A} \leftrightarrow \tilde{\boldsymbol{a}}$ is an isometry between $\mathrm{Lin}(V)$ and $\tilde{V}$.

2.3 Symmetric Second-Order Tensors

For $\boldsymbol{A} \in \mathrm{Lin}(V)$, we say that $\boldsymbol{A}$ is symmetric if $\boldsymbol{A} = \boldsymbol{A}^T$ (or in components, $A_{ij} = A_{ji}$) and skew symmetric if $\boldsymbol{A} = -\boldsymbol{A}^T$ (or in components, $A_{ij} = -A_{ji}$). Every second-order tensor $\boldsymbol{A}$ admits the unique decomposition

$$\boldsymbol{A} = \boldsymbol{A}^s + \boldsymbol{A}^a,$$

where $\boldsymbol{A}^s = \frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^T)$ is the symmetric part of $\boldsymbol{A}$ and $\boldsymbol{A}^a = \frac{1}{2}(\boldsymbol{A} - \boldsymbol{A}^T)$ is the skew-symmetric part of $\boldsymbol{A}$.

The set of all symmetric tensors $\mathrm{Sym}(V) := \{\boldsymbol{A} \in \mathrm{Lin}(V) \mid \boldsymbol{A} = \boldsymbol{A}^T\}$ is a subspace of $\mathrm{Lin}(V)$ of dimension $n(n+1)/2$. The symmetric second-order tensors

$$\left\{ \frac{1}{2^{\frac{1}{2}(1+\delta_{ij})}} (\boldsymbol{e}_i \otimes \boldsymbol{e}_j + \boldsymbol{e}_j \otimes \boldsymbol{e}_i) \right\}_{1\leq i\leq j\leq n},$$

where δ_{ij} is the Kronecker delta, form an orthonormal basis of $\mathrm{Sym}(V)$.

If $\boldsymbol{A} \in \mathrm{Sym}(V)$ then there exist $\{\lambda_i\}_{i=1,\ldots,n} \in \mathbb{R}$ and $\{\boldsymbol{u}_i\}_{i=1,\ldots,n}$ an orthonormal basis of V such that

$$\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \otimes \boldsymbol{u}_i.$$

The real number λ_i is called the eigenvalue of $\boldsymbol{A}$ associated with the eigenvector $\boldsymbol{u}_i$. Such a representation is called the spectral decomposition of $\boldsymbol{A}$. We have $\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^n \lambda_i$ and $\det(\boldsymbol{A}) = \prod_{i=1}^n \lambda_i$.

2.4 Symmetric Second-Order Tensors as Vectors

We consider an invertible map φ that assigns to each pair of indices (i, j) with $1 \le i \le j \le n$, a single index $\varphi(i, j)$ that ranges from 1 to $\frac{1}{2}n(n+1)$:

$$\begin{array}{l} \varphi : T_n \to [\![1 : \frac{1}{2}n(n+1)]\!] \\ \quad (i,j) \mapsto \varphi(i,j), \end{array} \tag{8}$$

where T_n is the set of ordered pairs (i, j), $1 \le i \le j \le n$. For $n = 3$, an example of such maps is

$$\begin{array}{c|cccccc} (i,j) & (1,1) & (2,2) & (3,3) & (2,3) & (1,3) & (1,2) \\ \hline \varphi(i,j) & 1 & 2 & 3 & 4 & 5 & 6 \end{array}. \tag{9}$$

For $1 \le i \le j \le n$, we set

$$\hat{\boldsymbol{e}}_{\varphi(i,j)} = \frac{1}{2^{\frac{1}{2}(1+\delta_{ij})}}(\boldsymbol{e}_i \otimes \boldsymbol{e}_j + \boldsymbol{e}_j \otimes \boldsymbol{e}_i). \tag{10}$$

Then $\{\hat{\boldsymbol{e}}_\alpha\}_{1 \le \alpha \le n(n+1)/2}$ is an orthonormal basis of $\hat{V} \equiv \mathrm{Sym}(V)$.

Therefore, any $\boldsymbol{A} \in \mathrm{Sym}(V)$ can be viewed as an n-dimensional symmetric second-order tensor $\boldsymbol{A} = A_{ij}\boldsymbol{e}_i \otimes \boldsymbol{e}_j$, or, as an $\frac{1}{2}n(n+1)$-dimensional vector

$$\boldsymbol{A} \leftrightarrow \hat{\boldsymbol{a}} = \hat{a}_\alpha \hat{\boldsymbol{e}}_\alpha, \quad \text{where } \hat{a}_{\varphi(i,j)} = A_{ij}, \quad 1 \le i \le j \le n.$$

We have $\mathrm{tr}(\boldsymbol{A}^T\boldsymbol{B}) = \langle \boldsymbol{A}, \boldsymbol{B} \rangle = \langle \hat{\boldsymbol{a}}, \hat{\boldsymbol{b}} \rangle = \hat{\boldsymbol{a}} \cdot \hat{\boldsymbol{b}}$ and $\|\boldsymbol{A}\| = \|\hat{\boldsymbol{a}}\|$. Therefore, the two-way map $\boldsymbol{A} \leftrightarrow \hat{\boldsymbol{a}}$ is an isometry between $\mathrm{Sym}(V)$ and $\hat{V}$.

2.5 Symmetric Positive-Definite Second-Order Tensors

A symmetric second-order tensor $\boldsymbol{A}$ is said to be positive definite if

$$\boldsymbol{u} \cdot \boldsymbol{A}\boldsymbol{u} > 0, \quad \forall\, \boldsymbol{u} \in V,\ \boldsymbol{u} \neq \boldsymbol{0}.$$

The set of all symmetric positive-definite second-order tensors $\mathrm{Sym}^{++}(V)$ is the interior of a convex cone.

On $\mathrm{Sym}^{++}(V)$, in addition to the Euclidean distance (5), we also have the Riemannian distance

$$\mathrm{d}_R(\boldsymbol{A}, \boldsymbol{B}) := [\mathrm{tr}(\mathrm{Log}^2(\boldsymbol{A}^{-1}\boldsymbol{B}))]^{1/2}. \tag{11}$$

It is important to note that the space $\mathrm{Sym}^{++}(V)$ is not complete with respect to the Euclidean metric (5), whereas $\mathrm{Sym}^{++}(V)$ is a complete metric space with respect to the Riemannian metric (11).

3 Fourth-Order Tensors

3.1 Fourth-Order Tensors as Linear Maps

A fourth-order tensor $\mathbb{A}$ is a linear map from $\mathrm{Lin}(V)$ to $\mathrm{Lin}(V)$, that is, it assigns to each second-order tensor $\boldsymbol{A}$ the second-order tensor $\mathbb{A}\boldsymbol{U}$ (see, e.g., [9]):

$$\begin{aligned}\mathbb{A} : \mathrm{Lin}(V) &\to \mathrm{Lin}(V)\\ \boldsymbol{U} &\mapsto \quad \mathbb{A}\boldsymbol{U}.\end{aligned}$$

Any fourth-order tensor $\mathbb{A}$ has the representation $\mathbb{A} = A_{ijkl}\boldsymbol{e}_i \otimes \boldsymbol{e}_j \otimes \boldsymbol{e}_k \otimes \boldsymbol{e}_l$, where $A_{ijkl} = \langle \boldsymbol{e}_i \otimes \boldsymbol{e}_j, \mathbb{A}(\boldsymbol{e}_k \otimes \boldsymbol{e}_l)\rangle$. The set of all fourth-order tensors, that is, the set of all linear maps from $\mathrm{Lin}(V)$ to $\mathrm{Lin}(V)$, is a linear space of dimension n^4 and is denoted by $\mathbb{L}\mathrm{in}(V)$. The product $\mathbb{A}\mathbb{B}$ of two fourth-order tensors is defined by composition

$$(\mathbb{A}\mathbb{B})\boldsymbol{U} = \mathbb{A}(\mathbb{B}\boldsymbol{U}), \quad \forall\, \boldsymbol{U} \in \mathrm{Lin}(V).$$

In components we have $(\mathbb{A}\mathbb{B})_{ijkl} = A_{ijpq}B_{pqkl}$.

The transpose of $\mathbb{A} \in \mathbb{L}\mathrm{in}(V)$ is the fourth-order tensor denoted $\mathbb{A}^T$ and defined by

$$\langle \boldsymbol{U}, \mathbb{A}^T\boldsymbol{V}\rangle = \langle \boldsymbol{V}, \mathbb{A}\boldsymbol{U}\rangle, \quad \forall\, \boldsymbol{U}, \boldsymbol{V} \in \mathrm{Lin}(V).$$

On $\mathbb{L}\mathrm{in}(V)$ we define the inner product

$$\langle \mathbb{A}, \mathbb{B}\rangle := \mathrm{tr}(\mathbb{A}^T\mathbb{B}) = A_{ijkl}B_{ijkl}, \tag{12}$$

the induced norm

$$\|\mathbb{A}\| = [\mathrm{tr}(\mathbb{A}^T\mathbb{A})]^{1/2}, \tag{13}$$

and metric

$$\mathrm{d}_E(\mathbb{A}, \mathbb{B}) = \|\mathbb{A} - \mathbb{B}\|. \tag{14}$$

With respect to the inner product (12), the set $\{\boldsymbol{e}_i \otimes \boldsymbol{e}_j \otimes \boldsymbol{e}_k \otimes \boldsymbol{e}_l\}_{1\le i,j,k,l\le n}$ forms an orthonormal basis of $\mathbb{L}\mathrm{in}(V)$.

3.2 Fourth-Order Tensors as Second-Order Tensors

For $1 \le i, j \le n$, we set $\tilde{\boldsymbol{e}}_{\psi(i,j)} = \boldsymbol{e}_i \otimes \boldsymbol{e}_j$. Then $\{\tilde{\boldsymbol{e}}_\alpha \otimes \tilde{\boldsymbol{e}}_\beta\}_{1\le\alpha,\beta\le n^2}$ is an orthonormal basis of $\mathrm{Lin}(\tilde{V}) \equiv \mathbb{L}\mathrm{in}(V)$.

Therefore, any $\mathbb{A} \in \mathbb{L}\mathrm{in}(V)$ can be viewed as an n-dimensional fourth-order tensor

$$\mathbb{A} = A_{ijkl}\boldsymbol{e}_i \otimes \boldsymbol{e}_j \otimes \boldsymbol{e}_k \otimes \boldsymbol{e}_l,$$

or, as an n^2-dimensional second-order tensor

$$\mathbb{A} \leftrightarrow \tilde{\boldsymbol{A}} = \tilde{A}_{\alpha\beta}\tilde{\boldsymbol{e}}_\alpha \otimes \tilde{\boldsymbol{e}}_\beta, \quad \text{where } \tilde{A}_{\psi(i,j)\psi(k,l)} = A_{ijkl}, \quad 1 \le i, j, k, l \le n.$$

We have

$$\mathrm{tr}(\mathbb{A}^T\mathbb{B}) = \langle \mathbb{A}, \mathbb{B}\rangle = \mathrm{tr}(\tilde{\boldsymbol{A}}^T\tilde{\boldsymbol{B}}) = \langle \tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}\rangle, \quad \text{and} \quad \|\mathbb{A}\| = \|\tilde{\boldsymbol{A}}\|,$$

and hence the two-way map $\mathbb{A} \leftrightarrow \tilde{\boldsymbol{A}}$ is an isometry.

4 Symmetries of Fourth-Order Tensors

For second-order tensors there is one notion of symmetry that represents invariance under changing the order of the two indices. For fourth-order tensors, however, there are several notions of symmetry that represent invariance under exchanging pairs of indices. For these tensors we distinguish three types of symmetries: major, minor, and total symmetries. These notions of symmetries are widely used in elasticity theory [10].

4.1 Major Symmetry

For $\mathbb{A} \in \mathbb{L}\mathrm{in}(V)$, we say that $\mathbb{A}$ is symmetric if $\mathbb{A} = \mathbb{A}^T$, or in components, $A_{ijkl} = A_{klij}$; and we say that $\mathbb{A}$ is skew symmetric if $\mathbb{A} = -\mathbb{A}^T$, or in components, $A_{ijkl} = -A_{klij}$. We have the decomposition $\mathbb{A} = \frac{1}{2}(\mathbb{A} + \mathbb{A}^T) + \frac{1}{2}(\mathbb{A} - \mathbb{A}^T)$.

The set of all symmetric fourth-order tensors,

$$\mathbb{S}\mathrm{ym}(V) := \{\mathbb{A} \in \mathbb{L}\mathrm{in}(V) \mid \mathbb{A} = \mathbb{A}^T\},$$

is a subspace of $\mathbb{L}\mathrm{in}(V)$ of dimension $n^2(n^2+1)/2$. In elasticity theory, this symmetry is often called *major symmetry*. In components, $A_{ijkl} = A_{klij}, \quad 1 \le i,j,k,l \le n$, and therefore this symmetry represents invariance under changing the pair of indices (i,j) and (k,l).

If $\mathbb{A} \in \mathbb{S}\mathrm{ym}(V)$ then there exist $\{\lambda_i\}_{i=1,\dots,n^2} \in \mathbb{R}$ and $\{\boldsymbol{U}_i\}_{i=1,\dots,n^2}$ an orthonormal basis of $\mathrm{Lin}(V)$ such that

$$\mathbb{A} = \sum_{i=1}^{n^2} \lambda_i \boldsymbol{U}_i \otimes \boldsymbol{U}_i.$$

The real number λ_i is called the eigenvalue of $\mathbb{A}$ associated with the eigentensor $\boldsymbol{U}_i$. Such a representation is called the spectral decomposition of $\mathbb{A}$. The trace and determinant of the symmetric fourth-order tensor $\mathbb{A}$ are defined as

$$\mathrm{tr}(\mathbb{A}) = \sum_{i=1}^{n^2} \lambda_i, \quad \text{and} \quad \det(\mathbb{A}) = \prod_{i=1}^{n^2} \lambda_i. \tag{15}$$

4.2 Minor Symmetry

The second type of symmetry is called *minor symmetry* and is defined by $\langle \boldsymbol{U}, \mathbb{A}\boldsymbol{V}\rangle = \langle \boldsymbol{U}^T, \mathbb{A}\boldsymbol{V}\rangle = \langle \boldsymbol{U}, \mathbb{A}\boldsymbol{V}^T\rangle, \ \forall \ \boldsymbol{U}, \boldsymbol{V} \in \mathrm{Lin}(V)$. In components, $A_{ijkl} = A_{jikl} = A_{ijlk}, \quad 1 \le i,j,k,l \le n$. Invariance under exchange of the first pair of indices is called the first minor symmetry and invariance under exchange of the second pair of indices is called the second minor symmetry.

The set of all fourth-order tensors that satisfy the minor symmetry will be denoted

$$\mathbb{S}^{\text{minor}}_{\text{ym}}(V) := \{\mathbb{A} \in \mathbb{L}\text{in}(V) \mid \mathbb{A} \text{ satisfies minor symmetry}\}.$$

The set $\{\hat{\boldsymbol{e}}_\alpha \otimes \hat{\boldsymbol{e}}_\beta\}_{1\leq\alpha,\beta\leq n(n+1)/2}$, where $\hat{\boldsymbol{e}}_\alpha$ is defined in (10), is an orthonormal basis of $\text{Lin}(\hat{V}) \equiv \mathbb{S}^{\text{minor}}_{\text{ym}}(V)$. Therefore, any $\mathbb{A} \in \mathbb{S}^{\text{minor}}_{\text{ym}}(V)$ can be viewed as an n-dimensional symmetric fourth-order tensor $\mathbb{A} = A_{ijkl}\boldsymbol{e}_i \otimes \boldsymbol{e}_j \otimes \boldsymbol{e}_k \otimes \boldsymbol{e}_l$, or, as an $\frac{1}{2}n(n+1)$-dimensional second-order tensor

$$\mathbb{A} \leftrightarrow \hat{\boldsymbol{A}} = \hat{A}_{\alpha\beta}\hat{\boldsymbol{e}}_\alpha \otimes \hat{\boldsymbol{e}}_\beta, \quad \text{where } \hat{A}_{\varphi(i,j)\varphi(k,l)} = A_{ijkl}, \quad 1 \leq i \leq j \leq n. \tag{16}$$

Fourth-order tensors that satisfy both the major and minor symmetries and in addition are positive definite will be denoted

$$\mathbb{S}\text{ym}^{++}(V) = \{\mathbb{A} \in \mathbb{S}^{\text{minor}}_{\text{ym}}(V) \cap \mathbb{S}\text{ym}(V) \mid \langle \boldsymbol{U}, \mathbb{A}\boldsymbol{U}\rangle > 0, \, \forall \boldsymbol{U} \in \text{Sym}(V), \, \boldsymbol{U} \neq \boldsymbol{O}\}.$$

On this set, which is the interior of a convex cone, besides the Euclidean distance (14), we have the Riemannian distance [14]

$$\text{d}_R(\mathbb{A}, \mathbb{B}) = [\text{tr}(\text{Log}^2(\mathbb{A}^{-1}\mathbb{B}))]^{1/2}. \tag{17}$$

It is worthy to note that this distance is invariant under inversion, and that $\mathbb{S}\text{ym}^{++}(V)$ is a complete metric space with respect to (17) but not with respect to (14).

Let $\mathbb{A}$ and $\mathbb{B}$ be two fourth-order tensors in $\mathbb{S}^{\text{minor}}_{\text{ym}}(V)$, and $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{B}}$ are the corresponding second-order tensors in $\mathbb{L}\text{in}(\hat{V})$. Let $\boldsymbol{U}$ be a symmetric second-order tensor and let $\hat{\boldsymbol{u}} \in \hat{V}$ be its vector representation. Then by using (16) and after straightforward algebra one can prove the following statements.

1. The fourth-order tensor $\mathbb{A}\mathbb{B}$ corresponds to the second-order tensor $\hat{\boldsymbol{A}}\hat{\boldsymbol{B}}$.
2. The second-order tensor $\mathbb{A}\boldsymbol{U}$ corresponds to the vector $\hat{\boldsymbol{A}}\hat{\boldsymbol{u}}$.
3. If $(\boldsymbol{U}, \lambda)$ is an eigenpair of $\mathbb{A}$ then $(\hat{\boldsymbol{u}}, \lambda)$ is an eigenpair of $\hat{\boldsymbol{A}}$.
4. If $\mathbb{A}$ is symmetric then $\hat{\boldsymbol{A}}$ is symmetric.

Furthermore, one can also show that the following equalities hold:

$$\begin{aligned} \langle \mathbb{A}, \mathbb{B}\rangle = \langle \hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}\rangle, \quad \|\mathbb{A}\| = \|\hat{\boldsymbol{A}}\|, \\ \det(\mathbb{A}) = \det(\hat{\boldsymbol{A}}), \quad \text{tr}(\mathbb{A}) = \text{tr}(\hat{\boldsymbol{A}}), \\ \text{d}_E(\mathbb{A}, \mathbb{B}) = \text{d}_E(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}), \quad \text{d}_R(\mathbb{A}, \mathbb{B}) = \text{d}_R(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}). \end{aligned}$$

The last equality holds whenever $\mathbb{A}$ and $\mathbb{B}$ are in $\mathbb{S}\text{ym}^{++}(V)$.

4.3 Total Symmetry

A fourth-order tensor $\mathbb{A}$ is said to satisfy the *total symmetry* if in addition to satisfying the major and minor symmetries it satisfies

$$\mathbb{A}\boldsymbol{u} \otimes \boldsymbol{U} \otimes \boldsymbol{v} = \mathbb{A}\boldsymbol{u} \otimes \boldsymbol{U}^T \otimes \boldsymbol{v}, \quad \forall\, \boldsymbol{U} \in \mathrm{Lin}(V), \boldsymbol{u}, \boldsymbol{v} \in V.$$

In components, $A_{ijkl} = A_{\sigma(i)\sigma(j)\sigma(k)\sigma(l)}$, for any permutation σ of $[\![1 : n]\!]$. The set of fourth-order totally symmetric tensors will be denoted

$$\mathbb{S}\mathrm{ym}^{\mathrm{total}}(V) := \{\mathbb{A} \in \mathbb{L}\mathrm{in}(V) \mid \mathbb{A} \text{ satisfies total symmetry}\}.$$

Any fourth-order tensor $\mathbb{A}$ satisfying the major and minor symmetries can be decomposed, in a unique manner, into its totally symmetric part $\mathbb{A}^s$ and its asymmetric part $\mathbb{A}^a$:

$$\mathbb{A} = \mathbb{A}^s + \mathbb{A}^a.$$

The components of the totally symmetric and asymmetric parts are [3]

$$A^s_{ijkl} = \tfrac{1}{3}(A_{ijkl} + A_{ikjl} + A_{ilkj}), \tag{18a}$$

$$A^a_{ijkl} = \tfrac{1}{3}(2A_{ijkl} - A_{ikjl} - A_{ilkj}). \tag{18b}$$

A fourth-order tensor $\mathbb{A}$ is totally symmetric if $\mathbb{A} = \mathbb{A}^s$, or equivalently if $\mathbb{A}^a = \mathbb{O}$, where $\mathbb{O}$ denotes the fourth-order null tensor. Similarly, a fourth-order tensor $\mathbb{A}$ is asymmetric if $\mathbb{A} = \mathbb{A}^a$, or equivalently if $\mathbb{A}^s = \mathbb{O}$.

Using (18) and the definition of the trace, it is straightforward to show that if $\mathbb{A}$ is totally symmetric and $\mathbb{B}$ is asymmetric then

$$\mathrm{tr}(\mathbb{A}\mathbb{B}) = 0.$$

Before we close this section we note that there is an isomorphism (one-to-one linear correspondence) between totally symmetric fourth-order tensors and homogeneous polynomials of degree 4 [3]. Batchelor's chapter of this book uses this isomorphism. If a totally symmetric fourth-order tensor $\mathbb{A}$ is traceless, that is, $\mathrm{tr}(\mathbb{A}) = 0$, then it is isomorphic to a harmonic polynomial of degree 4. For this reason, a totally symmetric and traceless tensor is called *harmonic tensor*. We also note that there is an isomorphism between harmonic tensors and spherical harmonics [2, 3].

5 Fourth-Order Tensors in 3D Space

From this section on we restrict attention to fourth-order tensors in three-dimensional space that satisfy the minor symmetry. As it has been customary in three-dimensional elasticity literature, we use a Voigt contracted notation [18], which is described by the indicial map (9). Let $\mathbb{A}$ be a fourth-order tensor satisfying the minor symmetry and let $\hat{\boldsymbol{A}}$ be the corresponding second-order tensor. The matrix of components

$$\begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} & \hat{A}_{13} & \hat{A}_{14} & \hat{A}_{15} & \hat{A}_{16} \\ \hat{A}_{21} & \hat{A}_{22} & \hat{A}_{23} & \hat{A}_{24} & \hat{A}_{25} & \hat{A}_{26} \\ \hat{A}_{31} & \hat{A}_{32} & \hat{A}_{33} & \hat{A}_{34} & \hat{A}_{35} & \hat{A}_{36} \\ \hat{A}_{41} & \hat{A}_{42} & \hat{A}_{43} & \hat{A}_{44} & \hat{A}_{45} & \hat{A}_{46} \\ \hat{A}_{51} & \hat{A}_{52} & \hat{A}_{53} & \hat{A}_{54} & \hat{A}_{55} & \hat{A}_{56} \\ \hat{A}_{61} & \hat{A}_{62} & \hat{A}_{63} & \hat{A}_{64} & \hat{A}_{65} & \hat{A}_{66} \end{bmatrix} \tag{19}$$

of $\hat{\boldsymbol{A}}$ with respect to the basis $\{\hat{\boldsymbol{e}}_\alpha\}_{1\leq\alpha\leq 6}$ is given in terms of the components A_{ijkl} of $\mathbb{A}$ with respect to the basis $\{\boldsymbol{e}_i\}_{1\leq i\leq 3}$ as

$$\begin{bmatrix} A_{1111} & A_{1122} & A_{1133} & \sqrt{2}A_{1123} & \sqrt{2}A_{1113} & \sqrt{2}A_{1112} \\ A_{2211} & A_{2222} & A_{2233} & \sqrt{2}A_{2223} & \sqrt{2}A_{2213} & \sqrt{2}A_{2212} \\ A_{3311} & A_{3322} & A_{3333} & \sqrt{2}A_{3323} & \sqrt{2}A_{3313} & \sqrt{2}A_{3312} \\ \sqrt{2}A_{2311} & \sqrt{2}A_{2322} & \sqrt{2}A_{2333} & 2A_{2323} & 2A_{2313} & 2A_{2312} \\ \sqrt{2}A_{1311} & \sqrt{2}A_{1322} & \sqrt{2}A_{1333} & 2A_{1323} & 2A_{1313} & 2A_{1312} \\ \sqrt{2}A_{1211} & \sqrt{2}A_{1222} & \sqrt{2}A_{1233} & 2A_{1223} & 2A_{1213} & 2A_{1212} \end{bmatrix}. \tag{20}$$

If, in addition, $\mathbb{A}$ satisfies the major symmetry then $\hat{\boldsymbol{A}}$ is symmetric. Thus a fourth-order tensor $\mathbb{A}$ that satisfies the major and minor symmetries requires 21 independent constants (the 21 independent components of $\hat{\boldsymbol{A}}$ which is symmetric).

As an example of such tensors we mention elasticity tensors, $\mathbb{C}$ or $\mathbb{S}$, that linearly relate stress $\boldsymbol{T}$ and infinitesimal strain $\boldsymbol{E}$ according to

$$\boldsymbol{T} = \mathbb{C}\boldsymbol{E}, \text{ or } \boldsymbol{E} = \mathbb{S}\boldsymbol{T}.$$

The tensors $\mathbb{C}$ and $\mathbb{S}$, which are inverses of one another, are called the stiffness tensor and compliance tensors, respectively. In addition to satisfying the major and minor symmetries, elasticity tensors must be positive-definite.

A fourth-order tensor $\mathbb{A}$ is totally symmetric if besides satisfying the minor and major symmetries it further satisfies

$$A_{1122} = A_{1212}, \quad A_{1133} = A_{1313}, \quad A_{2233} = A_{2323}, \tag{21a}$$
$$A_{1123} = A_{1213}, \quad A_{2213} = A_{1223}, \quad A_{3312} = A_{1323}. \tag{21b}$$

In elasticity literature (see, e.g. [6, 10]), these conditions are known as Cauchy relations, and the theory of elasticity based on totally symmetric tensors is called "Cauchy's elasticity theory" as opposed to the classical "Green's elasticity theory." As other examples of totally symmetric fourth-order tensors, we mention fourth-order orientation tensors and fourth-order diffusion tensors to be discussed in Sects. 7 and 8.

Comparison of (19) and (20) yields that total symmetry requires that the components of $\hat{\boldsymbol{A}}$ further satisfy the following six relations

$$\hat{A}_{44} = 2\hat{A}_{23}, \quad \hat{A}_{55} = 2\hat{A}_{13}, \quad \hat{A}_{66} = 2\hat{A}_{12}, \tag{22a}$$
$$\hat{A}_{45} = \sqrt{2}\hat{A}_{36}, \quad \hat{A}_{46} = \sqrt{2}\hat{A}_{25}, \quad \hat{A}_{56} = \sqrt{2}\hat{A}_{14}, \tag{22b}$$

and hence, a totally symmetric tensor requires at most 15 independent coefficients.

Let $\hat{\boldsymbol{A}}$, $\hat{\boldsymbol{A}}^s$, and $\hat{\boldsymbol{A}}^a$ be the second-order representation of $\mathbb{A}$, $\mathbb{A}^s$, and $\mathbb{A}^a$, respectively. Then $\hat{\boldsymbol{A}} = \hat{\boldsymbol{A}}^s + \hat{\boldsymbol{A}}^a$, where

$$\hat{\boldsymbol{A}} = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} & \hat{A}_{13} & \hat{A}_{14} & \hat{A}_{15} & \hat{A}_{16} \\ & \hat{A}_{22} & \hat{A}_{23} & \hat{A}_{24} & \hat{A}_{25} & \hat{A}_{26} \\ & & \hat{A}_{33} & \hat{A}_{34} & \hat{A}_{35} & \hat{A}_{36} \\ & \text{sym} & & \hat{A}_{44} & \hat{A}_{45} & \hat{A}_{46} \\ & & & & \hat{A}_{55} & \hat{A}_{56} \\ & & & & & \hat{A}_{66} \end{bmatrix},$$

$$3\hat{\boldsymbol{A}}^s = \begin{bmatrix} 3\hat{A}_{11} & \hat{A}_{12}+\hat{A}_{66} & \hat{A}_{13}+\hat{A}_{55} & \hat{A}_{14}+\sqrt{2}\hat{A}_{56} & 3\hat{A}_{15} & 3\hat{A}_{16} \\ & 3\hat{A}_{22} & \hat{A}_{23}+\hat{A}_{44} & 3\hat{A}_{24} & \hat{A}_{25}+\sqrt{2}\hat{A}_{46} & 3\hat{A}_{26} \\ & & 3\hat{A}_{33} & 3\hat{A}_{34} & 3\hat{A}_{35} & \hat{A}_{36}+\sqrt{2}\hat{A}_{45} \\ & \text{sym} & & 2(\hat{A}_{44}+\hat{A}_{23}) & 2(\hat{A}_{45}+\tfrac{1}{\sqrt{2}}\hat{A}_{36}) & 2(\hat{A}_{46}+\tfrac{1}{\sqrt{2}}\hat{A}_{25}) \\ & & & & 2(\hat{A}_{55}+\hat{A}_{13}) & 2(\hat{A}_{56}+\tfrac{1}{\sqrt{2}}\hat{A}_{14}) \\ & & & & & 2(\hat{A}_{66}+\hat{A}_{12}) \end{bmatrix}$$

and

$$3\hat{\boldsymbol{A}}^a = \begin{bmatrix} 0 & 2\hat{A}_{12}-\hat{A}_{66} & 2\hat{A}_{13}-\hat{A}_{55} & 2\hat{A}_{14}-\sqrt{2}\hat{A}_{56} & 0 & 0 \\ & 0 & 2\hat{A}_{23}-\hat{A}_{44} & 0 & 2\hat{A}_{25}-\sqrt{2}\hat{A}_{46} & 0 \\ & & 0 & 0 & 0 & 2\hat{A}_{36}-\sqrt{2}\hat{A}_{45} \\ & \text{sym} & & \hat{A}_{44}-2\hat{A}_{23} & \hat{A}_{45}-\sqrt{2}\hat{A}_{36} & \hat{A}_{46}-\sqrt{2}\hat{A}_{25} \\ & & & & \hat{A}_{55}-2\hat{A}_{13} & \hat{A}_{56}-\sqrt{2}\hat{A}_{14} \\ & & & & & \hat{A}_{66}-2\hat{A}_{12} \end{bmatrix}.$$

It follows that the trace of the totally symmetric part of $\mathbb{A}$ is

$$\operatorname{tr}(\mathbb{A}^s) = A_{1111}+A_{2222}+A_{3333}+\tfrac{2}{3}(A_{2233}+A_{1133}+A_{1122})+\tfrac{1}{3}(A_{2323}+A_{1313}+A_{1212})$$

and the trace of the asymmetric part of $\mathbb{A}$ is

$$\operatorname{tr}(\mathbb{A}^a) = \tfrac{2}{3}(A_{2233} - A_{2323} + A_{1133} - A_{1313} + A_{1122} - A_{1212}).$$

6 Material Symmetries

So far we have discussed various notions of symmetries of fourth-order tensors. These symmetries, which can be formulated as invariance of the components of the tensor with respect to an interchange of indices, are independent of the symmetry properties of the material which the tensor models. In this section we discuss the restrictions placed by the material symmetries on fourth-order tensors.

Given a fourth-order tensor $\mathbb{A}$, let $\mathscr{G}(\mathbb{A})$ denote the set of all special orthogonal second-order tensors $\boldsymbol{Q}$ such that

$$\mathbb{A}(\boldsymbol{Q}^T\boldsymbol{U}\boldsymbol{Q}) = \boldsymbol{Q}^T(\mathbb{A}\boldsymbol{U})\boldsymbol{Q},\ \forall\ \boldsymbol{U} \in \operatorname{Sym}(V). \tag{23}$$

It can be easily verified that $\mathscr{G}(\mathbb{A})$ is a subgroup of $SO(3)$, the group of special orthogonal transformations. The subgroup $\mathscr{G}(\mathbb{A})$ is called material symmetry

group of $\mathbb{A}$, which reflects the symmetries of a material property that is represented by the fourth-order tensor $\mathbb{A}$.

For a given fourth-order tensor $\mathbb{A}$, when (23) is satisfied for all $\boldsymbol{Q} \in SO(3)$, that is, when the material symmetry group $\mathscr{G}(\mathbb{A})$ is $SO(3)$, the tensor $\mathbb{A}$ is said to be isotropic. On the other hand, when there is no special orthogonal tensor, except the identity tensor, such that (23) is satisfied, that is, $\mathscr{G}(\mathbb{A}) = \{\boldsymbol{I}\}$, the tensor $\mathbb{A}$ is said to be triclinic (absence of material symmetry). The other classes of material symmetry are crystallographic symmetries (cubic, trigonal, tetragonal, orthotropic, and monoclinic) reflected by discrete subgroups of $SO(3)$, and transversely isotropy reflected by a one-parameter subgroup of $SO(3)$ consisting of all rotations about a fixed axis [8].

Next, we discuss some of these classes when the underlying fourth-order tensor satisfies the minor and major symmetries, and the total symmetry. For completeness, we give in Fig. 1 a diagram showing a sequence of increasing material symmetry classes from triclinic to isotropic.

6.1 Isotropy

The general form of an isotropic and symmetric fourth-order tensor is [14]

$$\mathbb{A}^{\text{iso}} = a\mathbb{J} + b\mathbb{K}, \tag{24}$$

where a and b are real numbers, and $\mathbb{J}$ and $\mathbb{K}$ are the two linearly independent symmetric fourth-order tensors defined by

$$\mathbb{J} = \tfrac{1}{3}\mathbf{I} \otimes \mathbf{I}, \qquad \mathbb{K} = \mathbb{I} - \mathbb{J}, \tag{25}$$

where $\mathbb{I}$ is the fourth-order identity tensor. The component forms of $\mathbb{I}$ and $\mathbb{J}$ are $I_{ijkl} = \frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$ and $J_{ijkl} = \frac{1}{3}\delta_{ij}\delta_{kl}$.

The tensor $\mathbb{A}^{\text{iso}}$ is positive definite when a and b are positive. The totally symmetric part and asymmetric part of $\mathbb{A}^{\text{iso}}$ are

$$(\mathbb{A}^{\text{iso}})^s = \tfrac{1}{3}(a + 2b)\mathbb{I}^s \text{ and } (\mathbb{A}^{\text{iso}})^a = \tfrac{1}{3}(-2a + 5b)\mathbb{I}^a,$$

where $\mathbb{I}^s$ and $\mathbb{I}^a$ are the symmetric and asymmetric parts of the identity tensor $\mathbb{I}$.

In the second-order tensor representation, we have

$$\hat{\boldsymbol{A}}^{\text{iso}} = a\hat{\boldsymbol{J}} + b\hat{\boldsymbol{K}},$$

where

$$\hat{\boldsymbol{J}} = \tfrac{1}{3}\begin{bmatrix} 1&1&1&0&0&0\\ 1&1&1&0&0&0\\ 1&1&1&0&0&0\\ 0&0&0&0&0&0\\ 0&0&0&0&0&0\\ 0&0&0&0&0&0 \end{bmatrix}, \quad \hat{\boldsymbol{K}} = \tfrac{1}{3}\begin{bmatrix} 2&-1&-1&0&0&0\\ -1&2&-1&0&0&0\\ -1&-1&2&0&0&0\\ 0&0&0&3&0&0\\ 0&0&0&0&3&0\\ 0&0&0&0&0&3 \end{bmatrix}.$$

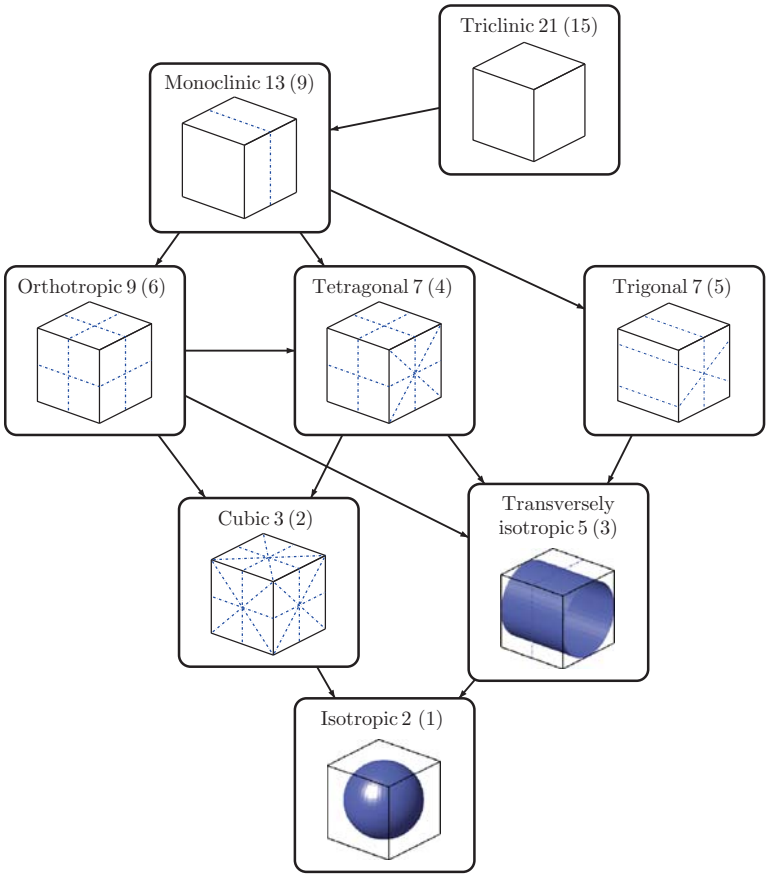


Fig. 1. Sequence of increasing material symmetry classes from triclinic to isotropic. For each material symmetry class the number of independent coefficients is listed. The numbers in parentheses represent the number of independent coefficients for totally symmetric tensors. Dashed lines represent the intersections of the planes of symmetry with the material element cube.

The totally symmetric part of $\hat{\boldsymbol{A}}^{\text{iso}}$ is $(\hat{\boldsymbol{A}}^{\text{iso}})^s = \frac{1}{3}(a+2b)\hat{\boldsymbol{I}}^s$, and the asymmetric part is $(\hat{\boldsymbol{A}}^{\text{iso}})^a = \frac{1}{3}(-2a+5b)\hat{\boldsymbol{I}}^a$, where $\hat{\boldsymbol{I}}^s$ and $\hat{\boldsymbol{I}}^a$ are the symmetric and asymmetric parts of the identity tensor $\hat{\boldsymbol{I}}$ and are given by

$$\hat{\boldsymbol{I}}^s = \frac{1}{3}\begin{bmatrix} 3&1&1&0&0&0\\ 1&3&1&0&0&0\\ 1&1&3&0&0&0\\ 0&0&0&2&0&0\\ 0&0&0&0&2&0\\ 0&0&0&0&0&2 \end{bmatrix}, \quad \hat{\boldsymbol{I}}^a = \frac{1}{3}\begin{bmatrix} 0&-1&-1&0&0&0\\ -1&0&-1&0&0&0\\ -1&-1&0&0&0&0\\ 0&0&0&1&0&0\\ 0&0&0&0&1&0\\ 0&0&0&0&0&1 \end{bmatrix}. \tag{26}$$

Therefore, the set of isotropic fourth-order tensors is a two-dimensional linear space spanned by a totally symmetric tensor and an asymmetric tensor.

A totally symmetric tensor $\mathbb{A}$ is isotropic if and only if $\mathbb{A} = \lambda \mathbb{I}^s$ for some real λ. In view of (26), the components of its second-order tensor representation $\hat{\boldsymbol{A}}$ must satisfy

$$\tfrac{1}{3}\hat{A}_{11} = \tfrac{1}{3}\hat{A}_{22} = \tfrac{1}{3}\hat{A}_{33} = \tfrac{1}{2}\hat{A}_{44} = \tfrac{1}{2}\hat{A}_{55} = \tfrac{1}{2}\hat{A}_{66} = \hat{A}_{23} = \hat{A}_{13} = \hat{A}_{12},$$

and all other components (not deduced by symmetry) are 0. Which when translated in terms of the components of $\mathbb{A}$ shows that a totally symmetric tensor $\mathbb{A}$ is isotropic if and only if its components in any Cartesian coordinate system satisfy

$$A_{1111} = A_{2222} = A_{3333} = 3A_{2223} = 3A_{1133} = 3A_{1122},$$

and all other components (not deduced by permutation of the indices) are 0.

6.2 Cubic Symmetry

A fourth-order tensor $\mathbb{A}^{\text{cub}}$ is said to have the cubic symmetry if its symmetry group $\mathscr{G}(\mathbb{A}^{\text{cub}})$ is the (finite) subgroup of the rotation group $SO(3)$, which leaves a cube invariant. This subgroup contains 24 elements: (i) the identity, (ii) eight rotations by $2\pi/3$ about axes along the diagonals of the cube, (iii) three rotations by π about axes perpendicular to face centers of the cube, (iv) six rotations by π about axes that pass through midpoints of opposite edges of the cube, (v) six rotations by $\pi/4$ about axes perpendicular to face centers of the cube.

A general symmetric fourth-order tensor with a cubic symmetry is given by [14, 19]

$$\mathbb{A}^{\text{cub}} = a\mathbb{J} + b\mathbb{L} + c\mathbb{M}. \tag{27}$$

Here $\mathbb{J}$, $\mathbb{L}$, and $\mathbb{M}$ are three linearly independent elementary fourth-order tensors. The fourth-order tensor $\mathbb{A}^{\text{cub}}$ is positive definite when $a > 0$, $b > 0$, and $c > 0$. The fourth-order tensor $\mathbb{A}^{\text{cub}}$ is totally symmetric when $b = \frac{2}{3}(a - c)$.

For this symmetry class, the number of independent constants for tensors that satisfy the major and minor symmetries is 3, and the number of independent constants for totally symmetric tensors is 2. In a Cartesian coordinate system whose axes are perpendicular to face centers of the cube, the matrix of components of $\mathbb{A}^{\text{cub}}$ is

$$\hat{A}^{\text{cub}} = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} & \hat{A}_{12} & 0 & 0 & 0 \\ \hat{A}_{12} & \hat{A}_{11} & \hat{A}_{12} & 0 & 0 & 0 \\ \hat{A}_{12} & \hat{A}_{12} & \hat{A}_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & \hat{A}_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \hat{A}_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & \hat{A}_{44} \end{bmatrix}.$$

6.3 Transverse Isotropy

A fourth-order tensor $\mathbb{A}^{\text{tis}}$ is said to be transversely isotropic if there is a preferred direction $\boldsymbol{n}$ such that $\mathscr{G}(\mathbb{A}^{\text{tis}})$ is the one-parameter subgroup of $SO(3)$ of rotations about $\boldsymbol{n}$. The plane normal to the preferred direction is called plane of transverse isotropy. This transverse plane has infinite planes of symmetry and thus, within this plane, the material properties are the same in all directions. Fibrous materials with fibers having circular cross sections are examples of transversely isotropic materials.

A general transversely isotropic and symmetric fourth-order tensor is given by [14, 19]

$$\mathbb{A}^{\text{tis}} = a\mathbb{E}_1 + b\mathbb{E}_2 + c(\mathbb{E}_3 + \mathbb{E}_4) + f\mathbb{F} + g\mathbb{G}, \tag{28}$$

where $\mathbb{E}_1$, $\mathbb{E}_2$, $\mathbb{E}_3$, $\mathbb{E}_4$, $\mathbb{G}$, and $\mathbb{G}$ are linearly independent elementary fourth-order tensors. The fourth-order tensor $\mathbb{A}^{\text{tis}}$ is positive definite when $ab - c^2 > 0$, $f > 0$, and $g > 0$. The fourth-order tensor $\mathbb{A}^{\text{tis}}$ is totally symmetric when $f = \frac{1}{2}b$ and $g = \sqrt{2}c$. Thus, for this type of material symmetry, the number of independent constants for tensors that satisfy the major and minor symmetries reduces to 5, and the number of independent constants for totally symmetric tensors reduces to 3. In a Cartesian coordinate system with $\boldsymbol{n}$ as the x_3-axis, the matrix of components of $\mathbb{A}^{\text{tis}}$ is

$$\hat{A}^{\text{tis}} = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} & \hat{A}_{13} & 0 & 0 & 0 \\ \hat{A}_{12} & \hat{A}_{11} & \hat{A}_{13} & 0 & 0 & 0 \\ \hat{A}_{13} & \hat{A}_{13} & \hat{A}_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & \hat{A}_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \hat{A}_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & \hat{A}_{11} - \hat{A}_{12} \end{bmatrix}.$$

6.4 Orthotropy

A fourth-order tensor $\mathbb{A}^{\text{ort}}$ is said to be orthotropic if its symmetry group $\mathscr{G}(\mathbb{A}^{\text{ort}})$ is generated by the rotations through angle π about three mutually orthogonal axes. The number of independent constants for an orthotropic tensor satisfying the major and minor symmetries is 9 and the number of independent constants for a totally symmetric and orthotropic tensor is 6. In a Cartesian coordinate, whose axes coincide with the axes of rotations that generate the symmetry group, the matrix of components of $\mathbb{A}^{\text{ort}}$ is

$$\hat{A}^{\text{ort}} = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} & \hat{A}_{13} & 0 & 0 & 0 \\ \hat{A}_{12} & \hat{A}_{22} & \hat{A}_{23} & 0 & 0 & 0 \\ \hat{A}_{13} & \hat{A}_{23} & \hat{A}_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & \hat{A}_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \hat{A}_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & \hat{A}_{66} \end{bmatrix}.$$

7 Orientation Distributions and Orientation Tensors

The microscopic description of fiber orientation in fibrous materials, which are made of a large collection of rod-like objects, is embodied in the orientation distribution function (ODF). The orientation distribution is a function ρ defined on the unit sphere S^2 of $\mathbb{R}^3$ and satisfies the antipodal-symmetry property $\rho(\boldsymbol{n}) = \rho(-\boldsymbol{n})$. It is normalized so that $\int_{S^2} \rho(\boldsymbol{n})\, d\sigma = 1$, where $d\sigma$ is area element in S^2.

Since the work of Advani and Tucker [1], orientation tensors of even orders have been used to describe the orientation of fibers at the macroscopic scale. The second- and fourth-order orientation tensors are the averages, with respect to the orientation distribution function ρ, over all orientations $\boldsymbol{n} \in S^2$:

$$\langle \boldsymbol{N} \rangle_\rho := \int_{S^2} \rho(\boldsymbol{n}) \boldsymbol{N}\, d\sigma, \quad \langle \mathbb{N} \rangle_\rho := \int_{S^2} \rho(\boldsymbol{n}) \mathbb{N}\, d\sigma,$$

where, for brevity $\boldsymbol{N} = \boldsymbol{n} \otimes \boldsymbol{n}$ and $\mathbb{N} = \boldsymbol{N} \otimes \boldsymbol{N}$. For the definitions, the second-order orientation tensor is symmetric, and the fourth-order orientation tensor is totally symmetric. Both have trace one and all their components are less than or equal to one in absolute value. Another important property is that the contraction, with respect to any two indices, of the fourth-order orientation tensor gives the second-order orientation tensor.

In what follows, we give the expressions of the orientation tensors $\langle \boldsymbol{N} \rangle_\rho$ and $\langle \mathbb{N} \rangle_\rho$ for orientation distribution functions with some prescribed material symmetry classes.

7.1 Isotropic Orientation Distribution

The isotropic orientation distribution function is given by

$$\rho^{\text{iso}}(\theta, \phi) = \frac{1}{4\pi}. \tag{29}$$

This distribution corresponds to fibers that are randomly oriented. Let $\boldsymbol{n} = (\cos\theta \sin\phi,\ \sin\theta \sin\phi,\ \cos\phi)^T$ be a generic vector on the unit sphere S^2 with $0 \leq \theta \leq 2\pi$ and $0 \leq \phi \leq \pi$. The isotropic orientation averages of $\boldsymbol{N}$ and $\mathbb{N}$ are given by

$$\langle \boldsymbol{N} \rangle^{\text{iso}} := \int_{S^2} \rho^{\text{iso}} \boldsymbol{N}\, d\sigma = \tfrac{1}{3} \boldsymbol{I},$$

and

$$\langle \mathbb{N} \rangle^{\text{iso}} := \int_{S^2} \rho^{\text{iso}} \mathbb{N}\, d\sigma = \frac{1}{5} \mathbb{I}^s.$$

7.2 Transversely Isotropic Orientation Distributions

The widely used uni-modal orientation distribution is the Fischer distribution with modal vector $\boldsymbol{m} \in S^2$ and concentration parameter $\kappa > 0$ that is given by (see, e.g., [11, 20])

$$\rho_{\kappa,\boldsymbol{m}}(\boldsymbol{n}) = \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa \boldsymbol{m}^T \boldsymbol{n}),$$

where $\boldsymbol{n} \in S^2$. However, this distribution is not invariant under the inversion of directions ($\boldsymbol{n} \mapsto -\boldsymbol{n}$). It can be modified in such a way that it becomes invariant under direction inversion. A plausible modification is given by

$$\rho^m_{\kappa,\boldsymbol{m}}(\boldsymbol{n}) = \frac{1}{a(\kappa)} \exp(\kappa(\boldsymbol{m}^T \boldsymbol{n})^2),$$

which is a special case of the well known Bingham (multimodal) distribution for axial data

$$\rho_{\boldsymbol{K}}(\boldsymbol{n}) = \frac{1}{b(\boldsymbol{K})} \exp(\boldsymbol{n}^T \boldsymbol{K} \boldsymbol{n}),$$

where $\boldsymbol{K}$ is a symmetric matrix and $b(\boldsymbol{K})$ is a normalization constant. The distribution $\rho^m_{\kappa,\boldsymbol{m}}(\boldsymbol{n})$ is transversely isotropic and is parametrized by the concentration parameter κ. For the purpose of getting analytical expressions for the second- and fourth-order orientation tensors, we shall, however, use another family of transversely isotropic orientation distribution functions.

The family (indexed by a positive integer n) of orientation distribution functions given by

$$\rho^{\text{tis}}_{n,3}(\theta, \phi) = \frac{2n+1}{4\pi} \cos^{2n} \phi \tag{30}$$

is independent of the azimuthal angle θ, and hence is transversely isotropic along the $\boldsymbol{e}_3$-axis (Fig. 2). The positive integer n acts as the concentration parameter of the modified Fisher distribution. The corresponding second-order orientation tensor is

$$\langle \boldsymbol{N} \rangle^{\text{tis}}_{n,3} := \int_{S^2} \rho^{\text{tis}}_{n,3} \boldsymbol{N} \, d\sigma = \frac{1}{2n+3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2n+1 \end{bmatrix}.$$

The corresponding fourth-order orientation tensor,

$$\langle \mathbb{N} \rangle^{\text{tis}}_{n,3} := \int_{S^2} \rho^{\text{tis}}_{n,3} \mathbb{N} \, d\sigma,$$

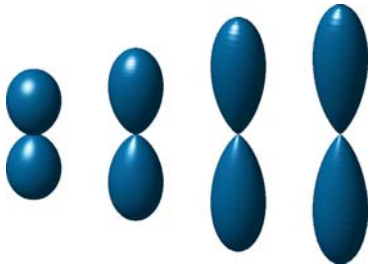


Fig. 2. Plots of the transversely isotropic ODF (30) for $n = 1$, 2, 4, and 5.

has the 6D second-order tensor representation

$$\frac{1}{(2n+3)(2n+5)}\begin{bmatrix} 3 & 1 & 2n+1 & 0 & 0 & 0 \\ 1 & 3 & 2n+1 & 0 & 0 & 0 \\ 2n+1 & 2n+1 & (2n+1)(2n+3) & 0 & 0 & 0 \\ 0 & 0 & 0 & 2(2n+1) & 0 & 0 \\ 0 & 0 & 0 & 0 & 2(2n+1) & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}.$$

Using a change of variables we can deduce that the family of orientation distribution functions

$$\rho^{\text{tis}}_{n,1}(\theta,\phi) = \frac{2n+1}{4\pi}\cos^{2n}\theta\sin^{2n}\phi \tag{31}$$

is transversely isotropic along the $\boldsymbol{e}_1$-axis. Similarly, the family of orientation distribution functions

$$\rho^{\text{tis}}_{n,2}(\theta,\phi) = \frac{2n+1}{4\pi}\sin^{2n}\theta\sin^{2n}\phi \tag{32}$$

is transversely isotropic along the $\boldsymbol{e}_2$-axis.

7.3 Orthotropic Orientation Distributions

A family of orthotropic orientation distribution functions can be obtained by combining, for example, the two transversely isotropic orientation distribution functions (31) and (32) (Fig. 3):

$$\rho^{\text{ort}}_{n,3}(\theta,\phi) = \frac{2n+1}{8\pi}(\cos^{2n}\theta + \sin^{2n}\theta)\sin^{2n}\phi. \tag{33}$$

The corresponding second-order orientation tensor is

$$\langle \boldsymbol{N}\rangle^{\text{ort}}_{n,3} := \int_{S^2}\rho^{\text{ort}}_{n,3}\boldsymbol{N}\,d\sigma = \frac{1}{2n+3}\begin{bmatrix} n+1 & 0 & 0 \\ 0 & n+1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

Fig. 3. Plots of the orthotropic ODF (33) for $n = 1$, 2, 4, and 5.

and the corresponding fourth-order orientation tensor

$$\langle \mathbb{N} \rangle_{n,3}^{\text{ort}} := \int_{S^2} \rho_{n,3}^{\text{ort}} \mathbb{N} \, d\sigma,$$

has the 6D second-order tensor representation (modulo the multiplicative factor $\dfrac{1}{(2n+3)(2n+5)}$)

$$\begin{bmatrix} 2(n+1)^2+1 & 2n+1 & n+1 & 0 & 0 & 0 \\ 2n+1 & 2(n+1)^2+1 & n+1 & 0 & 0 & 0 \\ n+1 & n+1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2(n+1) & 0 & 0 \\ 0 & 0 & 0 & 0 & 2(n+1) & 0 \\ 0 & 0 & 0 & 0 & 0 & 2(2n+1) \end{bmatrix}.$$

We note that at the second-order level, there is no distinction between transversely isotropic and orthotropic orientation tensors, whereas at the fourth-order level the structures of these types of orientation tensors differ.

Another family of orthotropic orientation distribution functions can be obtained by considering two transversely isotropic orientation distribution functions crossing at an angle ψ, for example, the ODF (31) rotated by $\pm\frac{1}{2}\psi$ about the $\boldsymbol{e}_3$-axis:

$$\rho_{n,\psi}^{\text{ort}}(\theta,\phi) = \frac{2n+1}{8\pi}(\cos^{2n}(\theta+\tfrac{1}{2}\psi) + \cos^{2n}(\theta-\tfrac{1}{2}\psi))\sin^{2n}\phi. \tag{34}$$

The corresponding second-order orientation tensor is

$$\langle \boldsymbol{N} \rangle_{n,\psi}^{\text{ort}} := \int_{S^2} \rho_{n,\psi}^{\text{ort}} \boldsymbol{N} \, d\sigma = \frac{1}{2n+3} \begin{bmatrix} 1+n(1+\cos\psi) & 0 & 0 \\ 0 & 1+n(1-\cos\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and the corresponding fourth-order orientation tensor

$$\langle \mathbb{N} \rangle_{n,\psi}^{\text{ort}} := \int_{S^2} \rho_{n,\psi}^{\text{ort}} \mathbb{N} \, d\sigma$$

has the 6D second-order tensor representation

$$\frac{1}{(2n+3)(2n+5)} \begin{bmatrix} a(n,\psi) & f(n,\psi) & e(n,\psi) & 0 & 0 & 0 \\ f(n,\psi) & b(n,\psi) & d(n,\psi) & 0 & 0 & 0 \\ e(n,\psi) & d(n,\psi) & c(n,\psi) & 0 & 0 & 0 \\ 0 & 0 & 0 & 2d(n,\psi) & 0 & 0 \\ 0 & 0 & 0 & 0 & 2e(n,\psi) & 0 \\ 0 & 0 & 0 & 0 & 0 & 2f(n,\psi) \end{bmatrix},$$

where

$$
\begin{aligned}
a(n,\psi) &= 1 + 2(n+1)^2 + 2n(n+2)\cos\psi - n(n-1)\sin^2\psi,\\
b(n,\psi) &= 1 + 2(n+1)^2 - 2n(n+2)\cos\psi - n(n-1)\sin^2\psi,\\
c(n,\psi) &= 3,\\
d(n,\psi) &= 1 + n(1-\cos\psi),\\
e(n,\psi) &= 1 + n(1+\cos\psi),\\
d(n,\psi) &= 1 + 2n + n(n-1)\sin^2\psi.
\end{aligned}
$$

Note that, as the orientation tensor has a unit trace, $a(n,\psi)+b(n,\psi)+c(n,\psi) + 2(d(n,\psi)+e(n,\psi)+f(n,\psi)) = (2n+3)(2n+5)$.

7.4 Cubic Orientation Distributions

A family of cubic orientation distribution functions can be obtained by combining, for example, the three transversely isotropic orientation distribution functions (30), (31), and (32) (Fig. 4):

$$
\rho_n^{\mathrm{cub}}(\theta,\phi) = \frac{2n+1}{12\pi}(\cos^{2n}\theta\sin^{2n}\phi + \sin^{2n}\theta\sin^{2n}\phi + \cos^{2n}\phi). \tag{35}
$$

Note that for $n = 1$, the ODF (35) degenerates to the isotropic ODF (29). The corresponding second-order orientation tensor is

$$
\langle \boldsymbol{N}\rangle_n^{\mathrm{cub}} := \int_{S^2} \rho_n^{\mathrm{cub}} \boldsymbol{N}\, d\sigma = \frac{1}{3}\boldsymbol{I},
$$

and the corresponding fourth-order orientation tensor

$$
\langle \mathbb{N}\rangle_n^{\mathrm{cub}} := \int_{S^2} \rho_n^{\mathrm{cub}} \mathbb{N}\, d\sigma,
$$

has the 6D second-order tensor representation (modulo the multiplicative factor $\dfrac{1}{(2n+3)(2n+5)}$)

Fig. 4. Plots of the cubic ODF (35) for $n = 1$, 2, 4, and 5.

$$\begin{bmatrix} 4(n+1)^2+5 & 4n+3 & 4n+3 & 0 & 0 & 0 \\ 4n+3 & 4(n+1)^2+5 & 4n+3 & 0 & 0 & 0 \\ 4n+3 & 4n+3 & 4(n+1)^2+5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2(4n+3) & 0 & 0 \\ 0 & 0 & 0 & 0 & 2(4n+3) & 0 \\ 0 & 0 & 0 & 0 & 0 & 2(4n+3) \end{bmatrix}.$$

We remark that for the ODF of cubic symmetry, the second-order tensor is isotropic and that one has to use the fourth-order orientation tensor to detect this symmetry.

In a similar way, by combining three transversely isotropic distribution functions, whose axes are not necessarily orthogonal, we can obtain orientation distribution functions of trigonal, tetragonal, monoclinic, or triclinic symmetries.

8 Fourth-Order Diffusion Tensors

8.1 Generalized Stejskal–Tanner Equation

In diffusion tensor imaging of biological tissues, the second-order diffusion tensor $\boldsymbol{D}$ is estimated using the Stejskal–Tanner equation [5]

$$\ln \frac{s_k}{s_0} = -b_k \boldsymbol{g}_k^T \boldsymbol{D} \boldsymbol{g}_k,$$

where $\boldsymbol{g}_k$ is the direction of the kth magnetic gradient, s_0 is the original signal intensity, s_k is the attenuated signal intensity after gradient $\boldsymbol{g}_k$ is applied, and b_k is a diffusion weighting factor. The second-order diffusion model has been used with success in characterizing the anisotropy and fiber direction within an imaging volume. However, this model has some limitations when several fibers with different orientations occur within a single voxel. To surmount these shortcomings, Tuch et al. [17] proposed the use of diffusion imaging with very large diffusion gradient directions known as high angular resolution imaging (HARDI). In [15], a method of computing the components of a high-order diffusion tensor from the diffusion profiles obtained from HARDI measurements is given. Their method is based on the following generalization of the Stejskal–Tanner equation:

$$\ln \frac{s(\boldsymbol{g})}{s_0} = -b D_{ijkl} g_i g_j g_k g_l, \tag{36}$$

where D_{ijkl} are the components of a fourth-order tensor $\mathbb{D}$ and g_i are the components of the gradient direction $\boldsymbol{g}$. In direct notation, the above equation writes

$$\ln \frac{s(\boldsymbol{g})}{s_0} = -b\mathrm{tr}(\mathbb{D}\mathbb{G}), \tag{37}$$

where $\mathbb{G}$ denotes the totally symmetric tensor $\mathbb{G} = \boldsymbol{g}\otimes\boldsymbol{g}\otimes\boldsymbol{g}\otimes\boldsymbol{g}$. The diffusion in the direction of $\boldsymbol{g}$ is given by

$$D(\boldsymbol{g}) = -\frac{1}{b}\ln\frac{s(\boldsymbol{g})}{s_0} = \mathrm{tr}(\mathbb{D}\mathbb{G}).$$

In terms of the second-order tensor representation of fourth-order tensors, the generalized Stejskal–Tanner equation (37) takes the form

$$\ln \frac{s(\boldsymbol{g})}{s_0} = -b\mathrm{tr}(\hat{\boldsymbol{D}}\hat{\boldsymbol{G}}) = -b\hat{D}_{\alpha\beta}\hat{G}_{\alpha\beta}, \tag{38}$$

where $\hat{\boldsymbol{D}}$ and $\hat{\boldsymbol{G}}$ are the second-order tensors corresponding to $\mathbb{D}$ and $\mathbb{G}$, respectively. Therefore, with this formalism the methods that have been used to estimate the second-order diffusion tensor can be adapted to devise rigorous methods for estimating the fourth-order diffusion tensor.

8.2 Diffusion Tensor Order Reduction

The mean diffusivity $D^{(2)}$ of a second-order tensor $\boldsymbol{D}$ is given by the (isotropic) average, over all orientations $\boldsymbol{n}\in S^2$, of the directional diffusivity $\boldsymbol{n}\cdot\boldsymbol{D}\boldsymbol{n} = \mathrm{tr}(\boldsymbol{D}\boldsymbol{N})$. Similarly, mean diffusivity $D^{(4)}$ of a fourth-order tensor $\mathbb{D}$ is given by the (isotropic) average, over all orientations $\boldsymbol{n}\in S^2$, of the directional diffusivity $\mathrm{tr}(\mathbb{D}\mathbb{N})$. Using the results of the orientation tensors with respect to the isotropic orientation distribution function, we can deduce that

$$D^{(2)} = \langle D_{ij}n_in_j\rangle = \mathrm{tr}(\boldsymbol{D}\langle\boldsymbol{N}\rangle) = \frac{1}{3}\mathrm{tr}(\boldsymbol{D}\boldsymbol{I}) = \frac{1}{3}\mathrm{tr}(\boldsymbol{D}),$$

and

$$D^{(4)} = \langle D_{ijkl}n_in_kn_kn_l\rangle = \mathrm{tr}(\mathbb{D}\langle\mathbb{N}\rangle) = \frac{1}{5}\mathrm{tr}(\mathbb{D}\mathbb{I}^s) = \frac{1}{5}\mathrm{tr}(\mathbb{D}^s).$$

Here $\langle\cdot\rangle$ designates average of its argument with respect to the isotropic ODF. Under the assumption that $\mathbb{D}$ is totally symmetric, the above becomes

$$D^{(4)} = \frac{1}{5}\mathrm{tr}(\mathbb{D}).$$

We note that $D^{(2)} > 0$ whenever $\boldsymbol{n}\cdot\boldsymbol{D}\boldsymbol{n} > 0$ for all $\boldsymbol{n}\in S^2$, and $D^{(4)} > 0$ whenever $\mathrm{tr}(\mathbb{D}\mathbb{N}) > 0$ for all $\boldsymbol{n}\in S^2$.

It is sometimes desirable to fit a high-order tensor model with a tensor model of lower order. To find the second-order tensor model that best fits a given fourth-order tensor model, we seek the second-order tensor $\boldsymbol{D}$ such that the difference between the signals given by the two models over all gradient directions is minimized

$$\min_{\boldsymbol{D}} \tfrac{1}{2} \int_{S^2} |\mathrm{tr}(\boldsymbol{G}\boldsymbol{D}) - \mathrm{tr}(\mathbb{G}\mathbb{D})|^2 \, d\sigma. \tag{39}$$

The second-order tensor solution of the above minimization problem is given by [13]

$$\boldsymbol{D} = \frac{3}{35} \left(10\mathbb{D}\boldsymbol{I} - \mathrm{tr}(\mathbb{D})\boldsymbol{I}\right). \tag{40}$$

In components, this second-order tensor is given by

$$D_{ij} = \frac{3}{35} \left(10D_{ijkk} - D_{kkll}\delta_{ij}\right). \tag{41}$$

This agrees with the spherical harmonic computations of [15]. One can easily verify that $\mathrm{tr}(\boldsymbol{D}) = \frac{3}{5}\mathrm{tr}(\mathbb{D})$, and therefore, the mean diffusivity of the second-order tensor that best fits a fourth-order tensor is equal to the mean diffusivity of that fourth-order tensor. This shows the consistency of the tensor-order reduction method. Although in the minimization problem (39) the constraint of positive definiteness of the second-order tensor $\boldsymbol{D}$ was not imposed, it was proved in [13] that $\boldsymbol{D}$ is indeed positive definite.

In this chapter we presented the elements of tensor algebra that are required for dealing with fourth-order tensors. The applications to such an algebra to high angular resolution diffusion tensor imaging have been discussed. In particular, the question of fitting diffusion models with different tensor order becomes straightforward by using results on orientation tensors. Furthermore, by using the six-dimensional representation, the estimation of fourth-order diffusion tensors can be done with the same methods that are used for the second-order diffusion tensor. The only difference is that one has to ensure that the tensor is totally symmetric by imposing the linear constrains given by Cauchy's relations. Another possible application of tensor algebra is the problem of resolving fiber crossings. This can be done by estimating fourth-order orientation tensors and then use the well established methods of detecting material symmetry axes to infer the structure of fibers.

References

[1] S. G. Advani and C. L. Tucker III. The use of tensors to describe and predict fiber orientation in short fiber composites. *J. Rheol.*, 31(8):751–784, 1987.

[2] J. Applequist. Traceless cartesian tensor forms for spherical harmonic functions: new theorems and applications to electrostatics of dielectric media. *J. Phys. A Math. Gen.*, 22:4303–4330, 1989.

[3] G. Backus. A geometric picture of anisotropic elastic tensors. *Reviews of Geophysics and Space Physics*, 8(3):633–671, 1970.

[4] A. Barmpoutis, B. Jian, and B. C. Vemuri. Symmetric positive 4th order tensors & their estimation from diffusion weighted MRI. Technical Report REP-2007-312, University of Florida, 2007.

[5] P. J. Basser, J. Mattiello, and D. Le Bihan. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson. B*, 123:247–254, 1994.

[6] A. Campanella and M. L. Tonon. A note on the Cauchy relations. *Meccanica*, 29:105–108, 1994.

[7] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche. Apparent diffusion coefficients from high angular resolution diffusion imaging: estimation and applications. *Magn. Reson. Med.*, 56:395–410, 2006.

[8] S. Forte and M. Vianello. Symmetry classes for elasticity tensors. *J. Elasticity*, 43:81–108, 1996.

[9] M. E. Gurtin. *An Introduction to Continuum Mechanics*, volume 158 of *Mathematics in Science and Engineering*. Academic Press, New York, 1981.

[10] A. E. H. Love. *A Treatise on the Mathematical Theory of Elasticity*, 4th edition. Dover, New York, 1944.

[11] K. V. Mardia. Statistics of directional data. *J. Royal Stat. Soc. Ser B*, 37(3):349–393, 1975.

[12] A. J. McConnell. *Applications of the Absolute Differential Calculus*. Blackie and Son, London, 1931.

[13] M. Moakher. Fourth-order Cartesian tensors: Old and new facts, notions and applications. *Q. J. Mech. Appl. Math.*, 61(2):181–203, 2008.

[14] M. Moakher and A. N. Norris. The closest elastic tensor of arbitrary symmetry to an elasticity tensor of lower symmetry. *J. Elasticity*, 85:215–263, 2006.

[15] E. Özarslan and T. H. Mareci. Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. *Magn. Reson. Med.*, 50:955–965, 2003.

[16] K. Reich. *Die Entwicklung des Tensorkalküls: Vom absoluten Differentialkalkül zur Relativitätstheorie*. Number 11 in Science Networks - Historical Studies. Birkhäuser Verlag, Basel/Boston, 1994.

[17] D. S. Tuch, R. M. Weisskoff, J. W. Belliveau, and V. J. Wedeen. High angular resolution diffusion imaging of the human brain. In *Proceedings of the 7th Annual Meeting of ISMRM*, Philadelphia, 1999, page 321.

[18] W. Voigt. *Lehrbuch der Kristallphysik*. Teubner, Leipzig, 1910.

[19] L. J. Walpole. Fourth rank tensors of the thirty-two crystal classes: multiplication tables. *Proc. R. Soc. Lond.*, A391:149–179, 1984.

[20] G. S. Watson. Distributions on the circle and sphere. *J. Appl. Probab.*, 19:265–280, 1982.

[21] J. Weickert and H. Hagen, editors. *Visualization and Processing of Tensor Fields*. Springer, Berlin, 2006.

Structure-Specific Statistical Mapping of White Matter Tracts*

Paul A. Yushkevich[1], Hui Zhang[2], Tony J. Simon[2], and James C. Gee[1]

[1] Penn Image Computing and Science Laboratory (PICSL), Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA
pauly2@mail.med.upenn.edu, gee@mail.med.upenn.edu

[2] Department of Psychiatry and Behavioral Sciences, M.I.N.D. Institute, University of California, Davis, CA, USA
huiz@seas.upenn.edu, tjsimon@ucdavis.edu

Summary. We present a new model-based framework for the statistical analysis of diffusion imaging data associated with specific white matter tracts. The framework takes advantage of the fact that several of the major white matter tracts are thin sheet-like structures that can be effectively modeled by medial representations. The approach involves segmenting major tracts and fitting them with deformable geometric medial models. The medial representation makes it possible to average and combine tensor-based features along directions locally perpendicular to the tracts, thus reducing data dimensionality and accounting for errors in normalization. The framework enables the analysis of individual white matter structures, and provides a range of possibilities for computing statistics and visualizing differences between cohorts. The framework is demonstrated in a study of white matter differences in pediatric chromosome 22q11.2 deletion syndrome.

1 Introduction

In recent years there has been increased interest in neuroimaging statistical mapping techniques that are structure-specific. The underlying belief is that analysis that takes into account the unique properties of specific anatomical structures and focuses its attention on these structures can be more sensitive than analysis performed over the whole brain. Furthermore, analysis that restricts its attention to structures of interest produces inferences that are also structure-specific. Such inferences can be communicated and visualized more effectively than whole-brain results, since they can be described and presented in the context of well-known structures. Thus, for instance, a significant cluster at the tail of the hippocampus is easier to describe and visualize than a cluster at a certain position in Talairach space. Another key feature of structure-specific analysis is the ability to combine or average data along anatomically

* Originally published in *Neuroimage*. 2008 Jun 1; 41(2):448–461.

meaningful directions while respecting the boundaries between structures, as opposed to uniform smoothing over the whole brain.

Structure-specific analysis requires a suitable representation for the anatomical structures of interest, and the main aim of this chapter is to provide a representation that is particularly well-suited for analyzing white matter structures in diffusion MRI studies. White matter tracts (fasciculi) are thin, sheet-like, or tube-like structures, and as such can be effectively described using skeletons. This fact is leveraged in the *tract-based spatial statistics (TBSS)* approach [44] although, in that approach, skeletons are derived for the entire white matter region and no distinction between white matter structures is made. In this chapter we propose to use the *continuous medial representation (cm-rep)* [56] to model individual white matter tracts. The advantage of using cm-reps, as opposed to skeletonization, is that the skeleton of the structure of interest is represented by a parametric surface. A parametric representation of the skeleton allows us to perform manifold-based statistical analyses similar to those used in cortical flat-mapping (e.g., [19, 48]).

One of the main motivations for representing white matter tracts using parametric geometrical models is to offer an alternative to smoothing in situations where it is necessary to increase the sensitivity of statistical analysis at the cost of specificity. While isotropic smoothing reduces the locality of detectable statistical differences equally in all directions, geometrical models provide a way to combine and collapse data in the context of the structure of interest, making it possible to reduce the locality along "less interesting" directions, while maintaining specificity along the directions that are more meaningful. One way to apply such shape-based averaging is to reduce the dimensionality of a volumetric dataset to a two-manifold that is representative of the overall shape of the dataset. For thin structures, the skeleton is an ideal manifold onto which to project volumetric data because the skeleton captures the overall shape of the structure.

In this chapter, the new structure-specific analysis framework is used to model major fasciculi with sheet-like geometry: the corpus callosum (CC), corticospinal tract (CST), inferior longitudinal fasciculus (ILF), superior longitudinal fasciculus (SLF), inferior fronto-occipital fasciculus (IFO), and uncinate fasciculus (UNC). The framework is demonstrated by application to statistical mapping of white matter tracts in chromosome 22q11.2 deletion syndrome (DS22q11.2), which encompasses most cases of DiGeorge and Velocardiofacial syndromes amongst others. Cluster analysis performed on the medial surfaces of white matter tracts yields a number of statistically significant findings, indicating that DS22q11.2 is associated with bilateral differences in diffusivity in three of the fasciculi [42].

The chapter is organized as follows. Section 2 reviews current approaches to white matter analysis and motivates the medial-based structure-specific mapping framework. The methods employed by the framework, including atlas-building, model initialization, fitting and statistical analysis, are described in Sect. 3. A demonstration of the method in the context of the

DS22q11.2 study is presented in Sect. 4.2. The discussion in Sect. 5 focuses on the next steps in the implementation of the structure-specific analysis framework, such as direct automatic fitting of medial tract models to diffusion tensor imaging (DTI) data.

2 Background

The traditional approach to structure-specific analysis of diffusion imaging data has been to identify regions of interest and integrate diffusion-based measurements over these regions, forming a small number of statistical features [21]. The advent of fiber tracking algorithms [36] has made it possible to map out DTI-based statistical features along the length of fiber bundles. A number of authors [14, 23, 29, 31] identify bundles of fibers with similar shape and make inferences on the basis of bundle centerlines; this approach is well-suited for tubular structures, but larger sheet-like structures like the corpus callosum have to be divided into several tubular bundles. Our approach is designed to analyze sheet-like structures as a whole, and is highly complementary to bundle-based methods: one could use our models for sheet-like fasciculi and bundle-based analysis for tube-like fasciculi.

Cascio et al. [12] and Sun et al. [46] collapse DTI-derived features onto the midsagittal cross-section of the corpus callosum and perform analysis in this two-dimensional space. Such structure-specific approaches are powerful because they take advantage of the specific geometry of the corpus callosum. However, the dimensionality reduction accomplished by these methods is very distinct from our approach: in the former, DTI data are averaged or combined along the length of fibers and projected onto a flat region on the midsagittal plane, while in our approach this data is collapsed onto a manifold representative of the overall shape of a white matter tract.

Another approach to white matter analysis involves adapting voxel-based and deformation-based morphometry [4] to diffusion data. Whole-brain morphometry is tailored for exploratory analysis, and serves different needs than structure-specific methods. A common approach is to normalize the subjects' structural images (T1, T2, or in some cases, fractional anisotropy (FA) images derived from diffusion-weighted MRI) to a standard whole-brain template (e.g., [18]). More recently, specialized deformable registration methods for diffusion data have been developed [11, 60]. Voxel-based morphometry in DTI has been applied in a large number of studies. One of the concerns raised about the standard VBM approach in DTI data is that structure boundaries and within-structure directional information are not taken into account during smoothing and analysis [22]. The *tract-based spatial statistics (TBSS)* framework for DTI morphometry addresses this limitation by incorporating the geometric properties of white matter tracts in the analysis. This is achieved by using registration to compute a "mean FA" image, thresholding this image, computing the skeleton of the thresholded region, projecting the subjects' FA

values onto the skeleton, and performing statistical analysis on the skeleton. TBSS leverages the fact that the skeleton is a natural representation for white matter structures, which are thin sheet-like or tube-like objects. One of the limitations, however, is that by operating on the mean FA image, TBSS ignores orientation information, which can lead to fasciculi that have different orientation but similar anisotropy being combined together into a single structure. Thus the skeleton computed by TBSS may not correspond to the skeletons of the individual fasciculi at such locations. Recent work by Kindlmann et al. [27] overcomes this limitation to some extent by formulating a way to automatically detect surfaces that separate adjacent fasciculi with differing fiber orientations, and providing a way to estimate the skeletons of adjacent fasciculi separately.

We also represent white matter structures medially. However, unlike TBSS, we use medial models to represent specific white matter structures of interest, rather than all of white matter. To distinguish between adjacent tracts, we derive segmentations of individual fasciculi using fiber tractography. The use of parametric medial models with predefined topology simplifies statistical analysis and makes it possible for the rich collection of statistical tools and visualization approaches developed in the cortical flat-mapping literature to be applied to white matter studies. This is illustrated by the manifold-based cluster analysis presented at the end of this chapter.

3 Methods and Materials

3.1 Analysis Framework Overview

The main aim of this chapter is to give a proof-of-concept demonstration of a new model-based framework for representing major fasciculi and analyzing associated DTI data. This framework, outlined in Fig. 1, includes four main components: *spatial normalization*, *tract segmentation*, *geometric tract modeling*, and *statistical analysis*. These components leverage prior work in DTI registration, atlas-based segmentation, deformable medial modeling, and nonparametric statistical analysis. Prior to describing these components in detail, we note that alternative approaches to each of them are available and emphasize that the focus of this chapter is on demonstrating the framework as a whole rather than on the implementation of each of its components.

In the current implementation, segmentation, modeling, and analysis all take place in *atlas space*. That is, all DTI images in a study are normalized to a common atlas, which is derived from the images themselves. Working in atlas space facilitates the segmentation and modeling of white matter structures. An alternative, discussed in Sect. 5, is to perform segmentation, modeling, and analysis in subject space. The main difference between these approaches is in the way that inter-subject correspondences are computed: on the basis of image-based normalization in the atlas-based approach or on the basis of shape in subject-space analysis.

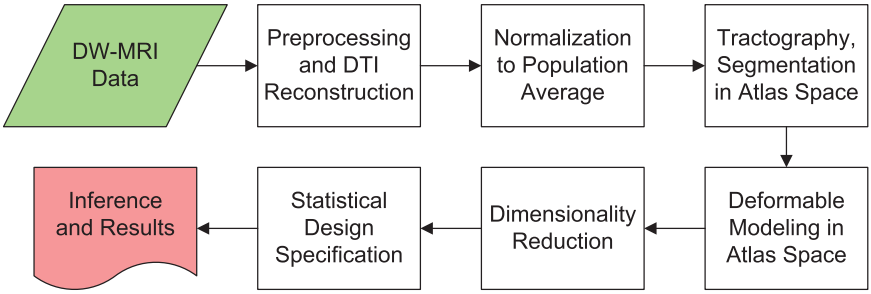


Fig. 1. An outline of the tract-specific analysis framework

3.2 Normalization to a DTI Atlas

The DTI atlas is constructed by normalizing all DTI images in a study to a common average image. A number of techniques for computing an unbiased average image from a cohort of images have been proposed [5, 20, 24, 49]. We use a similar approach [57] that leverages a new deformable DTI registration algorithm [60], in which image similarity is computed on the basis of full tensor images, rather than scalar features. When measuring similarity between tensor images, it is essential to take into account the fact that when a transformation is applied to a tensor field, the orientation of the tensors is changed [1]. A unique property of our registration algorithm is the ability to model the effect of deformation on tensor orientation as an analytic function of the Jacobian matrix of the deformation field. By using full tensor information in the similarity metric, the method aligns white matter regions better than scalar-based registration methods, as demonstrated by Zhang et al. [58] in a task-driven evaluation study.

When building the atlas, the initial average image is computed as a log-Euclidean mean [3] of the input diffusion tensor images. The average is then iteratively refined by repeating the following procedure: register the subject images to the current average, then compute a refined average for the next iteration as the Euclidean mean of the normalized images. This procedure is repeated until the average image converges.

To allow asymmetry analysis and to simplify the modeling of bilateral fasciculi, a symmetric average is constructed by simply applying the above averaging procedure to the nonsymmetric population average and its mirror image across the midsagittal plane (the midsagittal plane is estimated automatically by solving a three-parameter rigid registration problem, similar to [2]). The images of the subjects are then registered to the symmetric average. The combination of the symmetric average image and the warps from the average to each of the subject's DTI images forms the symmetric DTI atlas. This atlas captures the correspondences between the white matter regions of different subjects as well as between left and right hemispheres of each subject.

3.3 Tract Segmentation

Six major fasciculi (CC, CST, IFO, ILF, SLF, UNC) were segmented in the white matter atlas using an established protocol by Wakana et al. [50, 51], which is based on fiber tracking. The FACT fiber tracking algorithm [33] was applied to the DTI white matter atlas with the FA threshold of 0.15 and an inner product threshold of 0.7, which prevents angles larger than 45° during tracking. A fiber was tracked from the center of any voxel with a FA > 0.2. The threshold of 0.2 was chosen as suggested by Smith et al. [44], who point out that this value is low enough to include all the major white matter structures (that are the focus of this chapter) but high enough to exclude the majority of the voxels that are composed of primarily grey matter tissue or CSF. A lower FA value can potentially include additional finer structures near the cortex but also likely to include more voxels that are not white matter. More sophisticated segmentation procedures than simple FA threshold are required to delineate such finer cortical connections. In addition, the accuracy of the intersubject alignment for such connections requires further investigation.

Fibers of interest were extracted using a multiple region of interest (ROI) based approach [13, 34]. Two types of ROIs were defined: those through which all fibers in a tract must pass, and those through which none of the fibers may pass. Since in this chapter a 2D manifold-based model of fasciculi is employed, only those fasciculi that have a major portion that is sheet-like were segmented. Fasciculi that are more appropriately represented by tubular models, such as the tapetum of the CC, the cingulum, the fornix, and the optic tract were not segmented. Fasciculi in the cerebellum and brain stem were not considered either. Only the arcuate portion of the SLF, which can be tracked consistently, was segmented. Binary 3D segmentations of individual fasciculi were generated by labeling voxels in the white matter atlas through which at least one fiber passed. The binary segmentations were further edited using ITK-SNAP [55] by one of the authors (HZ) to remove extraneous connections: the portion of the CST extending to the cerebellum was removed; regions that can not be disambiguated between CC and CST were attributed to both structures. The second column of Fig. 7 shows the binary segmentations for the six selected fasciculi. The segmentation of all six fasciculi required less than 2 h of user time, and no slice-by-slice or voxel-by-voxel editing was performed.

3.4 Geometrical Modeling of Major Fasciculi

Geometrical modeling of the fasciculi involves fitting deformable medial models (cm-reps) to binary segmentations of the fasciculi in atlas space. These models describe the skeleton[3] and the boundary of a geometrical object as parametric digital surfaces with predefined topology. Furthermore, the models

[3] In this chapter, the term *skeleton* refers to the Blum skeleton [10, 15, 16]. The Blum skeleton of a geometrical object in $\mathbb{R}^3$ is a surface or set of surfaces that result from thinning the object uniformly, that is, evolving the boundary at uniform

describe the geometrical relationship between the skeleton and the boundary, such that deformations applied to the model's skeleton can be propagated to the model's boundary. The medial models used in this chapter have skeletons consisting of a single surface patch, that is, nonbranching skeletons. For most fasciculi, a single-surface model appears sufficient, as our fitting results in Sect. 4.2 demonstrate.

The deformable medial model is constructed as follows. We start with an initial mesh representation of the skeleton, denoted $\mathcal{S}_0$. The mesh $\mathcal{S}_0$ is *triangular*, which allows maximum flexibility for modeling complex skeleton shapes. Each vertex i in the mesh $\mathcal{S}_0$ is a tuple $\{\mathbf{m}_{0,i}, R_{0,i}\}$, where $\mathbf{m}_{0,i} \in \mathbb{R}^3$ is the coordinate of the vertex and $R_{0,i} \in \mathbb{R}^+$ is the *radius value*, which describes the local thickness of the model. The initial mesh $\mathcal{S}_0$ is fairly sparse (see Fig. 4d). Loop subdivision [30] is used to derive a sequence of successively finer meshes $\mathcal{S}_1, \mathcal{S}_2, \ldots$, which converge to a continuous limit surface as $k \to \infty$. In practice, only two or three levels of subdivision are used when modeling the skeleton surface. To deform a medial model to optimize some cost function, we change the values of $\mathbf{m}_{0,i}$ and $R_{0,i}$ in the sparse mesh $\mathcal{S}_0$. The corresponding deformation of the finer-level meshes is simply a linear function of the update to the sparse mesh.

Let $\mathcal{S}_k$ be the mesh representing the skeleton of a deformable medial model at subdivision level k. The boundary of the medial model is derived analytically from the skeleton using *inverse skeletonization*, that is, finding a closed surface mesh $\mathbf{b}_k$ whose Blum skeleton is approximately $\mathcal{S}_k$. The boundary surface mesh consists of two halves, each a surface patch of disk topology. These patches lie on the opposite sides of the skeleton. The boundary half-patches are denoted $\mathbf{b}_k^-$ and $\mathbf{b}_k^+$ (see Fig. 2) and are given by the following inverse skeletonization expression [15, 16, 56]:

$$\mathbf{b}^{\pm} = \mathbf{m} + R\,\mathbf{U}^{\pm}\,, \tag{1}$$

$$\mathbf{U}^{\pm} = -\nabla_{\mathbf{m}} R \pm \sqrt{1 - ||\nabla_{\mathbf{m}} R||^2}\,\mathbf{N}_{\mathbf{m}}\,, \tag{2}$$

where $\mathbf{N}_{\mathbf{m}}$ denotes the unit normal to the skeleton and $\nabla_{\mathbf{m}} R$ is the Riemannian gradient of R on the manifold $\mathbf{m}$. The normal and the Riemannian gradient at node i are approximated from one-ring neighborhood of i (i.e., all nodes connected to i by an edge) using the expression for the tangent vectors to the limit surface of Loop subdivision, as given by Xu [53].

However, the inverse skeletonization expression (1) on its own does not guarantee that the surface patches $\mathbf{b}_k^-$ and $\mathbf{b}_k^+$ together form the boundary of an object in $\mathbb{R}^3$. For some $\mathbf{m}$ and R, these patches may intersect or self-intersect; and the edges of the patches may be disjoint. In order for the patches $\mathbf{b}_k^-$ and $\mathbf{b}_k^+$ to form the boundary of an object, $\mathbf{m}$ and R must satisfy a set of sufficient conditions [15, 16, 56]. These conditions include inequality

---

speed along the inward normal until the object becomes infinitely thin. An example of a skeleton is shown in Fig. 4a.

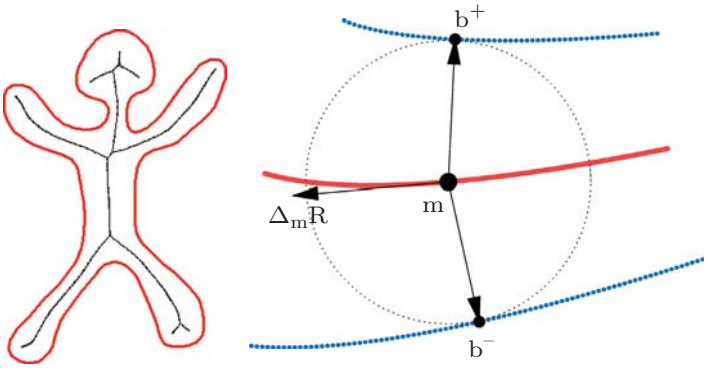


Fig. 2. 2D medial geometry. *Left*: the skeleton of a 2D object. *Right*: local geometry. The red curve represents the skeleton $\mathbf{m}$. The radial scalar field R defines the radius of the maximal inscribed disk. The boundary, *in blue*, consists of two parts, $\mathbf{b}^+$ and $\mathbf{b}^-$, derived from $\mathbf{m}$ and R by inverse skeletonization in (1). The vector $\nabla_{\mathbf{m}} R$ lies in the tangent plane of $\mathbf{m}$ and points in the direction of greatest change in R. The arrows pointing from $\mathbf{m}$ to $\mathbf{b}^+$ and $\mathbf{b}^-$ are called *spokes*

constraints that prevent intersection and self-intersection, as well as a single equality constraint, that ensures that $\mathbf{b}_k^-$ and $\mathbf{b}_k^+$ share a common edge:

$$\|\nabla_{\mathbf{m}} R\| = 1 \qquad \text{on } \partial\Omega\,. \tag{3}$$

In previous work, this constraint has been satisfied in different ways: by adapting the domain Ω such that (3) is satisfied along $\partial\Omega$ [54]; by defining R as a solution of a Poisson partial differential equation with the boundary condition identical to (3) [56] or by using specially designed subdivision rules for quadrilateral meshes to model the surface $\mathbf{m}$ [47]. In this chapter, we simply admit skeleton meshes $\mathcal{S}_k$ that slightly violate the equality constraint (3) and correct these solutions by local adjustment. During fitting, a penalty is imposed on $(\|\nabla_{\mathbf{m}} R\| - 1)^2$ along edge nodes in the mesh $\mathcal{S}_k$, which forces $\|\nabla_{\mathbf{m}} R\|$ to be very close to 1. The correction procedure consists of simply applying (1) as if $\|\nabla_{\mathbf{m}} R\|$ were actually equal to 1 at edge nodes. While this approach may not be as elegant as the previous ones, it behaves well in practice and leads to a very efficient fitting algorithm. The remaining sufficient conditions of inverse skeletonization are inequality constraints and are handled in this chapter the same way as in [56], that is, by using additional penalty terms during model fitting.

Now, as we apply deformations to the nodes of the sparse-level skeleton $\mathcal{S}_0$, we can use inverse skeletonization to generate a boundary of the medial model at subdivision level k. Furthermore, the correspondence between points on the skeleton and boundary points established by (1) allows us to parameterize the region enclosed by the boundary, as discussed in Sect. 3.5. Overlap between the medial model and a binary image can be computed efficiently.

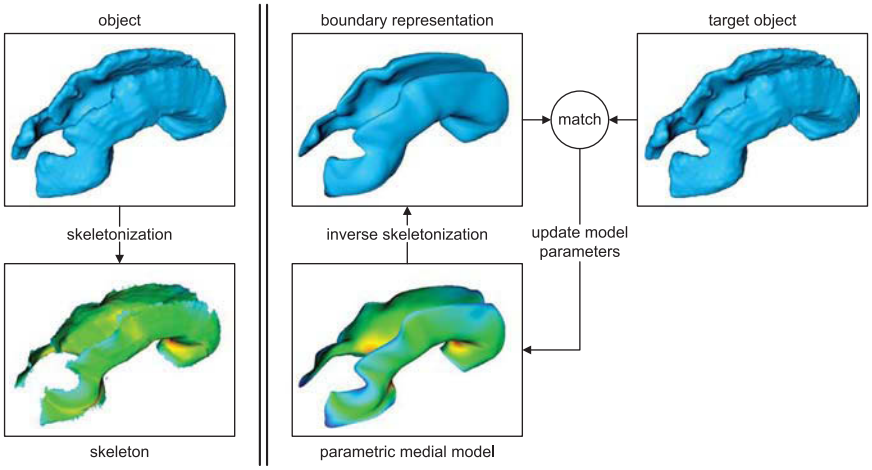


Fig. 3. Skeletonization vs. medial modeling. *Left*: A 3D object and the skeleton derived by skeletonization. The color map on the skeleton is the "radius scalar field" R or, equivalently, the distance to the closest boundary point. *Right*: medial modeling, which is, essentially, the opposite of skeletonization. A deformable parametric medial model is defined as a surface or set of surfaces, and the boundary is derived analytically using "inverse skeletonization," (see (1)). The model is then deformed to maximize fit between its boundary and the object of interest. Since the medial model is parametric, it can be flattened, that is, mapped to a 2D domain; not so for skeletonization. The key difference between skeletonization and medial modeling is that the former computes exact skeletons, but does not guarantee that the branching topology of the skeletons is consistent across individuals; the latter computes approximate skeletons, but guarantees the same topology for all individuals, allowing effective statistical analysis

Thus, fitting a model to a binary segmentation of a fasciculus is an optimization problem where overlap between the model's interior and the segmentation is maximized and the penalty terms required for inverse skeletonization to be well-posed are minimized. Model fitting is discussed in greater detail in [56]. The overall concept of deformable medial modeling, as opposed to deterministic skeletonization, is illustrated in Fig. 3.

3.4.1 Automatic Model-Building for Initialization

Prior to fitting a deformable medial model to a target structure, an initial model must be generated. While it is possible to begin with a simple stock model, the freedom to choose an arbitrary domain Ω makes it possible to build data-driven initial models. With the freedom to define cm-reps over arbitrary domains comes the problem of finding the right domain and the right mesh configuration for a particular anatomical structure. We accomplish this by essentially flattening the skeleton of the structure, under constraints that minimize local distortion.

Automatic domain generation and model building employ the following pipeline. Given a binary image of the structure of interest, we first compute the Voronoi skeleton of the structure using the Quickhull algorithm [7] (as implemented in *qhull* software, www.qhull.org) and use simple heuristics (similar to [37]) to prune away the least salient branches. Unlike earlier work on automatic medial model initialization [45], we do not require the Voronoi skeleton to be pruned down until it becomes a manifold. Instead, we use the Maximum Variance Unfolding (MVU) technique [52], which finds a two-dimensional embedding of the skeletons' vertices that minimally distorts the distances between neighboring vertices. The two-dimensional embedding assigns a pair of coordinates $(u_i, v_i) \in \mathbb{R}^2$ to each of the vertices $\mathbf{x}_i$ in the skeleton. A manifold can then be fitted to the skeleton by fitting low-degree polynomial surfaces to the data $x_i(u_i, v_i), y_i(u_i, v_i), z_i(u_i, v_i)$. The domain Ω is computed as a region homeomorphic to a disc in $\mathbb{R}^2$ that encloses all of the points (u_i, v_i). This is achieved by rasterizing the scatter plot of (u_i, v_i) and applying dilation and erosion operations. To produce a quality triangulation of Ω, we sample a set of 20–40 points along its boundary and use the meshing program *triangle* to produce conforming constrained Delaunay triangulations [41]. The initialization pipeline is illustrated in Fig. 4 using the CC as the example.

3.5 Shape-Based Coordinate System

A key property of cm-rep medial models is the ability to parameterize the entire interior of the model using a shape-based coordinate system [56]. Let us refer to the vectors $R\mathbf{U}^{\pm}$ with tail on the medial surface $\mathbf{m}$ as *spokes*. If $\mathbf{m}$ is a continuous surface and the constraints of inverse skeletonization are met, the spokes span the interior of the model, that is, the region enclosed by the surface $\mathbf{b}^- \cup \mathbf{b}^+$. Every point inside this region can belong to only one spoke. Let us use coordinates (u, v) to describe where the tail of a spoke is on the medial surface, and use the coordinate $\xi \in [-1, 1]$ to describe a location along a spoke (when $\xi > 0$, it references the spoke $R\mathbf{U}^+$, and when $\xi < 0$, it references the spoke $R\mathbf{U}^-$). Then every point $\mathbf{x}$ on the model's interior can be assigned a set of coordinates (u, v, ξ), as follows:

$$\mathbf{x}(u, v, \xi) = \begin{cases} \mathbf{m}(u, v) + \xi R(u, v)\, U^+(u, v) & \text{if } \xi \geq 0 \\ \mathbf{m}(u, v) - \xi R(u, v)\, U^-(u, v) & \text{o/w.} \end{cases} \tag{4}$$

This assignment is unique, except when $(u, v) \in \partial\Omega$, in which case the spokes $R\mathbf{U}^+$ and $R\mathbf{U}^-$ coincide. This 3D parametrization of the model is a powerful tool, as it allows us to analyze data on the interior of structures in a shape-based coordinate system.

3.6 Statistical Analysis on Medial Manifolds

This section describes one of the many possible ways to perform structure-specific white matter analysis in cm-rep coordinate space. We assume a simple

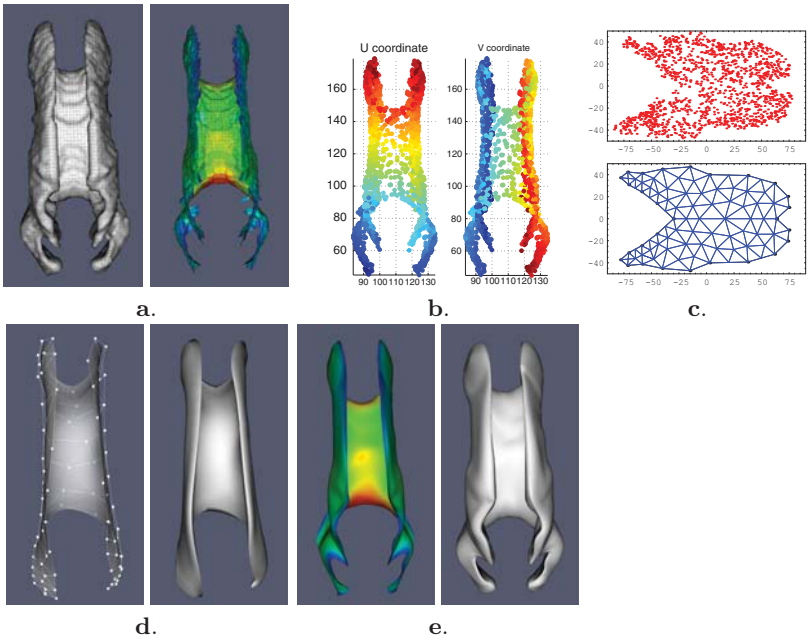


Fig. 4. Model building and fitting, illustrated for the CC. (**a**) The boundary and the pruned skeleton of the binary segmentation of the CC (in atlas space). The color map plots the distance from the skeleton to the boundary, that is, thickness. *Red* indicates greatest thickness (at the genu and splenium); *blue* indicates places where the CC is thinnest. Note that the skeleton is not a manifold since it may contain any number of branches. (**b**) The two-dimensional embedding of the skeleton using maximum variance unfolding [52]. The two plots use color to show the u and v parameters assigned to each point on the skeleton. *Blue* represents low u, v values and *red* represents higher u, v. (**c**) A scatter plot of the skeleton's points in u, v space and, underneath, the conforming constrained Delaunay triangulation of the region containing the points. (**d**) The initial cm-rep model for the corpus callosum: on the left is the skeleton surface with the control mesh displayed as a wireframe; on the right is the boundary surface derived by inverse skeletonization. (**e**) The result of fitting the cm-rep model to the binary segmentation (the skeleton and the boundary of the model are shown). Again, thickness is plotted in color

scenario, where we are interested in analyzing differences between two groups of subjects. To further simplify the analysis, we restrict our attention only to the differences in the apparent diffusion coefficient (ADC). ADC is a scalar measurement of overall diffusivity equal to the mean of the eigenvalues of a diffusion tensor.

The key feature of our statistical analysis approach is the projection of tensor-derived quantities lying on the interior of a fasciculus onto its medial manifold. This projection results in a dimensionality reduction along the spoke

direction, that is, the direction orthogonal to the boundary of the fasciculus. Like smoothing, dimensionality reduction increases sensitivity at the cost of decreased specificity. However, unlike isotropic Gaussian smoothing, data reduction along spoke directions respects the boundaries between structures. Furthermore, isotropic Gaussian smoothing causes equal loss of specificity in all directions, while, arguably, dimensionality reduction along spokes in thin sheet-like fasciculi causes loss of specificity along the least interesting direction.

We propose and compare two approaches to dimensionality reduction; other approaches are also possible. In both approaches, we map the three-dimensional field of diffusion tensors $D(\mathbf{x})$ defined over the image domain to two-dimensional tensor fields defined over the medial manifold of each fasciculus. For each fasciculus, this mapping is achieved by sampling diffusion tensors along spokes $RU^{\pm}(u,v)$ and deriving a single tensor $\tilde{D}(u,v)$ for each point on the medial manifold. Such a two-dimensional tensor field is computed for each fasciculus in each subject.

The first strategy is an adaptation of the dimensionality reduction strategy used in TBSS [44] and will be referred to as the "Max-FA" strategy. In this strategy, we select the diffusion tensor that has the largest FA of all the tensors sampled along the pair of spokes:

$$\tilde{D}(u,v) = D(\mathbf{x}(u,v,\xi^*(u,v)))\,, \quad \text{where}$$
$$\xi^*(u,v) = \arg\max_{\xi\in[-1,1]} \mathrm{FA}(D(\mathbf{x}(u,v,\xi)))\,. \quad (5)$$

In TBSS, this strategy is adopted to correct for registration errors in atlas building. Since FA tends to be larger on the interior of fasciculi, by searching for the position with the largest FA value along vectors, which extend from the interior to the fasciculus boundary, one may recover intersubject correspondences distorted by imperfect registration [44]. However, there is a slight but important difference between our Max-FA approach and TBSS. The latter computes the skeleton of the entire white matter region (defined by an FA threshold) and so the search for the maximum FA value may span multiple fasciculi if they are not separated by voxels with low FA. Our approach searches within the interior of each fasciculus, as defined by our tracking-based segmentation protocol. The effects of this difference are examined in Sect. 4.2.

The second proposed strategy is to compute the average tensor along each pair of spokes. Following Arsigny et al. [1], we compute the tensor average using the log-Euclidean formulation:

$$\tilde{D}(u,v) = \exp[\frac{1}{2}\int_{-1}^{1} \log D(\mathbf{x}(u,v,\xi))\ \mathrm{d}\xi\,]\,. \quad (6)$$

This strategy assumes that the registration quality is sufficiently high and uses averaging as the way to increase sensitivity of the analysis at the cost of specificity in the spoke direction, which is assumed to be of least interest.

A third strategy, which is not studied here, is not to perform dimensionality reduction, but to generate statistical maps in the full u, v, ξ coordinate system. This approach has the advantage of greatest specificity.

The analysis of statistical feature maps defined over manifolds has been studied extensively in the neuroimaging literature and many of the techniques previously proposed for cortical surface analysis can be directly applied to the analysis of white matter differences on medial manifolds of fasciculi (e.g., [19, 48]). As an illustration, we analyze the ADC differences between two groups using nonparametric cluster-based analysis with family-wise error rate (FWER) correction [38]. First, we compute a two-sample t-test at each point on the medial manifold of each fasciculus. Then, given an arbitrary threshold t_0, the set of clusters with $t > t_0$ is extracted. In this context, a cluster $C \in \Omega$ is a simply connected subset in the domain of the medial manifold that satisfies $t(u, v) \geq t_0$ for all $(u, v) \in C$. We record the mass of each cluster C, defined as

$$\text{mass}(C) = \int_C t(u, v) \mathrm{d}A \,.$$

We compare the mass of each cluster to the histogram of maximum cluster masses generated from a large number of identical experiments in which the labels of the subjects are randomly permuted. This histogram lookup yields a corrected p-value for each cluster. The histogram of cluster masses is pooled over all structures of interest, so that the FWER correction takes into account the number of structures that we examine.

4 Experimental Evaluation

4.1 Evaluation Experiment: Subjects and Imaging Protocol

We demonstrate the proposed structure-specific white matter analysis paradigm in an ongoing study of DS22q11.2 conducted at the Department of Psychiatry and Behavioral Science, M.I.N.D. Institute of University of California, Davis. DS22q11.2 is associated with reduced brain volume, and in particular, has been linked to morphological changes in white matter [17, 26, 42]. We are interested in testing the hypothesis that white matter structure is different between children with DS22q11.2 and typically developing children (TD).

The participants in the study include 13 children with DS22q11.2 and 18 children with typical development. Diffusion-weighted and structural MRI were acquired for each participant on a 3 Tesla Siemens Trio scanner. A single-shot, spin-echo, diffusion-weighted echo-planar imaging (EPI) sequence was used for the diffusion-weighted MRI. The diffusion scheme was as follows: one image with minimum possible diffusion gradient, referred to as the $b = 0$ image, followed by twelve images measured with twelve non-collinear and noncoplanar diffusion encoding directions isotropically distributed in space ($b = 1{,}000\,\mathrm{s\,mm^{-2}}$). Additional imaging parameters for the diffusion-weighted sequence are: TR = 6,500 ms, TE = 99 ms, 90° flip angle,

number of averages = 6, matrix size = 128 × 128, slice thickness = 3.0 mm, spacing between slices = 3.0 mm, 40 axial slices with in-plane resolution of $1.72 \times 1.72\,\text{mm}^2$.

Diffusion-weighted images were corrected for motion and eddy-current artifacts using the method described in [32], prior to extracting brain parenchyma with the Brain Extraction Tool [43]. Diffusion tensor images were then reconstructed from the diffusion-weighted images using the standard linear regression approach [8]. Finally, the resulting tensor volumes were resampled to a voxel space of 128 × 128 × 64 with voxel dimensions equal to $1.72 \times 1.72 \times 2.5\,\text{mm}^3$. The resampled volume, with axial dimension equal to a power of 2, is better suited for registration algorithms that require the construction of standard multiresolution image pyramids.

4.2 Results

The pipeline described in Sect. 3 was applied to all 31 images in the study. First, a symmetric unbiased population atlas was constructed using image registration; this atlas is illustrated in Fig. 5. A qualitative assessment of registration accuracy is presented in Fig. 6, which shows the DTI images from eight subjects warped into the space of the atlas. No large-scale registration errors are evident, although there are small-scale differences.

Fiber tracking and tract labeling were used to segment six fasciculi of interest in atlas space. Fiber tracking results and the corresponding binary segmentations are shown in the first two columns of Fig. 7, and the skeletons and boundary surfaces of medial models fitted to each of the binary segmentations are shown in last two columns. The model fitting accuracy for the six fasciculi is listed in Table 1, in terms of Dice overlap between the binary segmentation of a fasciculus and the fitted cm-rep model, as well as root mean square displacement from the boundary of the fitted cm-rep to the boundary of the

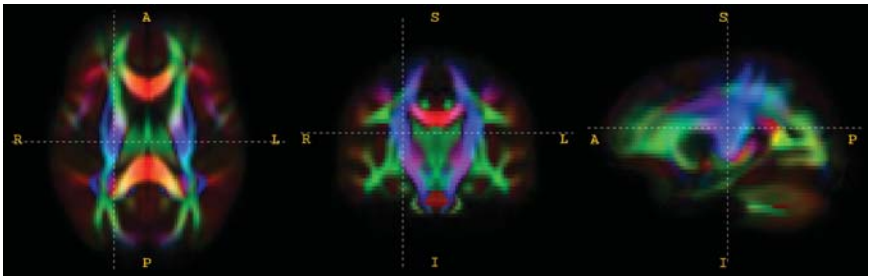


Fig. 5. Three orthogonal views through the right CST in the white matter atlas constructed from 31 subjects (13 children with DS22q11.2 and 18 typically developing children). The standard color-coding scheme for visualizing the principal direction of diffusion is used: red for left–right, green for anterior–posterior, and blue for inferior–superior. The intensity is scaled by FA

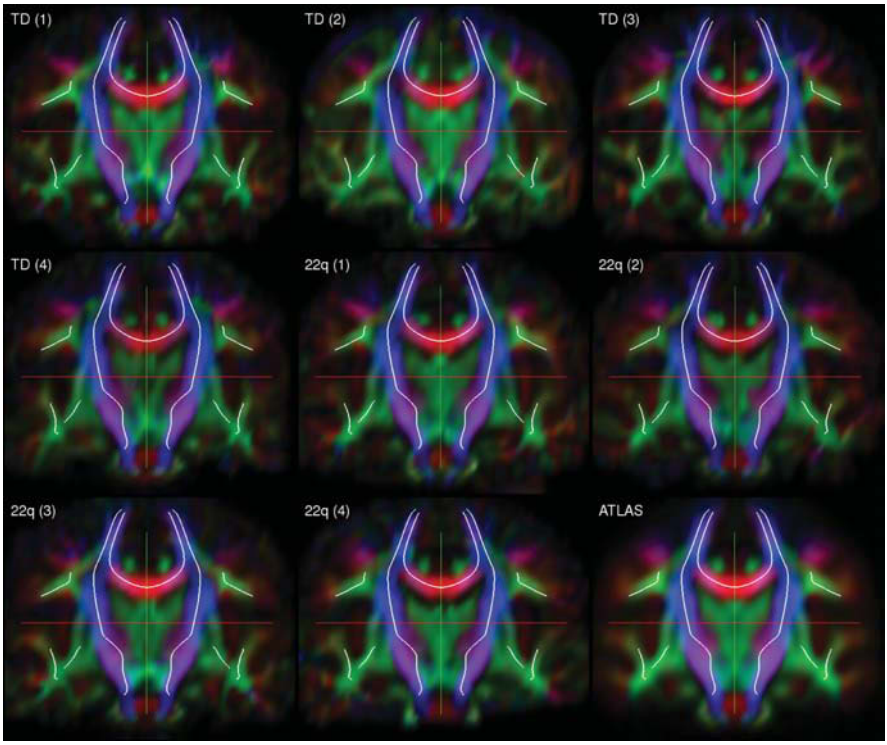


Fig. 6. Visual assessment of registration accuracy. Each plot above shows a coronal slice through one of the subjects' DTI images registered to the DTI atlas, except for the last plot, which shows the atlas itself. The projection of the medial models of the fasciculi into the cut plane are shown as white curves in each plot. These curves are exactly the same across all nine plots and are provided as a reference

binary segmentation. The fitting accuracy is high (∼0.95 Dice overlap) for five of the six fasciculi. For the ILF, fitting accuracy is lower (0.90 Dice overlap), due to branching fibers in the posterior part of the fasciculus (see Fig. 7). Fitting this structure more accurately would require extending our medial modeling approach to allow multi-manifold skeletons, which is the subject of ongoing research. A combined view of the six fasciculi and the medial models fitted to them is shown in Fig. 8.

Figure 9 plots the distribution of thickness along the medial models fitted to each of the fasciculi. Thickness is also shown as a color map in Fig. 7. The CC and CST are the thickest fasciculi, with thickness reaching 16 mm. ILF is the thinnest, with maximum thickness below 9 mm. The voxel size of our DTI data is $1.72 \times 1.72 \times 3.0\,\mathrm{mm}^3$. Hence, the dimensionality reduction strategies proposed above do indeed combine data from multiple voxels.

For each fasciculus, we tested the hypothesis that ADC is greater in TD than in DS22q11.2. This hypothesis was tested using both Max-FA and tensor

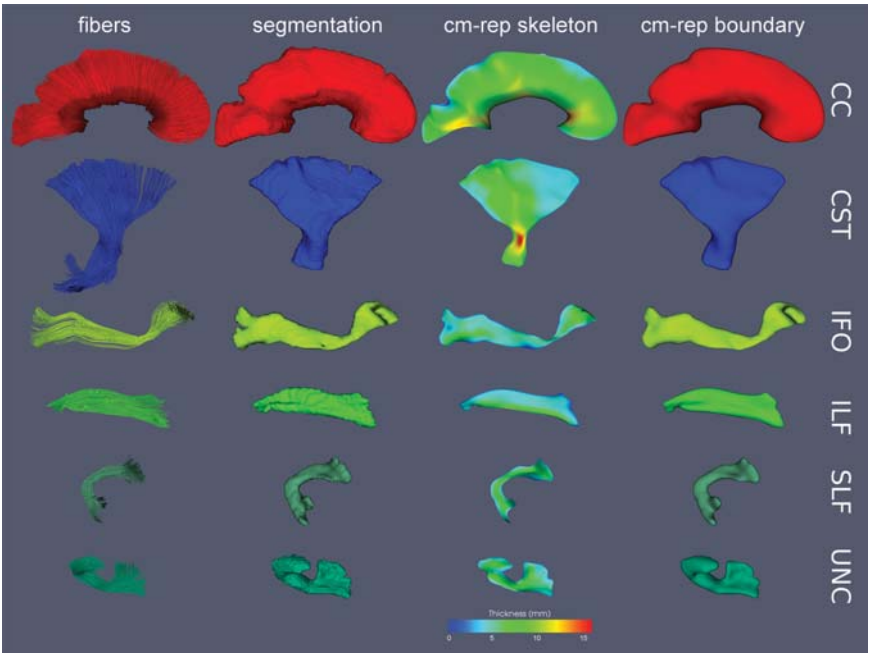


Fig. 7. Model-fitting results for six white matter tracts. *First column*: fiber tracking and fiber labeling result. *Second column*: binary segmentation derived by rasterizing the fiber tracking result. *Third column*: the skeleton surface **m** of the cm-rep model fitted to the binary segmentation. The color map along the skeleton surface plots the local thickness of each model, equal to twice the value of the radius scalar field R. *Fourth column*: the boundary surface **b** of the cm-rep model

Table 1. Accuracy of cm-rep model fitting to binary segmentations of the six selected fasciculi, in terms of Dice overlap and root mean square boundary displacement (RMSBD)

Fasciculus	Dice Overlap	RMSBD (mm)
CC	0.957	0.332
CST	0.954	0.285
ILF	0.902	0.320
SLF	0.944	0.246
IFO	0.950	0.226
UNC	0.951	0.219

averaging dimensionality reduction strategies. In each case, a t-map was generated over the medial manifold of each fasciculus. These t-maps are plotted in three dimensions in Fig. 10 as color maps over the medial manifolds. An alternative visualization in the top row of Fig. 11 displays the t-maps in the two-dimensional u, v domain of each fasciculus, making it easy to display the

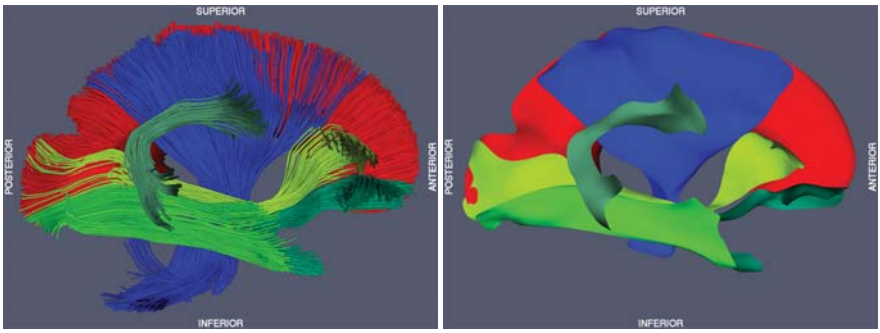


Fig. 8. *Left*: fiber tracking results for the six selected fasciculi. *Right*: skeletons of the cm-rep models fitted to the six fasciculi

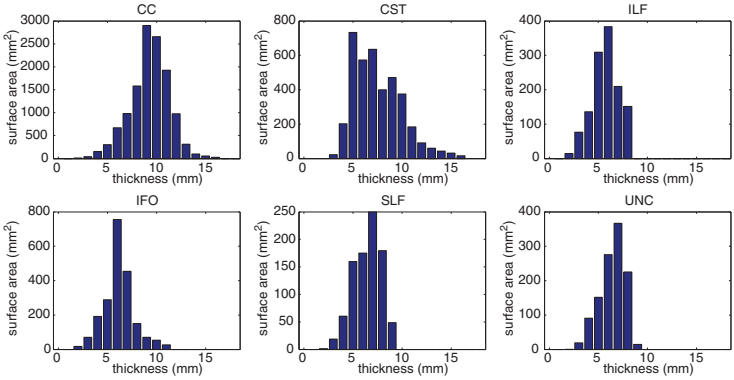
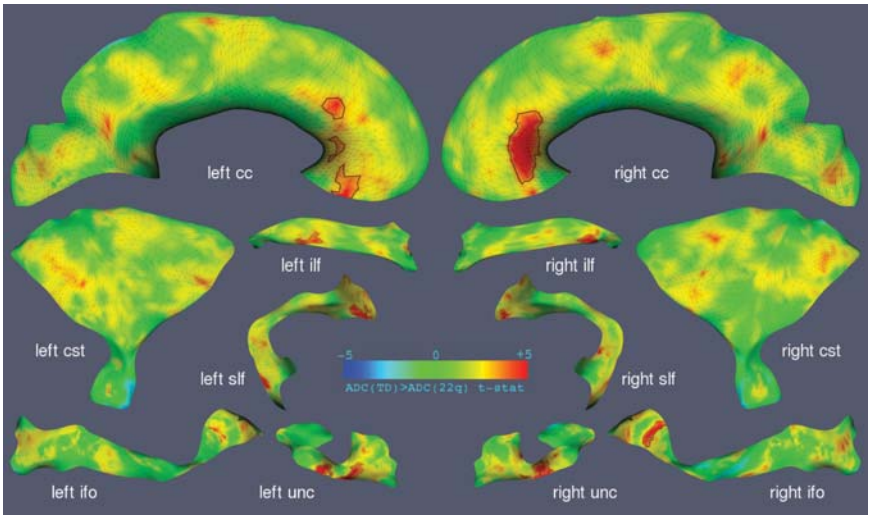


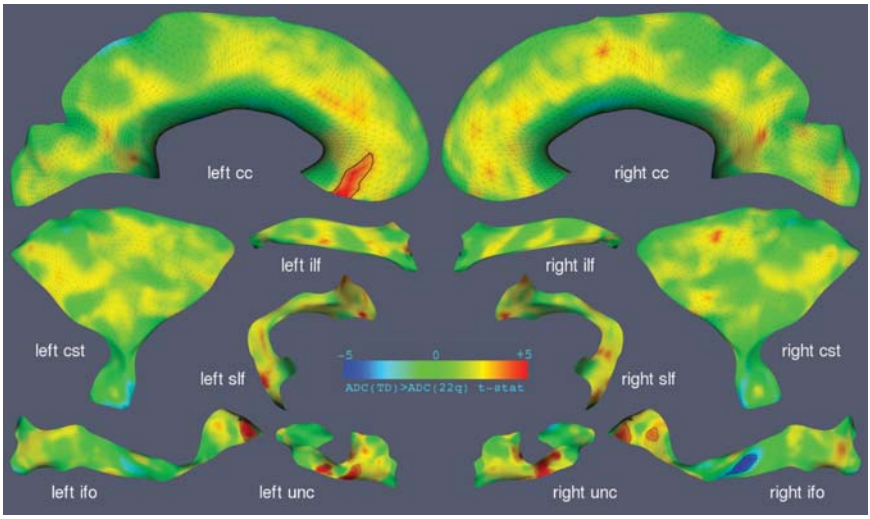
Fig. 9. Thickness profiles of the medial models fitted to the six fasciculi in atlas space. The bars in the thickness profiles represent the surface area of the region on the medial manifold where the thickness falls within the range $[i - 0.5, i + 0.5]$ for $i = 0, 1, \ldots, 18$. Thickness is defined as the diameter of the maximal inscribed ball centered at the given point on the medial axis (i.e., thickness is twice the radius function $R(u, v)$)

results of hypothesis testing on a single page. Note that the two-dimensional projection of each fasciculus roughly resembles the shape of the fasciculus in three dimensions, making it easier to interpret the two-dimensional t-maps. The 2D visualization is also an effective tool for comparing the results of different studies or different contrasts within a study.

To find the regions where ADC differences between TD and DS22q11.2 groups are statistically significant, we performed permutation-based cluster analysis with 10,000 random permutations. The cluster threshold t_0 was set to 3.40, corresponding to an uncorrected p-value of 0.001. The sets of clusters computed for each contrast and each dimensionality reduction strategy are shown in a two-dimensional flattened view in the bottom row of Fig. 11. The clusters detected using tensor averaging dimensionality reduction are listed



Max-FA Strategy



Tensor Averaging Strategy

Fig. 10. Results of cluster analysis comparing ADC between DS22q11.2 and TD, displayed in three dimensions. The medial models are colored by the t-score for the hypothesis ADC(TD) > ADC(DS22q11.2). Statistically significant clusters are marked by a black outline. Results for both dimensionality reduction strategies (Max-FA and tensor averaging) are shown

in Table 2 and the clusters detected with Max-FA are listed in Table 3. The results consistently point to ADC differences in CC, IFO, and UNC. However, the study is ongoing and we intend to report the complete results, including FA differences, in a separate clinically oriented chapter.

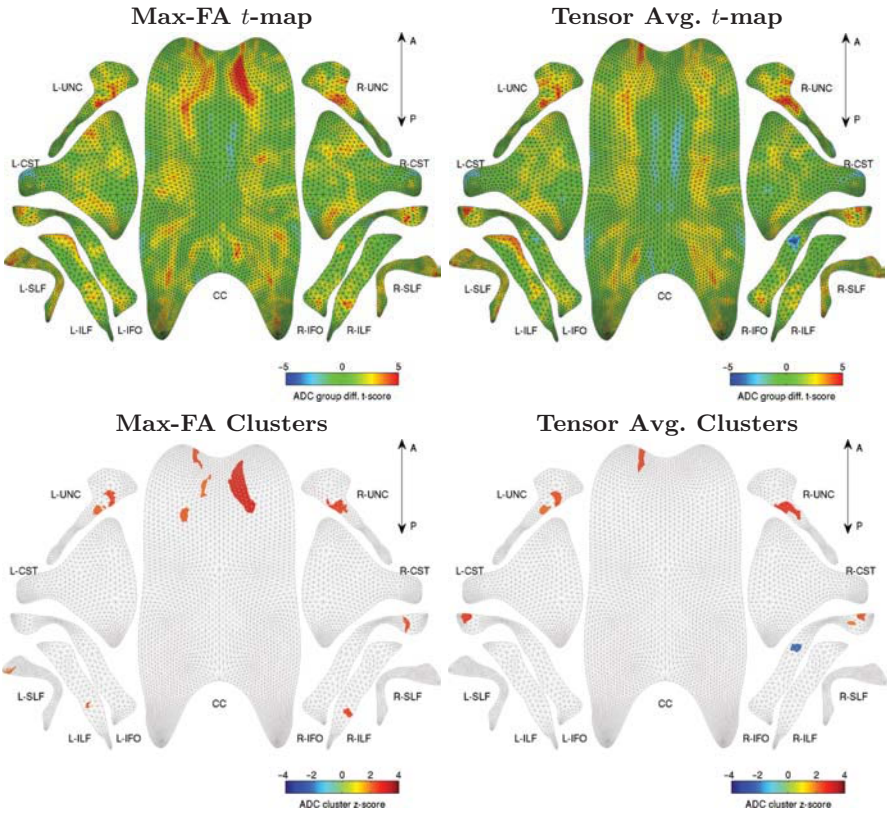


Fig. 11. A statistical analysis of ADC differences between TD vs. DS22q11.2, computed using two different dimensionality reduction strategies (Max-FA and tensor averaging, see Sect. 3.6) and visualized in two dimensions. *Top row*: t-maps for the TD vs. DS22q11.2 group comparison ($t > 0$ indicates greater ADC in TD, and vice versa). The t-map in each plot is the same as in Fig. 10, but is projected onto a 2D domain, allowing the t-maps associated with all 11 fasciculi to be viewed simultaneously. *Bottom row*: a similar visualization of the statistically significant clusters in ADC comparison (clusters thresholded at $t = \pm 3.40$ with FWER-corrected p-value below 0.05). The 2D domain for each fasciculus is displayed as a gray mesh (see Fig. 4c), and the significant clusters are displayed in color. *Red* clusters correspond to increased ADC in DS22q11.2 and *blue* clusters correspond to increased ADC in TD

More clusters are obtained using the Max-FA strategy than using tensor averaging. In particular, a large cluster is detected in the right genu of the CC, for which there is no equivalent in the clusters generated by tensor averaging. However, looking at the differences between lists of clusters can be deceptive because a cluster's presence or absence can be sensitive to small differences in the topography of the t-map from which the cluster is generated. A more fair assessment when comparing t-maps generated by two strategies is to examine

Table 2. Statistically significant clusters for the ADC comparison between DS22q11.2 and typical development (TD), *using the tensor averaging dimensionality reduction strategy*

Fasciculus	Location	Area (mm^2)	$\bar{t}$	p_{corr}	$\bar{t}_{maxfa}$
cc	L. anterior	82.9	3.89	0.0096	3.42
Left ifo	Anterior	88.8	4.49	0.0085	3.11
Left unc	Anterior medial	54.7	4.23	0.0184	3.99
Left unc	Medial	34.4	4.14	0.0346	4.13
Right ifo	Mid-fasciculus	58.4	−4.44	0.0266	−1.26*
Right ifo	Anterior	26.0	3.60	0.0492	4.16
Right ifo	Anterior	48.3	3.84	0.0213	2.98
Right unc	Medial	103.7	4.19	0.0070	3.62

Clusters are defined as connected regions with $|t| > 3.40$ on the fascicular medial manifold. For each cluster, four values are shown. The first value is the surface area of the cluster on the medial manifold. Next, $\bar{t}$ is the average value of the t-statistic inside the cluster. Positive values of t indicate greater ADC in TD than in DS22q11.2. Third, the family-wise error rate (FWER) corrected p-value of the cluster is given (the correction treats pointwise hypothesis testing across all 11 fasciculi as a single multiple comparison problem). Last, $\bar{t}_{maxfa}$ gives the average over the cluster of the t-map computed using the "Max-FA" dimensionality reduction strategy. The difference between $\bar{t}$ and $\bar{t}_{maxfa}$ gives an indication of how sensitive the cluster is to the choice of dimensionality reduction. Asterisk (*) indicates the clusters where this difference is most striking.

Table 3. Statistically significant clusters for the ADC comparison between DS22q11.2 and TD, *using the Max-FA dimensionality reduction strategy*

Fasciculus	Location	Area (mm^2)	$\bar{t}$	p_{corr}	$\bar{t}_{tenav}$
cc	L. anterior	50.7	3.75	0.0162	3.66
cc	L. anterior	23.9	3.52	0.0427	2.38
cc	L. anterior	34.0	3.79	0.0274	2.92
cc	R. anterior	208.9	4.49	0.0015	1.71*
Left ilf	Posterior	40.4	3.80	0.0213	2.39
Left unc	Anterior medial	53.2	4.07	0.0149	4.18
Left unc	Medial	34.5	4.17	0.0268	4.08
Left slf	Inferior	26.4	4.15	0.0372	3.06
Right ilf	Posterior	49.9	4.10	0.0164	2.52
Right ifo	Anterior	48.3	4.25	0.0169	3.32
Right unc	Medial	74.4	3.99	0.0090	4.16

See footnote in Table 2 for details

the t-values from different strategies over the extent of the same cluster. This comparison is made in Tables 2 and 3. For each cluster, the average t in the "native" t-map is reported (i.e., the same t-map that was used to generate the cluster), along with the average t in the "opposite" t-map. In all cases, the native $|\bar{t}| > 3.40$, since 3.40 is the clustering threshold. For almost all

clusters, the opposite $\bar{t}$ is also fairly large, indicating agreement between the two strategies. The two exceptions are the cluster for the tensor averaging strategy located in the right IFO ($\bar{t}_{\text{tenav}} = -4.44, \bar{t}_{\text{maxfa}} = -1.26$) and the cluster for Max-FA strategy located in the right genu of the CC ($\bar{t}_{\text{maxfa}} = 4.49, \bar{t}_{\text{tenav}} = 1.71$).

Figure 12 helps explain why the two strategies can produce different t-maps. It looks at the aforementioned cluster in the right IFO, and examines where the tensors with maximal FA are in relation to the medial manifold of the fasciculus. It turns out that over the region of the cluster, the tensors with maximal FA are located not along the medial manifold but towards the end of the spokes that point in the direction of the adjacent ILF, which passes very close to the IFO. The tensors in the ILF have larger FA than tensors in the IFO, and the Max-FA strategy chooses the former. Thus, although it seems to produce larger and more numerous clusters, the Max-FA strategy can also fail to detect statistically significant differences in places where one fasciculus passes close to another.

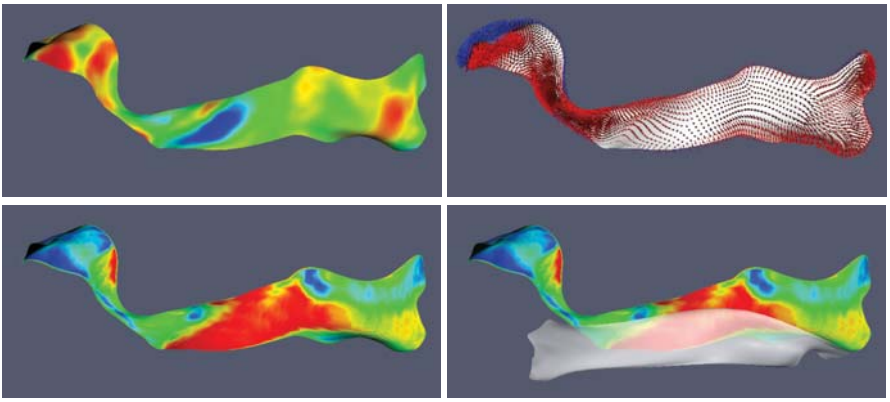


Fig. 12. Closer examination of the cluster in the right IFO that appears under the tensor averaging strategy but has no equivalent under the Max-FA strategy. *Top left*: the t-map for the ADC(TD) vs. ADC(DS22q11.2) comparison produced by the tensor averaging strategy (this is the same t-map as in Fig. 10). The cluster of interest is the blue spot in the t-map (increased ADC in DS22q11.2). *Top right*: the spokes along the medial manifold of the right IFO. Each spoke is a line segment extending from the medial manifold to the corresponding (and also nearest) point on the boundary (see Fig. 2). Spoke RU^+ is shown in *red* and spoke RU^- in *blue*. *Bottom left*: a color map showing where the Max-FA strategy samples the DTI image in relation to the medial manifold of the right IFO. *Red* indicates that the tensor with maximal FA is located at the end of spoke RU^+ (on average over all 31 subjects) and *blue* indicates that the tensor with maximal FA is at the end of RU^-. *Green* indicates that the tensor with Max-FA is, on average, located along the medial manifold. *Bottom right*: the medial manifold of the right ILF is shown next to the right IFO. The two fasciculi pass close to each other. The Max-FA strategy appears to be sampling tensors from the right ILF, which has greater FA values, and causes the statistically significant differences in the adjacent IFO to be missed

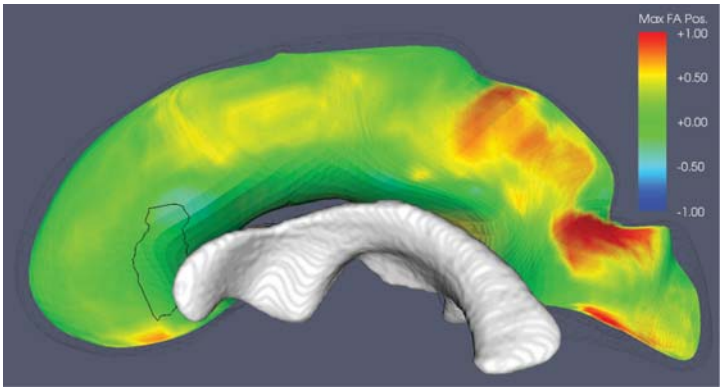


Fig. 13. A closer examination of the cluster in the right anterior CC that appears under the Max-FA strategy but has no equivalent under the tensor averaging strategy. The location of the cluster in question is outlined on the medial surface of the CC. The medial manifold is colored by the relative position along the spokes of the tensor with maximal FA. Positive values (*red*) indicate that, on average, the maximal FA tensor is located on the side of the medial manifold away from the midsagittal plane; negative values indicate that maximal FA tensors tend to be located on the side facing the midsagittal plane; and values close to zero indicate no bias in the location of the maximal FA tensor. The lateral ventricle is in close proximity to the cluster, which may help explain why tensor averaging, which is more prone to partial volume errors, is less sensitive in this region

The other cluster that is inconsistent between the two strategies is located in the right genu of the CC. As Fig. 13 shows, this cluster is located where the CC is adjacent on one side to the lateral ventricles. Unlike in the case of the right IFO, there is no systematic shift in the position of the diffusion tensor with maximal FA along the spokes in that area of the CC; on average, the position of the maximal FA tensor is right on the medial manifold. It is likely that the low sensitivity of the tensor averaging strategy is due to the partial volume effects between the CC and adjacent highly isotropic gray and CSF structures.

5 Discussion and Conclusions

This work demonstrates the feasibility of using deformable medial models for structure-specific statistical mapping of DTI data in major white matter fasciculi. The cornerstone of this approach is the shape-based parametrization of the fasciculi. Medial models are a natural representation for thin sheet-like structures because the medial manifold effectively summarizes the overall shape of structures and presents an ideal target for dimensionality reduction. Collapsing tensors by taking the average along spokes is an attractive alternative to spatial smoothing of DTI data. Both approaches trade specificity

for sensitivity by averaging data from multiple locations, but spoke-based averaging also reduces data dimensionality and explicitly strives to ensure that averaging is applied within the structure of interest, not across its boundaries. However, since averaging takes place in atlas space and normalization is never perfect, there is still a possibility for tensor data outside of the structure of interest to be included in the analysis. The TBSS-inspired approach of summarizing data along the spokes by selecting the tensor with maximal FA is also feasible in our framework and, as shown earlier, can lead to increased sensitivity, albeit at the cost of potentially missing significant differences in areas of adjoining fasciculi.

The use of skeletons in DTI analysis was pioneered by Smith et al. [44] in the TBSS approach. Our approach is distinct from TBSS in several important ways. First, our approach is directed at statistical mapping of individual fasciculi, while TBSS aims to draw inferences about white matter as a whole. In TBSS, deterministic skeletonization is applied to the set of all voxels with above-threshold fractional anisotropy. This results in a skeleton consisting of hundreds or thousands of branches, some of which may be spurious and sensitive to noise. In contrast, the cm-rep approach yields a single approximating medial surface for each of the six selected fasciculi. This simple parametric representation allows statistical mapping of individual fasciculi and provides an easy and attractive way to visualize and interpret statistical differences. It establishes a canonical two-dimensional coordinate space for the fasciculi (Fig. 11), where the results from different white matter studies can be compared.

There are a number of ways to improve the proposed analysis framework. An obvious improvement is to expand the set of structures included in the analysis. Wakana et al. [50] identify the set of fasciculi that can be reliably segmented using the tractography-based protocol in high-quality DTI data. In addition to the six cortical white matter structures modeled in this work, they include the anterior thalamic radiation (ATR), tapetum of the corpus callosum (TP), the projection of the cingulum to the cingulate gyrus (CgC), the hippocampal projection of the cingulum (CgH), fornix (FX), and four fasciculi in the cerebellum and brainstem: medial lemniscus (ML) and superior, middle and inferior cerebellar peduncles (SCP, MCP, and ICP). Of these, four (ATR, SCP, MCP, and ICP) have sheet-like geometry and may be modeled using the current framework. One of these structures, ATR, is partially included in our definition of the CST, which combines (see Fig. 14) projections to Brodmann areas 1–3 (primary somatosensory cortex), 4 (primary motor cortex), 5 (somatosensory association cortex), 6 (pre-motor and supplementary motor cortex), 8 and 9 (dorsolateral prefrontal cortex). The remaining structures identified by Wakana et al. as reliable, such as the fornix and the cingulum, have tube-like rather than sheet-like geometry. Established curve-based geometrical tract models [14, 23, 29, 31] are particularly well-suited for modeling these structures, and can be used in parallel with our approach.

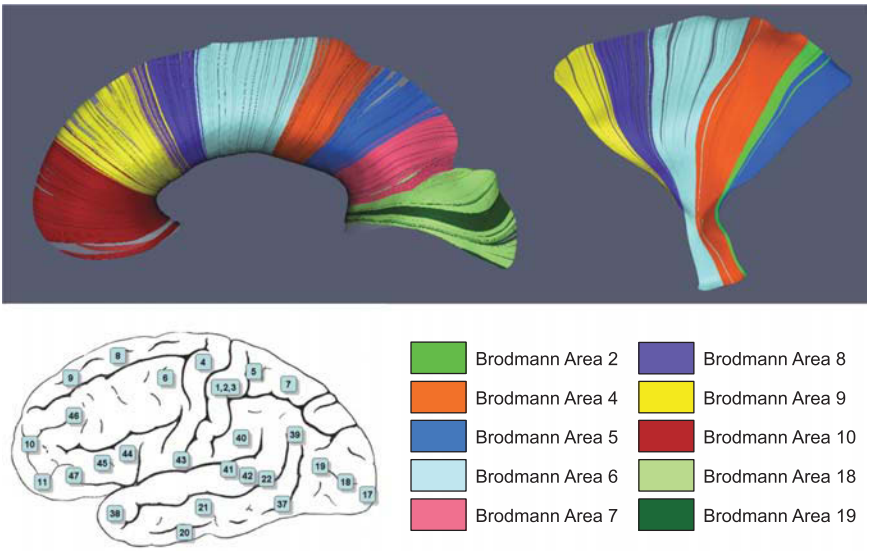


Fig. 14. Surface-based representations of the CC (*top left*) and the CST (*top right*) labeled according to their cortical connections to Brodmann areas (*bottom left*). The color encoding of the Brodmann areas is shown in the bottom right panel. Tracking was performed directly on the medial manifolds, using a new algorithm [59] that extends the concept of fiber tractography to 2D medial manifolds. Tracking on medial manifolds is akin to geodesic shooting, except that path taken at each point $\mathbf{m}(u, v)$ is parallel to the estimate of the principal direction of diffusion computed from the mean diffusion tensor $D(u, v)$ (which is derived by (6) above). Cortical connectivity labels are approximated as follows: first, the ICBM-MNI template [35], which includes Brodmann area labels, is coregistered to the DTI atlas described in Sect. 3.2 using affine registration; second, each fiber is assigned the label of the Brodmann area closest to its endpoint, in terms of Euclidean distance

In the current experiment, cm-rep models were fit to tract segmentations in *atlas space*. The dimensionality reduction of tensor data along spokes also occurs in atlas space, meaning that in *subject space* the averaging takes place along curves that are not guaranteed to extend from the medial axis to the boundary, since the deformations yielded by registration algorithms do not preserve the geometrical relationship between the boundaries and skeletons of structures. An alternative approach is to fit cm-rep models to tract segmentations in subject space. This can be achieved by warping the binary segmentations from the template to each of the subjects, and fitting a cm-rep model to each of these warped segmentations. In this way, the registration component would be leveraged to align the boundary a structure, while the normalization of the structure's interior would be based on the geometrical correspondences generated by cm-rep fitting.

Reliance on expert-driven segmentation of the fasciculi is another limitation. Although the segmentation of the fasciculi in atlas space requires less than 2 h of expert time, performing segmentation in subject space would be prohibitively expensive for many studies. Furthermore, the simple six-parameter tensor-based representation of diffusion in our dataset and simple streamline-based fiber tracking employed by FACT limit the accuracy with which white matter tracts can be segmented. For instance, the data near the cortex is too noisy to accurately model the cortical projections of the fasciculi. However, the limitations of segmentation method do not preclude us from performing statistical analysis, as long as it is safe to assume that the segmentation error is independent of the differences between cohorts in a study. Recent advances in DTI segmentation [6, 28, 39] and higher resolution imaging may indeed make it possible to more accurately label fasciculi in atlas space, as well as in individual subject images, and to do so automatically or with minimal user interaction. However, given the fact that deformable medial models are themselves a very powerful segmentation tool [25, 40], a truly attractive approach would be to fit the deformable medial models employed in this chapter directly to DTI data, rather than to segmentations obtained with other techniques. This would require incorporating tensor-based and connectivity-based likelihood metrics into cm-rep fitting, as well as training shape priors for each of the fasciculi. In addition to shape, priors may incorporate the variation in the angle between spokes of the medial model and the principal direction of diffusion in the underlying DTI data. Model-based segmentation of the fasciculi is one of the key long-term aims of our research.

The statistical methods used in this chapter represent just one of the many possible approaches to structure-specific inference. For example, instead of permutation-based cluster analysis, one could use point-wise hypothesis testing with false discovery rate correction [9]. It is also possible to subdivide the fasciculi according to cortical regions to which individual fibers connect (Fig. 14), and to compute region-wise statistics. Instead of using features derived from diffusion tensors averaged along spokes, multivariate analysis may be applied, where each point on the medial manifold is associated with a tuple of diffusion tensors. Such a framework would be more sensitive to group differences occurring on opposite sides of the medial manifold.

In conclusion, we have presented a novel structure-specific framework for analyzing white matter differences between populations. The technique can be seen both as a model-based structure-specific alternative to the TBSS method [44] and as an extension of the cortical flat mapping paradigm to white matter structures. The six sheet-like fasciculi in the human brain are modeled using deformable medial models, which allows the diffusion tensor data associated with these structures to be projected onto parametric manifolds or even flattened to a two-dimensional domain. This approach simplifies white matter analysis and provides a compelling way to present results of white matter studies.

Acknowledgements

This work was supported by the NIH grants AG027785 (PY), NS061111 (PY), HD042974 (TJC), HD046159 (TJC), MH068066 (V. Megalooikonomou), and NS045839 (J.A. Detre).

References

[1] D. C. Alexander, C. Pierpaoli, P. J. Basser, and J. C. Gee. Spatial transformations of diffusion tensor magnetic resonance images. *IEEE Trans Med Imaging*, 20(11):1131–1139, 2001.

[2] B. A. Ardekani, J. Kershaw, M. Braun, and I. Kanno. Automatic detection of the mid-sagittal plane in 3D brain images. *IEEE Trans Med Imaging*, 16(6):947–952, 1997.

[3] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, 2006.

[4] J. Ashburner and K. J. Friston. Voxel-based morphometry – the methods. *Neuroimage*, 11(6 Pt 1):805–821, 2000.

[5] B. Avants and J. C. Gee. Geodesic estimation for large deformation anatomical shape averaging and interpolation. *Neuroimage*, 23(Suppl 1): S139–S150, 2004.

[6] S.P. Awate, H. Zhang, and J.C. Gee. A fuzzy, nonparametric segmentation framework for DTI and MRI analysis: With applications to dti-tract extraction. *IEEE Transactions on Medical Imaging*, 26(11):1525–1536, 2007.

[7] C. Bradford Barber, D. P. Dobkin, and H. Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4):469–483, 1996.

[8] P. J. Basser, J. Mattiello, and D. LeBihan. Estimation of the effective self-diffusion tensor from the nmr spin echo. *Journal of Magnetic Resonance B*, 103(3):247–254, 1994.

[9] Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188, 2001.

[10] H. Blum. A transformation for extracting new descriptors of shape. In *Models for the Perception of Speech and Visual Form*. MIT Press, 1967.

[11] Y. Cao, M. I. Miller, R. L. Winslow, and L. Younes. Large deformation diffeomorphic metric mapping of vector fields. *IEEE Trans Med Imaging*, 24(9):1216–1230, 2005.

[12] C. Cascio, M. Styner, R. G. Smith, M. D. Poe, G. Gerig, H. C. Hazlett, M. Jomier, R. Bammer, and J. Piven. Reduced relationship to cortical white matter volume revealed by tractography-based segmentation of the corpus callosum in young children with developmental delay. *American Journal of Psychiatry*, 163(12):2157–2163, 2006.

[13] T. E. Conturo, N. F. Lori, T. S. Cull, E. Akbudak, A. Z. Snyder, J. S. Shimony, R. C. McKinstry, H. Burton, and M. E. Raichle. Tracking neuronal fiber pathways in the living human brain. *Proceedings of the National Academy of Sciences U S A*, 96(18):10422–10427, 1999.

[14] I. Corouge, P. Thomas Fletcher, S. Joshi, S. Gouttard, and G. Gerig. Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. *Medical Image Analysis*, 10(5):786–798, 2006.

[15] J. Damon. On the smoothness and geometry of boundaries associated with skeletal structures II: Geometry in the Blum case. *Compositio Mathematica*, 140(6):1657–1674, 2004.

[16] J. Damon. Determining the geometry of boundaries of objects from medial data. *International Journal of Computer Vision*, 63(1):45–64, 2005.

[17] S. Eliez, J. E. Schmitt, C. D. White, and A. L. Reiss. Children and adolescents with velocardiofacial syndrome: a volumetric mri study. *America Journal of Psychiatry*, 157(3):409–415, 2000.

[18] S. H. Eriksson, F. J. Rugg-Gunn, M. R. Symms, G. J. Barker, and J. S. Duncan. Diffusion tensor imaging in patients with epilepsy and malformations of cortical development. *Brain*, 124(Pt 3):617–626, 2001.

[19] B. Fischl, M. I. Sereno, and A. M. Dale. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2):195–207, 1999.

[20] A. Guimond, J. Meunier, and J.-P. Thirion. Average brain models: a convergence study. *Computer Vision and Image Understanding*, 77(9):192–210, 2000.

[21] M. A. Horsfield and D. K. Jones. Applications of diffusion-weighted and diffusion tensor mri to white matter diseases – a review. *NMR Biomed*, 15(7/8):570–577, 2002.

[22] D. K. Jones, M. R. Symms, M. Cercignani, and R. J. Howard. The effect of filter size on vbm analyses of dt-mri data. *Neuroimage*, 26(2):546–554, 2005.

[23] D. K. Jones, A. R. Travis, G. Eden, C. Pierpaoli, and P. J. Basser. Pasta: pointwise assessment of streamline tractography attributes. *Magnetic Resonance in Medicine*, 53(6):1462–1467, 2005.

[24] S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage*, 23(Suppl 1): S151–S160, 2004.

[25] S. Joshi, S. Pizer, P. Thomas Fletcher, P. Yushkevich, A. Thall, and J. S. Marron. Multiscale deformable model segmentation and statistical shape analysis using medial descriptions. *IEEE Transactions on Medical Imaging*, 21(5):538–550, 2002.

[26] W. R. Kates, C. P. Burnette, E. W. Jabs, J. Rutberg, A. M. Murphy, M. Grados, M. Geraghty, W. E. Kaufmann, and G. D. Pearlson. Regional cortical white matter reductions in velocardiofacial syndrome: a volumetric mri analysis. *Biological Psychiatry*, 49(8):677–684, 2001.

[27] G. Kindlmann, X. Tricoche, and C.-F. Westin. Delineating white matter structure in diffusion tensor mri with anisotropy creases. *Medical Image Analysis*, 11(5):492–502, 2007.

[28] C. Lenglet, M. Rousson, and R. Deriche. DTI segmentation by statistical surface evolution. *IEEE Transactions on Medical Imaging*, 25(6):685–700, 2006.

[29] F. Lin, C. Yu, T. Jiang, K. Li, X. Li, W. Qin, H. Sun, and P. Chan. Quantitative analysis along the pyramidal tract by length-normalized parameterization based on diffusion tensor tractography: application to patients with relapsing neuromyelitis optica. *Neuroimage*, 33(1):154–160, 2006.

[30] C. Loop and T. DeRose. Generalized b-spline surfaces of arbitrary topology. In *Computer Graphics (ACM SIGGRAPH Proceedings)*, pp. 347–356, 1990.

[31] M. Maddah, W. M. Wells III, S. K. Warfield, C.-F. Westin, and W. E. L. Grimson. Probabilistic clustering and quantitative analysis of white matter fiber tracts. In *International Conference on Information Processing in Medical Imaging*, pp. 372–383, 2007.

[32] J.-F. Mangin, C. Poupon, C. Clark, D. Le Bihan, and I. Bloch. Distortion correction and robust tensor estimation for mr diffusion imaging. *Medical Image Analysis*, 6(3):191–198, 2002.

[33] S. Mori, B. J. Crain, V. P. Chacko, and P. C. van Zijl. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology*, 45(2):265–269, 1999.

[34] S. Mori, W. E. Kaufmann, C. Davatzikos, B. Stieltjes, L. Amodei, K. Fredericksen, G. D. Pearlson, E. R. Melhem, M. Solaiyappan, G. V. Raymond, H. W. Moser, and P. C. M. van Zijl. Imaging cortical association tracts in the human brain using diffusion-tensor-based axonal tracking. *Magnetic Resonance in Medicine*, 47(2):215–223, 2002.

[35] S. Mori, K. Oishi, H. Jiang, L. Jiang, X. Li, K. Akhter, K. Hua, A. V. Faria, A. Mahmood, R. Woods, A. W. Toga, G. Bruce Pike, P. Rosa Neto, A. Evans, J. Zhang, H. Huang, M. I. Miller, P. van Zijl, and J. Mazziotta. Stereotaxic white matter atlas based on diffusion tensor imaging in an icbm template. *Neuroimage*, 40(2):570–582, 2008.

[36] S. Mori and P.C. M. van Zijl. Fiber tracking: principles and strategies – a technical review. *NMR Biomed*, 15(7/8):468–480, 2002.

[37] M. Näf, O. Kübler, R. Kikinis, M. Shenton, and G. Székely. Characterization and recognition of 3D organ shape in medical image analysis using skeletonization. In *Workshop on Mathematical Methods in Biomedical Image Analysis*, IEEE Computer Society, pp. 139–150, 1996.

[38] T. E. Nichols and A. P. Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, 15(1):1–25, 2002.

[39] L. J. O'Donnell and C. F. Westin. Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Transactions on Medical Imaging*, 26(11):1562–1575, 2007.

[40] S. M. Pizer, D. S. Fritsch, P. A. Yushkevich, V. E. Johnson, and E. L. Chaney. Segmentation, registration, and measurement of shape variation via image object shape. *IEEE Transactions on Medical Imaging*, 18(10):851–865, 1999.

[41] J. Richard Shewchuk. Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. In M. C. Lin and D. Manocha, editors, *Applied Computational Geometry: Towards Geometric Engineering*, volume 1148 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 203–222, 1996. From the First ACM Workshop on Applied Computational Geometry.

[42] T. J. Simon, L. Ding, J. P. Bish, D. M. McDonald-McGinn, E. H. Zackai, and J. Gee. Volumetric, connective, and morphologic changes in the brains of children with chromosome 22q11.2 deletion syndrome: an integrative study. *Neuroimage*, 25(1):169–180, 2005.

[43] S. M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–55, 2002.

[44] S. M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. E. Nichols, C. E. Mackay, K. E. Watkins, O. Ciccarelli, M. Zaheer Cader, P. M. Matthews, and T.E. J. Behrens. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage*, 31(4):1487–1505, 2006.

[45] M. Styner, G. Gerig, S. Joshi, and S.M. Pizer. Automatic and robust computation of 3D medial models incorporating object variability. *International Journal of Computer Vision*, 55(2):107–122, 2003.

[46] H. Sun, P. A. Yushkevich, H. Zhang, P. A. Cook, J. T. Duda, T. J. Simon, and J. C. Gee. Shape-based normalization of the corpus callosum for DTI-based connectivity analysis. *IEEE Transactions on Medical Imaging*, 2007. in Press.

[47] T. B. Terriberry and G. Gerig. A continuous 3-d medial shape model with branching. In *International Workshop on Mathematical Foundations of Computational Anatomy MFCA-2006, in conjunction with MICCAI 2006*, 2006.

[48] P. M. Thompson, M. S. Mega, R. P. Woods, C. I. Zoumalan, C. J. Lindshield, R. E. Blanton, J. Moussai, C. J. Holmes, J. L. Cummings, and A. W. Toga. Cortical change in alzheimer's disease detected with a disease-specific population-based brain atlas. *Cerebral Cortex*, 11(1):1–16, 2001.

[49] P. M. Thompson and A. W. Toga. Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations. *Medical Image Analysis*, 1(4):271–94, 1997.

[50] S. Wakana, A. Caprihan, M. M. Panzenboeck, J. H. Fallon, M. Perry, R. L. Gollub, K. Hua, J. Zhang, H. Jiang, P. Dubey, A. Blitz, P. van Zijl, and S. Mori. Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage*, 36(3):630–644, 2007.

[51] S. Wakana, H. Jiang, L. M. Nagae-Poetscher, P.C. M. van Zijl, and S. Mori. Fiber tract-based atlas of human white matter anatomy. *Radiology*, 230(1):77–87, 2004.

[52] K. Q. Weinberger and L. K. Saul. Unsupervised learning of image manifolds by semidefinite programming. *International Journal of Computer Vision*, 70(1):77–90, 2006.

[53] Guoliang Xu. Discrete laplace-beltrami operators and their convergence. *Computer Aided Geometry Design*, 21(8):767–784, 2004.

[54] P. Yushkevich, P. T. Fletcher, S. Joshi, A. Thall, and S. M. Pizer. Continuous medial representations for geometric object modeling in 2D and 3D. *Image and Vision Computing*, 21(1):17–28, 2003.

[55] P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, 2006.

[56] P. A. Yushkevich, H. Zhang, and J. Gee. Continuous medial representation for anatomical structures. *IEEE Transactions on Medical Imaging*, 25(2):1547–1564, 2006.

[57] H. Zhang, P. A. Yushkevich, D. Rueckert, , and J. C. Gee. Unbiased white matter atlas construction using diffusion tensor images. In *Medical Image Computing and Computer-Assisted Intervention, MICCAI*, 2007.

[58] H. Zhang, B. B. Avants, P. A. Yushkevich, J. H. Woo, S. Wang, L. F. McCluskey, L. B. Elman, E. R. Melhem, and J. C. Gee. High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences: an example study using amyotrophic lateral sclerosis. *IEEE Transactions in Medical Imaging*, 26(11):1585–1597, 2007.

[59] H. Zhang, P. Yushkevich, T. Simon, and J. Gee. Surface-based modeling of white matter fasciculi with orientation encoding. In *IEEE International Symposium on Biomedical Imaging: Macro to Nano*, pp. 708–711, 2008.

[60] H. Zhang, P. A. Yushkevich, D. C. Alexander, and J. C. Gee. Deformable registration of diffusion tensor MR images with explicit orientation optimization. *Medical Image Analysis*, 10(5):764–785, 2006.

Analysis of Distance/Similarity Measures for Diffusion Tensor Imaging

T.H.J.M. Peeters, P.R. Rodrigues, A. Vilanova, and B.M. ter Haar Romeny

Department of Biomedical Engineering, Eindhoven University of Technology,
WH 2.103, 5600 MB Eindhoven, The Netherlands
{T.H.J.M.Peeters, P.R.Rodrigues, A.Vilanova, B.M.terHaarRomeny}@tue.nl

Summary. Many different measures have been proposed to compute similarities and distances between diffusion tensors. These measures are commonly used for algorithms such as segmentation, registration, and quantitative analysis of Diffusion Tensor Imaging data sets. The results obtained from these algorithms are extremely dependent on the chosen measure. The measures presented in literature can be of complete different nature, and it is often difficult to predict the behavior of a given measure for a specific application. In this chapter, we classify and summarize the different measures that have been presented in literature. We also present a framework to analyze and compare the behavior of the measures according to several selected properties. We expect that this framework will help in the initial selection of a measure for a given application and to identify when the generation of a new measure is needed. This framework will also allow the comparison of new measures with existing ones.

1 Introduction

Diffusion tensor imaging (DTI) is a magnetic resonance technique that measures the diffusion of water in tissue. If the tissue is fibrous, the water molecules diffuse more in directions along the fibers than perpendicular to them. To capture the anisotropic behavior, the diffusion is often represented by a symmetric positive definite second-order tensor. Using tractography (e.g., see Vilanova et al. [15]) it is possible to reconstruct connections in the brain or the fibrous structure of muscle tissue such as the heart (e.g., see Zhukov et al. [19]). In several applications, for example, comparison between subjects, it is interesting to segment structures with a higher level of meaning, for example, white matter bundles, that is [14, 16, 20], and also to register different DTI data sets [1, 10, 18]. It is often also necessary to derive statistical properties of diffusion tensors (DTs) to identify differences, for example, between healthy and pathology areas [11].

In all the previous methods, it is often needed to define the difference between diffusion tensors, that is, to compare diffusion tensors. In segmentation and registration, similarity measures are applied to match DTs in voxels in a certain region, and between regions of different data sets. In quantitative analysis or DT statistics, distance or similarity measures of DTs in neighboring voxels can be used to classify the amount of variability in a selected voxel [13] or volume of interest. The results of these applications are highly dependent on the choice of measure.

Alexander et al. [1] listed several measures and analyzed their results for segmentation. However, since then, various people have introduced new measures for comparing DTs. These measures are of different nature and it is very difficult to predict which measure will give better or similar results. There exist numerous measures, but to our best knowledge, there is no overview that compares and classifies them in a structured way. This comparison can help to support researchers in choosing a measure, and being able to predict the behavior of the measures for their concrete application.

In this chapter, we provide this analysis and improve the intuition in the behavior of the measures. The intrinsic characteristics of a measure are analyzed without having a specific application in mind. This allows an evaluation of the nature of the measure in itself. It is out of the scope of this chapter to make an application-oriented analysis, (e.g., finding the best measure for DTI adult brain registration). However, this chapter aims to help make a first selection of the possible measures that could be used for such application by looking at the characteristics of the problem and the characteristics of the measures. We expect that it will also help to identify when a new measure is necessary, and compare its behavior with existing ones.

First, we present the notations used in this paper. In Sect. 3, we describe the properties that will be used for the analysis of the measures. In Sect. 4, we give an overview of existing measures from literature. In Sect. 5, we explain how we evaluate the properties of the measures and show some simple results to illustrate our methods. Section 6 presents the results of the experiments. Finally, in Sect. 7, conclusions and summarized results are described.

2 Notation

We represent symmetric positive definite second-order tensors, Sym_3^+, by capital bold letters, for example, $\mathbb{D} \in Sym_3^+$. The scalar components of a tensor $\mathbb{D}$ are denoted by $\mathbb{D}_{ij}$:

$$\mathbb{D} = \begin{pmatrix} \mathbb{D}_{11} & \mathbb{D}_{12} & \mathbb{D}_{13} \\ \mathbb{D}_{12} & \mathbb{D}_{22} & \mathbb{D}_{23} \\ \mathbb{D}_{13} & \mathbb{D}_{23} & \mathbb{D}_{33} \end{pmatrix}$$

Eigenvalues of tensor $\mathbb{D}$ are $\lambda_1^{\mathbb{D}} \geq \lambda_2^{\mathbb{D}} \geq \lambda_3^{\mathbb{D}} \geq 0$ and the corresponding eigenvectors are $\overline{e}_1^{\mathbb{D}}$, $\overline{e}_2^{\mathbb{D}}$, and $\overline{e}_3^{\mathbb{D}}$. We denote the trace $(\sum_{i=1}^{3} \mathbb{D}_{ii}) = (\sum_{i=1}^{3} \lambda_i^{\mathbb{D}})$ of $\mathbb{D}$ with $\mathrm{tr}(\mathbb{D})$. The determinant of $\mathbb{D}$ will be denoted by $\det(\mathbb{D})$.

With *measure* we refer to a function m that has two tensors $\mathbb{A}$, $\mathbb{B}$ as input, and returns a nonnegative scalar value:

$$m : Sym_3^+ \times Sym_3^+ \mapsto \mathbb{R}_0^+. \tag{1}$$

If a measure returns how similar $\mathbb{A}$ and $\mathbb{B}$ are, then we call the measure a *similarity measure*. If it returns how different $\mathbb{A}$ and $\mathbb{B}$ are, we call it a *distance measure*. We denote similarity measures with s and distance measures with d.

3 Properties

In this section, we present a list of properties that can be evaluated for the different measures. Diffusion tensors can be classified by their size, orientation, and shape. We evaluate the measures according to their sensitivity to changes in these properties. These changes are illustrated in Fig. 1. We also include as a property how robust the measures are to noise, and the fact that a measure is a metric or not.

3.1 Size

We understand as the size of a DT the mean diffusivity $MD = \mathrm{tr}(\mathbb{D})/3$. This is illustrated in Fig. 1(a). We consider a measure to be size-invariant if it is uniform scaling invariant, that is, if it fulfills

$$m(s\mathbb{A}, t\mathbb{B}) = m(\mathbb{A}, \mathbb{B}), \tag{2}$$

where s and t are scalar values.

3.2 Orientation

A measure m is rotation invariant if the value of m does not change when the input tensors are rotated:

$$m(R^T\mathbb{A}R, P^T\mathbb{B}P) = m(\mathbb{A}, \mathbb{B}), \tag{3}$$

where R and P are rotation matrices. The orientation invariance can be divided into two. One is whether the measure is sensitive, in general, to the difference in orientation between tensors. Orientation changes are illustrated in Fig. 1(b, c).

The other invariance included in the previous is invariance to image rotation. If we define a DTI image as $f : \mathbb{R}^3 \mapsto Sym_3^+$ in most of the cases, we want that our measure is invariant to rigid body transformations of f (i.e., rotation and translation). In the case of DTI images, the image transformation also has to be applied to the tensor. From these transformations the rotation is the only one that affects the tensor. Being invariant to image rotation means that we want to fulfill (3) when $R = P$. If the image f is transformed with other transformation (e.g., nonuniform scaling, skewing), it is not clear how this should affect the tensor and therefore we consider it out of the scope of this chapter.

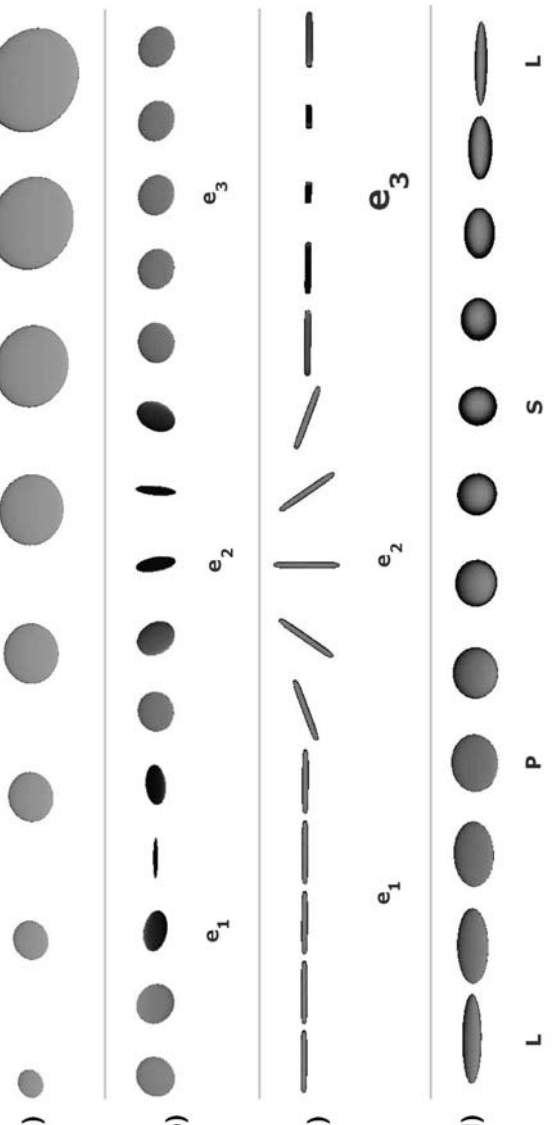


Fig. 1. Ellipsoidal glyphs showing smoothly varying DTs. (**a**) A planar-shaped DT where the size is increased smoothly, that is, MD increases; (**b**) a planar-shaped DT is rotated smoothly, until π, first around $\bar{e}_1$, then around $\bar{e}_2$, and finally around $\bar{e}_3$; (**c**) a linear-shaped DT is rotated smoothly as in (b); (**d**) DTs where the shape changes from linear (L), to planar (P), to spherical (S), and back to linear (L)

3.3 Shape

The shape of a tensor can be defined as *linear*, *planar*, *spherical*, or as an interpolation between these types. The shape is given by the ratio between the different eigenvalues. A graphical representation of different tensor shapes is shown in Fig. 1(d). A measure m is shape-invariant if the value of $m(\mathbb{A}, \mathbb{B})$ does not change when changing the shape (i.e., the ratio between eigenvalues) of $\mathbb{A}$, $\mathbb{B}$, or both.

3.4 Robustness

Measures are never completely invariant to noise. However, if small changes in the input produce small changes in output, then we consider the measures to be robust under noise. Therefore, we can define

$$|m(\mathbb{A} + \mathbb{E}1, \mathbb{B} + \mathbb{E}2) - m(\mathbb{A}, \mathbb{B})| \leq \varepsilon, \tag{4}$$

where ε is a very small scalar value and the components $\mathbb{E}\{1,2\}_{ij}$ of $\mathbb{E}\{1,2\} \in Sym_3$ are also very small values.

3.5 Metric

A distance measure d is a semi-metric if, for two tensors $\mathbb{A}$ and $\mathbb{B}$, it satisfies the following conditions:

$$\mathbb{A} = \mathbb{B} \Leftrightarrow d(\mathbb{A}, \mathbb{B}) = 0 \tag{5}$$

$$d(\mathbb{A}, \mathbb{B}) = d(\mathbb{B}, \mathbb{A}). \tag{6}$$

Condition (5) is important because it allow us to distinguish between equal and nonequal tensors. Condition (6) is necessary if we do not want the results to depend on the order in which we deal with the DTs in a volume. If the measure has to be a Riemannian metric, it also has to fulfill, for infinitesimally close $\mathbb{A}$ and $\mathbb{B}$,

$$d(\mathbb{A}, \mathbb{B}) \leq d(\mathbb{A}, \mathbb{C}) + d(\mathbb{C}, \mathbb{B}). \tag{7}$$

Condition (7) is important in applications where you need to take the mean or do interpolation between tensors [2, 12].

4 Measures

In this section, we present a classification of similarity and distance measures for diffusion tensors (DTs) that have been used in literature. This classification is based on the nature of the derivation of the measure: measures based on scalar indices; measures that make use of the angles between eigenvectors; measures based in linear algebra; measures based on imposing the preservation of positive definiteness of the tensor, that is, Riemannian geometry; measures

considering the DTs as a representation of a probability density function and, finally, measures that combine different measures from the previous classes.

4.1 Scalar Indices

Given a scalar index $g : Sym_3^+ \mapsto \mathbb{R}_0^+$, the simplest way to obtain a difference between two DTs $\mathbb{A}$ and $\mathbb{B}$ is by using the absolute difference $|g(\mathbb{A}) - g(\mathbb{B})|$ of the scalar index of the two tensors. There exist numerous scalar indices that can be chosen for g. Two well-known examples are fractional anisotropy (FA) and linear anisotropy (C_l). For a selection of scalar indices, see Table 1 and refer to Westin et al. [17] and Vilanova et al. [15]. These indices reduce the 6D information in a DT to a scalar value. In the computation of the scalar value, only the rotationally invariant eigenvalues of the tensors are used, thus they do not depict the directional variation of the diffusion anisotropy. The measures created from scalar indices will be denoted by ds, with the short name of the index as subscript, for example, ds_{FA}, ds_{C_l}, ds_{MD}. Thus

$$ds_{FA}(\mathbb{A}, \mathbb{B}) = |FA(\mathbb{A}) - FA(\mathbb{B})|. \tag{8}$$

When using ds, a lot of information is lost. Each DT is represented by one scalar value, while six scalar values are needed to represent the full DT. Thus, the measures based on scalar indices can be very limited.

More scalar indices can be derived from tensors. For example, several DTI literature recognized the benefit of *tensor invariants* as measure of the diffusion tensor shape that do not require diagonalization. Kindlmann [8] used these invariants, like the mean, variance, and skewness, which are invariant to rotation, to measure the shape gradients in tensor fields. However, using them for constructing a distance measure will give similar results to ds and will not solve the problem that just one aspect is being shown. Thus, we do not treat them separately here.

Table 1. Scalar indices for diffusion tensors [15, 17]

Name	Abbrev.	Equation
Mean diffusivity	MD	$= \mathrm{tr}(\mathbb{D})/3 = (\lambda_1 + \lambda_2 + \lambda_3)/3$
Fractional anisotropy	FA	$= \frac{\sqrt{(\lambda_1-\lambda_2)^2+(\lambda_2-\lambda_3)^2+(\lambda_1-\lambda_3)^2}}{\sqrt{2(\lambda_1^2+\lambda_2^2+\lambda_3^2)}}$
Relative anisotropy	RA	$= \frac{\sqrt{(\lambda_1-\lambda_2)^2+(\lambda_2-\lambda_3)^2+(\lambda_1-\lambda_3)^2}}{\sqrt{2}(\lambda_1+\lambda_2+\lambda_3)}$
Linear anisotropy	c_l	$= (\lambda_1 - \lambda_2)/(\lambda_1 + \lambda_2 + \lambda_3)$
Planar anisotropy	c_p	$= 2(\lambda_2 - \lambda_3)/(\lambda_1 + \lambda_2 + \lambda_3)$
Isotropy	c_s	$= 3\lambda_3/(\lambda_1 + \lambda_2 + \lambda_3)$
Volume Ratio	VR	$= \lambda_1\lambda_2\lambda_3/MD^3$

4.2 Angular Difference

Angular difference d_{ang_i} of the main eigenvectors $\overline{e}_i^{\mathbb{D}}$ is often used as a distance between tensors that measures change in orientation [21]:

$$d_{ang_i}(\mathbb{A}, \mathbb{B}) = \arccos(\overline{e}_i^{\mathbb{A}} \cdot \overline{e}_i^{\mathbb{B}}). \tag{9}$$

Using d_{ang_1} only makes sense for tensors where the diffusion is mainly linear. If the tensors have planar shape then d_{ang_3} can be used. For tensors with spherical shape, any d_{ang_i} can be considered random and should not be used.

4.3 Linear Algebra

A class of measures deal with the diffusion tensors components as vector elements. A typical distance measure is the L^n-norm of the componentwise difference of two vectors:

$$d_{Ln}(\mathbb{A}, \mathbb{B}) = \sqrt[n]{\sum_{i=1}^{3}\sum_{j=1}^{3}(\mathbb{A}_{ij} - \mathbb{B}_{ij})^n}. \tag{10}$$

In DTI literature, the L^2-norm, d_{L2}, is most commonly used for computing a distance measure (see Batchelor et al. [4]); therefore, we only treat d_{L2} in this chapter. d_{L2} is the same as the *Frobenius distance* [21], which is computed using $d_F(\mathbb{A}, \mathbb{B}) = \sqrt{\mathrm{tr}((\mathbb{A} - \mathbb{B})^2)}$.

One can also compute the *scalar product* of two tensors by summing the products of components of the tensors [1]. The result can be used as a similarity measure s_{sp}:

$$s_{sp}(\mathbb{A}, \mathbb{B}) = \sum_{i=1}^{3}\sum_{j=1}^{3}\mathbb{A}_{ij}\mathbb{B}_{ij}. \tag{11}$$

Measures s_{sp} and d_{Ln} treat the DTs as simple vectors and ignore the matrix or tensor nature of them. Another class of measures use the fact that we have matrices. Pierpaoli and Basser [13] propose to use the sum of the squared vector dot products of the eigenvectors weighted by the product of the eigenvalues as a tensor scalar product [7]: s_{tsp} includes the colinearity of the orientation of the tensors weighted by its eigenvalues. The value is maximized if the tensors are aligned.

$$s_{tsp}(\mathbb{A}, \mathbb{B}) = \sum_{i=1}^{3}\sum_{j=1}^{3}\lambda_i^{\mathbb{A}}\lambda_j^{\mathbb{B}}(\overline{e}_i^{\mathbb{A}}\overline{e}_j^{\mathbb{B}})^2. \tag{12}$$

This measure is also called tensor dot product [3]. It is used to construct the *lattice index*, which we show in Sect. 4.6. Jonasson et al. [7] use the normalized

tensor scalar product s_{ntsp} in order to make it invariant to scaling of the tensors:

$$s_{ntsp}(\mathbb{A}, \mathbb{B}) = \frac{s_{tsp}(\mathbb{A}, \mathbb{B})}{\text{tr}(\mathbb{A})\text{tr}(\mathbb{B})}. \tag{13}$$

Instead of applying the aforementioned measures to the tensors directly, they can also be applied to the *deviatoric* of the DTs (see, e.g., [1]). The deviatoric $\tilde{\mathbb{D}}$ of tensor $\mathbb{D}$ represents the nonisotropic part of $\mathbb{D}$. It expresses just the shape and orientation of the DT, independent of the size. It can be computed as follows:

$$\tilde{\mathbb{D}} = \mathbb{D} - \frac{1}{3}\text{tr}(\mathbb{D})\mathbb{I}, \tag{14}$$

where $\mathbb{I}$ is the identity matrix. Note that $\tilde{\mathbb{D}}$ is not a positive definite tensor everywhere. This means that it can have negative eigenvalues, and some of the measures will also give negative values.

4.4 Riemannian Geometry

If we constrain the matrices to positive definite matrices, we get another class of measures based on Riemannian geometry. Batchelor et al. [4] introduced a geometric-based distance d_g that measures the distance between two tensors in the space of positive definite tensors:

$$d_g(\mathbb{A}, \mathbb{B}) = N(\mathbb{A}^{-\frac{1}{2}}\mathbb{B}\mathbb{A}^{-\frac{1}{2}}), \tag{15}$$

where

$$N(\mathbb{D}) = \sqrt{\sum_{i=1}^{3}(\log(\lambda_i^{\mathbb{D}}))^2}. \tag{16}$$

This measures the distances along geodesics in the manifold of symmetric positive defined matrices. Additionally, it is invariant to any linear change of coordinates. Pennec et al. [12] introduce a similar framework with the same distance measure, and extend it with methods for filtering and regularization of tensor fields.

Arsigny et al. [2] introduce a new *Log-Euclidian* framework. It has similar theoretical properties as Pennec et al.'s framework, but with simpler and faster calculations. They derive the following Log-Euclidian distance measure d_{LE}:

$$d_{LE}(\mathbb{A}, \mathbb{B}) = \sqrt{\text{tr}((\log(\mathbb{A}) - \log(\mathbb{B}))^2)}. \tag{17}$$

This measure is equivalent to the d_{L2} of the logarithm of the matrices. The details of its computation and derivation can be found in Arsigny et al. [2].

4.5 Statistics

A diffusion tensor can be interpreted as the covariance matrix of a Gaussian distribution describing the local diffusion. Thus, a natural family of dissimilarity measures between DTs would be the statistical divergence that

measures the overlapping of probability density functions. Given a diffusion tensor $\mathbb{D}$, the displacement $\mathbf{r}$ of water molecules at time t is a random variable with the following probability density function (pdf):

$$p(\mathbf{r}|t,\mathbb{D}) = \frac{1}{\sqrt{(2\pi)^n \det(2t\mathbb{D})}} \mathrm{e}^{-(\mathbf{r}^T D^{-1}\mathbf{r})/(4t)},$$

where n is the dimensionality of the square matrix $\mathbb{D}$.

Wang and Vemuri [16] proposed to use the square-root of the J-divergence (symmetrized Kullback–Leibler) as a new definition of DT distance d_{KL}:

$$d_{KL}(\mathbb{A},\mathbb{B}) = \frac{1}{2}\sqrt{\mathrm{tr}(\mathbb{A}^{-1}\mathbb{B} + \mathbb{B}^{-1}\mathbb{A}) - 2n}, \tag{18}$$

where the dimensionality n is 3 for DTs.

In probability theory, class separability can be measured by the overlap between the corresponding pdfs. Therefore, the overlap of pdfs can also be used as a similarity measure between tensors. The calculation of the overlap cannot be done analytically and often approximations are being used. The Chernoff bound [6] gives us the upper bound of the probability error, $P(error)$, of a Bayesian classifier for two classes, w_1 and w_2, given their pdfs $P(w_1)$ and $P(w_2)$. For normal distributions we have

$$P(error) \leq P^{\beta}(w_1)P^{1-\beta}(w_2)\mathrm{e}^{-k\beta},$$

where β is a parameter that needs to be optimized to find the Chernoff bound. A special case is the Bhattacharyya bound where $\beta = 1/2$. This bound is never looser than the optimal Chernoff bound and can be directly calculated. For DTs, it becomes the following similarity measure:

$$s_{Bhat}(\mathbb{A},\mathbb{B}) = \mathrm{e}^{-\frac{1}{2}ln\left(\frac{\det(\frac{\mathbb{A}+\mathbb{B}}{2})}{\sqrt{\det(\mathbb{A})\det(\mathbb{B})}}\right)}. \tag{19}$$

4.6 Composed

As mentioned in Sect. 4.1, each scalar measure in itself can give very limited information of the difference between DTs (e.g., FA just gives information about the anisotropy). Usually, a measure that reflects the changes of a combination of these properties is necessary. Therefore, several authors have tried to combine simple measures to obtain a more complete measure. Often, the measures that are combined have quite different natures and therefore ad hoc normalizations and weighting factors are needed.

Pollari et al. [10] introduced shape-dependent similarity measures, which are used depending on the DT shape:

$$
\begin{aligned}
s_l(\mathbb{A},\mathbb{B}) &= |\overline{e}_1^{\mathbb{A}} \cdot \overline{e}_1^{\mathbb{B}}| = \cos(d_{ang_1}(\mathbb{A},\mathbb{B})) \\
s_p(\mathbb{A},\mathbb{B}) &= |\overline{e}_3^{\mathbb{A}} \cdot \overline{e}_3^{\mathbb{B}}| = \cos(d_{ang_3}(\mathbb{A},\mathbb{B})) \\
s_s(\mathbb{A},\mathbb{B}) &= 1 - \tfrac{|\mathrm{tr}(\mathbb{A})-\mathrm{tr}(\mathbb{B})|}{max(\mathrm{tr}(\mathbb{A}),\mathrm{tr}(\mathbb{B}),1)} \\
s_{T_2}(a,b) &= 1 - \tfrac{|g^a-g^b|}{max(g^a,g^b,1)},
\end{aligned}
$$

where a and b are the voxels with tensors $\mathbb{A}$ and $\mathbb{B}$. g^a, g^b are the grey-levels in a, b in the T_2 MRI data. Using s_{l,p,s,T_2}, Pollari et al. introduce a DT distance measure for registration of DTI brain datasets that looks at the overlap between diffusion shapes and weighs this with the most reliable information for that shape:

$$
\begin{aligned}
I(a,b) = \hat{c}_l^{\mathbb{A}}\hat{c}_l^{\mathbb{B}} s_l(\mathbb{A},\mathbb{B}) + \hat{c}_p^{\mathbb{A}}\hat{c}_p^{\mathbb{B}} s_p(\mathbb{A},\mathbb{B}) + \\
\gamma * \hat{c}_s^{\mathbb{A}}\hat{c}_s^{\mathbb{B}} \left(s_s(\mathbb{A},\mathbb{B}) + s_{T_2}(a,b)\right)/2,
\end{aligned} \tag{20}
$$

where γ is $\frac{1}{2}$ in all of their experiments because they want to give less weight to isotropic voxels. The anisotropy measures are defined as $\hat{c}_l = \frac{\lambda_1-\lambda_2}{\lambda_1}$, $\hat{c}_p = \frac{\lambda_2-\lambda_3}{\lambda_1}$, $\hat{c}_s = \frac{\lambda_3}{\lambda_1}$, which is a variation of the measures proposed by Westin [17] listed in Table 1. Because we are analyzing measures for DTs only, in Sect. 6 we use a modified similarity measure s_{pnl} that disregards the s_{T_2} term in (20):

$$
\begin{aligned}
s_{pnl}(\mathbb{A},\mathbb{B}) = \hat{c}_l^{\mathbb{A}}\hat{c}_l^{\mathbb{B}} s_l(\mathbb{A},\mathbb{B}) + \hat{c}_p^{\mathbb{A}}\hat{c}_p^{\mathbb{B}} s_p(\mathbb{A},\mathbb{B}) + \\
\gamma * \hat{c}_s^{\mathbb{A}}\hat{c}_s^{\mathbb{B}} s_s(\mathbb{A},\mathbb{B}).
\end{aligned} \tag{21}
$$

We also use $\gamma = \frac{1}{2}$, although a more precise analysis, of the robustness of this measure to the changes of gamma, would be needed.

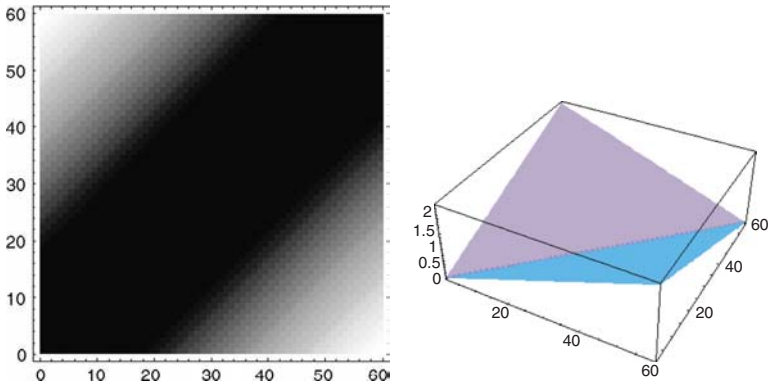
Pierpaoli and Basser [13] introduced the *lattice index* as an intervoxel anisotropy measure that takes the DTs in neighboring voxels into account. For the computation of the lattice index they defined a measure s_{LI} that gives a similarity between two tensors:

$$
s_{LI}(\mathbb{A},\mathbb{B}) = \frac{\sqrt{3}}{\sqrt{8}} \frac{\sqrt{s_{tsp}(\tilde{\mathbb{A}},\tilde{\mathbb{B}})}}{\sqrt{s_{tsp}(\mathbb{A},\mathbb{B})}} + \frac{3}{4} \frac{s_{tsp}(\tilde{\mathbb{A}},\tilde{\mathbb{B}})}{\sqrt{s_{tsp}(\mathbb{A},\mathbb{A})}\sqrt{s_{tsp}(\mathbb{B},\mathbb{B})}}, \tag{22}
$$

with s_{tsp} as defined in (12) and $\tilde{\mathbb{A}}, \tilde{\mathbb{B}}$ as in (14). Because $s_{tsp}(\tilde{\mathbb{A}}, \tilde{\mathbb{B}})$ can be negative, s_{LI} can give negative or imaginary values, which do not fulfill the basic description of a measure as we defined it. Therefore, we do not use s_{LI} in further analysis.

5 Methods

For analyzing the properties of the measures, we want to show the behavior of each measure for the different properties in a global way. So, we show the results of each measure for sets of pairs of DTs where one property is changed. We change each property gradually and analyze the behavior of the measures.



(a) (b)

Fig. 2. Size comparison plot of ds_{MD} for tensors with planar shape. On the axes from left to right, and from bottom to top, the size of the tensors increase while the shape and orientation are invariant. See Fig. 1(a). (**a**) A grey-value plot; (**b**) The height field

To do this analysis, we use plots as shown in Fig. 2. The axes of the plots have smoothly varying DTs and in the plot we show the similarity or difference of corresponding DTs. In Fig. 2(a) the results of the measure are shown as a grey-scale image. Figure 2(b) shows the same results as a height field, which gives a more clear impression about the evolution of the measure.

Furthermore, we compared the different measures by means of the root mean square difference (RMSD) of their normalized results. This allows us to grasp the similarities between the measures.

5.1 Size

Size is simple to evaluate because it can be captured with only one scalar value (mean diffusivity MD, see Table 1). Figure 2 shows a size comparison plot for ds_{MD}. From left to right and bottom to top, we increase the size of the tensor by multiplying the eigenvalues of a base tensor with increasing values. This is illustrated in Fig. 1(a). It can be seen from Fig. 2 that tensors with the same size (on the diagonal of the plot) have zero distance, and tensors of which the sizes differ have larger distances. Some measures (e.g., ds_{FA}) are invariant to scaling. So this plot will not be used for those measures.

5.2 Orientation

For orientation, we consider the sensitivity of the measure to rotation of the tensors, that is, rotation around any axis. For tensors with linear shape, the measure should be invariant to rotations around $\overline{e}_1$. For tensors with planar shape, the measure should be invariant to rotations around $\overline{e}_3$. For tensors

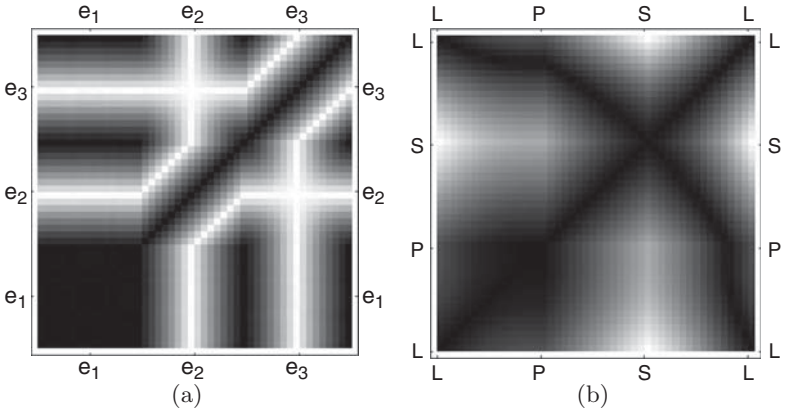


Fig. 3. (**a**) Comparison plot for tensor with linear shape rotated around $\overline{e}_1$, $\overline{e}_2$, and $\overline{e}_3$ showing d_{ang_1}. See Fig. 1(c); (**b**) ds_{FA} comparison plot for tensors with shape changing from linear (L) to planar (P) to spherical (S) to linear (L). See Fig. 1(d)

with spherical shape, the measure should be invariant to any rotation. We created plots for multiple types (linear, planar, spherical) of tensors, where on both axes, we gradually rotate the tensor around $\overline{e}_1$ until π. Then, on the middle part of the horizontal and vertical axes of the plots, we rotate around $\overline{e}_2$ until π. Finally, in the top and right of the two axes we rotate around $\overline{e}_3$ until π. The tensors on the axes of the plots are illustrated in Figs. 1(b, c). Figure 3(a) shows results for d_{ang_1}. The tensor used for this image was linear. Thus, for rotation around $\overline{e}_1$, the distance does not change. This can be seen in the image because in the lower-left part, the distances stay zero. When rotating the tensor around $\overline{e}_2$ and $\overline{e}_3$, it can be seen that the distance between measures gradually increases for a rotation up to $\pi/2$ and then decreases again until it is zero at π.

Furthermore, we tested the rotation invariance of the measures to the situation when we rotate the volume. We did this by applying the same rotation to every tensor in a set, and then computing the RMSD of these results to the corresponding ones without rotation.

5.3 Shape

We consider that DTs can have linear, planar, or spherical shape or a shape that is a combination of these shapes. To study the behavior of the measure under changes in shape, we start with tensors that have linear shape, and then gradually change the shape to respectively planar, spherical, and back to linear. This is illustrated in Fig. 1(d). To make sure that we are only evaluating shape, we do not change the size and orientation of the tensors in the same plot. Results for ds_{FA} are shown in Fig. 3(b). As can be seen from the black

areas in the plot, which are not in the diagonal, tensors with different shapes can have the same value for FA. This is a known property of FA.

5.4 Robustness

From the results of the previous methods, we can derive whether a measure is sensitive to small changes in one of the properties. In addition, we introduce a small variation to the set of tensors in our experiments. To each component of the input tensors (on both axes) for making the size, shape, and orientation plots, we add a uniformly distributed random value. Then we analyze this robustness by computing the root mean square difference (RMSD) between the plots with and without the added noise. We consider the measures robust to noise if its plots do now show sharp changes or discontinuities, and the computed RMSDs are relatively small.

5.5 Metric

The conditions that need to be fulfilled for a measure to be a Riemannian metric can be derived from its definition. Thus, no experiments are needed to evaluate this property. However, we summarize whether the properties in (5)–(7) are fulfilled for each of the measures.

6 Experiments

In this section, we analyze and categorize behavior of the different measures using the methods described in the previous section. The behavior of the measures is summarized in Table 2.

6.1 Size

We can observe four different behaviors for the measures with respect to the size difference of the tensors. All scalar measures listed in Table 1, except MD, are invariant to scaling one or both input tensors with a scalar s. d_{ang_i}, s_{ntsp}, and s_{LI} are also invariant to scaling.

The measures ds_{MD} and d_{L2} show a behavior as illustrated in Fig. 2. The relation between the differences in size (MD) and the computed difference behaves as follows

$$d(\mathbb{A}, \mathbb{B}) = s \times |MD(\mathbb{A}) - MD(\mathbb{B})|, \tag{23}$$

where s is a scalar value. Scaling both $\mathbb{A}$ and $\mathbb{B}$ with a scalar value will change the outcome of ds_{MD} and d_{L2}, and so (24) is not valid for those measures if $\mathbb{A} \neq \mathbb{B}$:

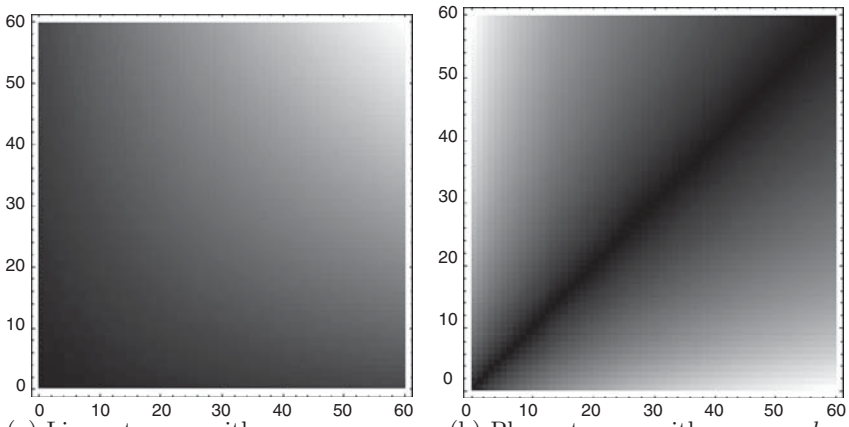
$$d(s\mathbb{A}, s\mathbb{B}) = d(\mathbb{A}, \mathbb{B}). \tag{24}$$

This behavior is listed as *add* in Table 2.

Table 2. Overview of properties for similarity and distance measures for diffusion tensors

	Equation	Eqn	Cite	Size	Orientation	Shape	Robustness	Metric
ds_{FA}	$\lvert FA(\mathbb{A}) - FA(\mathbb{B})\rvert$	(8)	[17]	Invariant	Invariant	Not good	Shape only	No
ds_{MD}	$\lvert MD(\mathbb{A}) - MD(\mathbb{B})\rvert$	(8)	[17]	Add	Invariant	Invariant	Size only	No
d_{ang_1}	$\arccos(\overline{e}_1^{\mathbb{A}} \cdot \overline{e}_1^{\mathbb{B}})$	(9)	[21]	Invariant	Linear only	Linear only	Linear only	No
d_{L2}	$\sqrt[n]{\sum_{i=1}^{3}\sum_{j=1}^{3}(\mathbb{A}_{ij} - \mathbb{B}_{ij})^n}$	(10)	[4]	Add	Smooth	Ok	All	Yes
s_{sp}	$\sum_{i=1}^{3}\sum_{j=1}^{3}\mathbb{A}_{ij}\mathbb{B}_{ij}$	(11)	[1]	Increases	Smooth	Not self-similar (1)	All	No
s_{tsp}	$\sum_{i=1}^{3}\sum_{j=1}^{3}\lambda_i^{\mathbb{A}}\lambda_j^{\mathbb{B}}(\overline{e}_i^{\mathbb{A}}\overline{e}_j^{\mathbb{B}})^2$	(12)	[7]	Increases	Smooth	Not self-similar (1)	All	No
s_{ntsp}	$\frac{s_{tsp}(\mathbb{A},\mathbb{B})}{\mathrm{tr}(\mathbb{A})\mathrm{tr}(\mathbb{B})}$	(13)	[7]	Invariant	Smooth	Not self-similar (1)	Shape and orientation only	no
s_{pnl}	$\hat{c}_l^{\mathbb{A}}\hat{c}_l^{\mathbb{B}}s_l(\mathbb{A},\mathbb{B}) + \hat{c}_p^{\mathbb{A}}\hat{c}_p^{\mathbb{B}}s_p(\mathbb{A},\mathbb{B}) + \gamma * \hat{c}_s^{\mathbb{A}}\hat{c}_s^{\mathbb{B}}s_s(\mathbb{A},\mathbb{B})$	(21)	[10]	mult	smooth	not self-similar (2)	shape and orientation only	no
d_g	$N(\mathbb{A}^{-\frac{1}{2}}\mathbb{B}\mathbb{A}^{-\frac{1}{2}})$	(15)	[4]	Mult	Smooth	Sensitive	Sensitive	Yes
d_{LE}	$\sqrt{\mathrm{tr}((\log(\mathbb{A}) - \log(\mathbb{B}))^2)}$	(17)	[2]	Mult	Smooth	Sensitive	Sensitive	Yes
d_{KL}	$\frac{1}{2}\sqrt{\mathrm{tr}(\mathbb{A}^{-1}\mathbb{B} + \mathbb{B}^{-1}\mathbb{A}) - 2n}$	(18)	[16]	Mult	Smooth	Sensitive	Sensitive	Yes
s_{Bhat}	$e^{-\frac{1}{2}ln\left(\frac{\det(\frac{\mathbb{A}+\mathbb{B}}{2})}{\sqrt{\det(\mathbb{A})\det(\mathbb{B})}}\right)}$	(19)	[6]	mult	smooth	sensitive	sensitive	yes

The ds measures that are not listed have the same properties as ds_{FA}



(a) Linear tensors with measure s_{tsp}. (b) Planar tensors with measure d_g.

Fig. 4. Comparison plots with tensors changing size

Measures s_{sp} and s_{tsp} have behavior as shown in Fig. 4(a). They return bigger values if the MD is larger. As a consequence, the result of comparing two tensor that are exactly the same is not constant and depends on the size of the tensor. There is no upper limit for the similarity measure that can be given. We call the behavior of the measures *increasing* in Table 2.

The remaining measures are listed as *mult*. This means that they behave as shown in Fig. 4(b). The relation between the output of the measure and the ratio $MD(\mathbb{A})/MD(\mathbb{B})$ of the size of the two tensors is linear, and (24) is valid.

6.2 Orientation

Measure d_{ang_1} only works well for tensors with linear shape. All scalar-index-based measures (ds) are invariant to rotation. All measures except ds and d_{ang_i} have similar behavior under rotation. For tensors with linear shape, they have the same behavior as d_{ang_1}, which is shown in Fig. 3(a). Results for s_{KL} for tensors with planar shape is shown in Fig. 5. The other measures show similar behavior. It is similar to that in Fig. 3(a). Except for d_{ang_i}, all measures are invariant to rotations if at least one of the two tensors that are being compared has spherical shape, that is,

$$m(\mathbb{S}, \mathbb{A}) = m(\mathbb{S}, R^T \mathbb{A} R) \tag{25}$$

for spherical tensor $\mathbb{S}$, and $\mathbb{A} \in Sym^+(3)$ and rotation R. For tensors whose shape is not purely linear, planar, or spherical, the resulting plots are a weighted average of the plots of the respective tensor types. This is shown in Fig. 6 for d_{LE}. All measures, except d_{ang_i} in the areas where its result is random, are invariant to rotations of both tensors, thus, for any rotation R

$$m(\mathbb{A}, \mathbb{B}) = m(R^T \mathbb{A} R, R^T \mathbb{B} R) \tag{26}$$

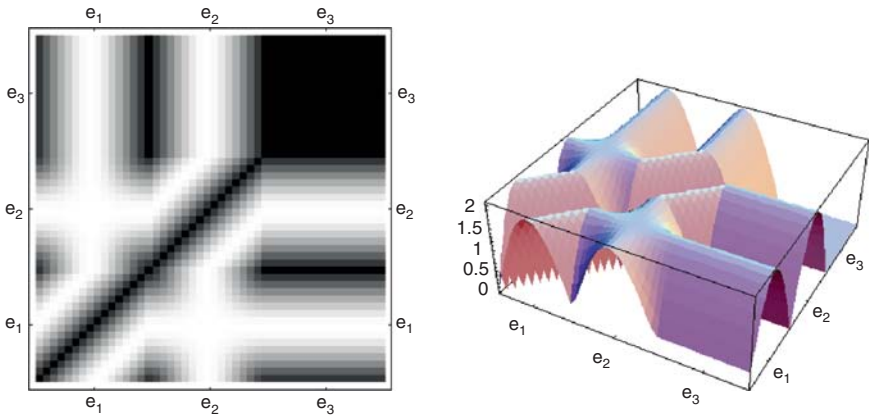


Fig. 5. Comparison plot of planar tensors rotated around λ_1, λ_2, and λ_3 for d_{KL}. See Fig. 1(b)

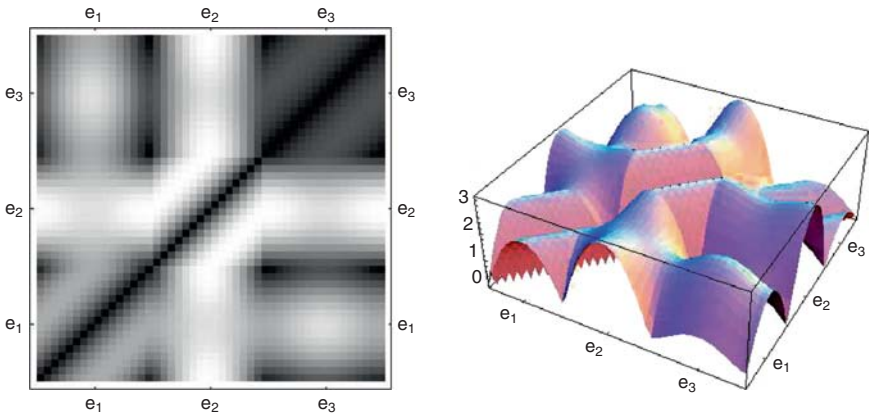


Fig. 6. Comparison plot for tensors rotated around $\bar{e}_1$, $\bar{e}_2$, and $\bar{e}_3$ for d_{LE}. The tensors do not have pure linear, planar, or spherical shape, but eigenvalues $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, and $\lambda_3 = 0.1$

To refine the classification of these measures, we compared their results by computing the RMSD between them. Measures d_{L2}, d_g, d_{LE}, d_{KL}, and s_{Bhat} (since s_{Bhat} it is not a distance, we inverted the result, $d_{Bhat} = 1 - s_{Bhat}$, before the comparison) are similar to each other ($RMSD \approx 0$). We can define another subgroup with the measures s_{sp}, s_{tsp}, and s_{ntsp}. These measures give the same result, $RMSD = 0$.

6.3 Shape

Of the measures that we analyzed, only ds_{MD} is invariant to shape changes. The behavior for ds_{FA} is shown in Fig. 3(b). The other ds measures show similar behavior where tensors that differ can have a distance of zero depending on which anisotropy measure is used. d_{ang_i} can give random values depending on the shape of the diffusion. The behavior of d_{L2} is shown in Fig. 7, The diagonal is black, and the greatest distance occurs between linear and spherical tensors.

Measures s_{sp}, s_{tsp}, and s_{ntsp} all behave similar to what is shown in Fig. 8. Tensors with linear shape are very similar to themselves. However, tensors with planar or spherical shapes are less similar to themselves. Thus, the similarity between a tensor and itself depends on its shape. Because of this

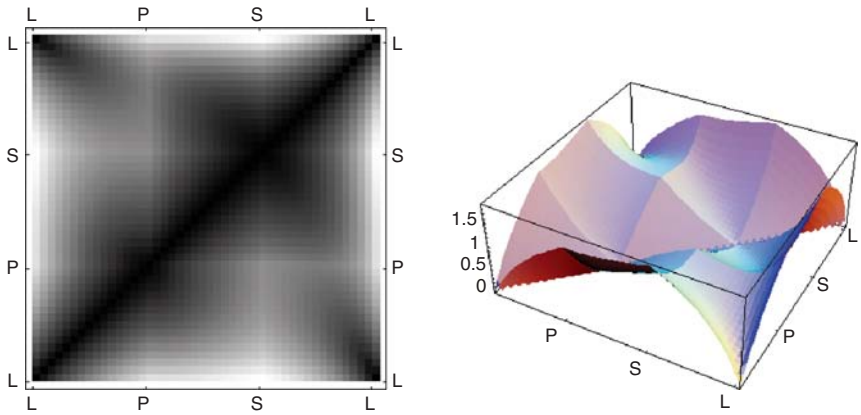


Fig. 7. Comparison plot of d_{L2} for shapes changing from linear to planar to spherical to linear

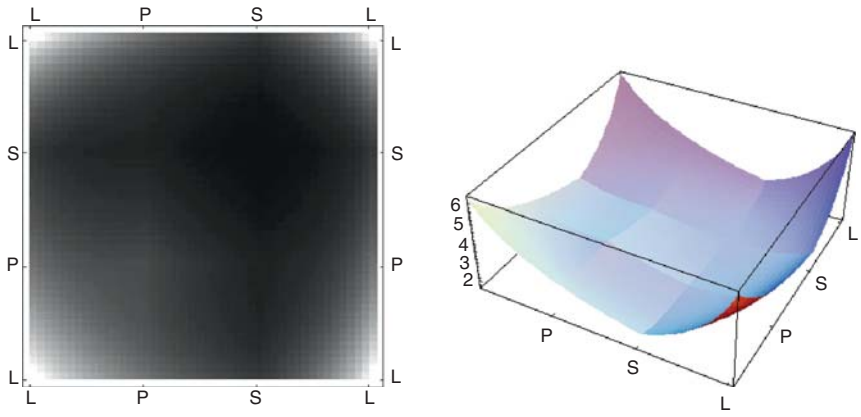


Fig. 8. Comparison plot for changing shapes for S_{tsp}

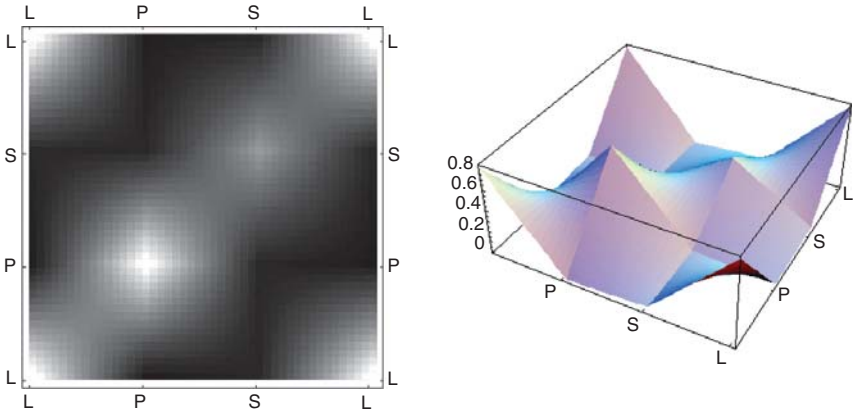


Fig. 9. Comparison plot for changing shapes for S_{pnl}

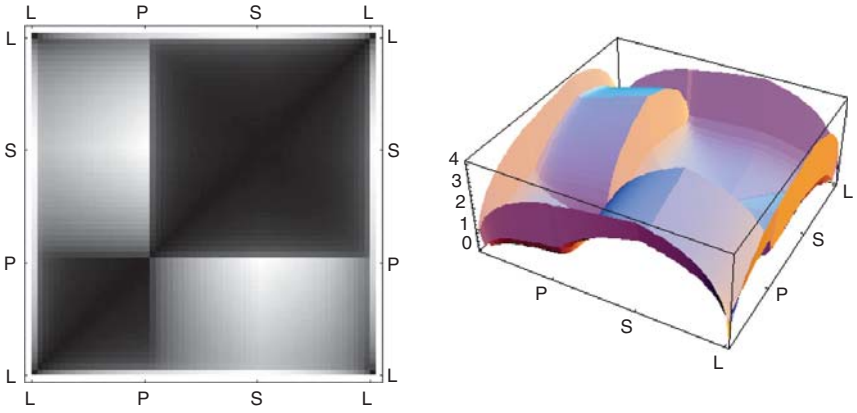


Fig. 10. Comparison plot for shapes changing from linear (L) to planar (P) to spherical (S) to linear (L). The plot shows d_{KL}. $\lambda_1/\lambda_3 = 100$ for tensors with "pure" linear and planar shape

behavior, we cannot convert these similarity measures to distance measures, which fulfill metric condition (5). We list this in Table 2 as *not self-similar (1)*.

The behavior under shape changes for s_{pnl} is shown in Fig. 9. The values on the diagonal are brighter than the values next to it because tensors are similar to themselves. However, the actual values on the diagonal are not all the same. Thus, for s_{pnl} the similarity between a tensor $\mathbb{D}$ and itself also depends on the shape of $\mathbb{D}$. We list this in Table 2 as *not self-similar (2)*.

The plots for measures based on Riemannian geometry and statistics (see Sects. 4.4 and 4.5) show steep edges in areas where at least one of the eigenvalues is very small. This is shown for d_{KL} in Fig. 10, where $\lambda_1/\lambda_3 = 100$ for

tensors with "pure" linear and planar shape. This behavior is listed as *sensitive* in Table 2. In medical data, the chance to be exactly on the very steep part is small because the fractions between eigenvalues are not that large. However, it is always possible that two similar tensors are on opposite sides of this edge, which will result in a large difference. Also, noise in medical data can change the fractions of the eigenvalues in such a way that the tensors come closer to the steep edges, that is, small variations in the shape results in large variation in the measures.

6.4 Robustness

We repeated the experiments of the previous sections after adding noise as described in Sect. 5.4 to the input tensors. The noise consists of uniformly distributed random values $\varepsilon \in [-0.01, 0.01]$, which are added to the components of the tensors. We then compare the root mean square difference (RMSD) between the output of the normalized plots with and without noise. The results are shown in Table 3. The more robust the measures are to noise, the lower the values in the table.

The shape experiments were done with tensors that have varying eigenvalues but with constant mean diffusivity, $\frac{\lambda_1+\lambda_2+\lambda_3}{3} = 1$. The eigenvalues are changed from linear ($\lambda_1 > \lambda_2 = \lambda_3$) to planar shape ($\lambda_1 = \lambda_2 > \lambda_3$), from planar to spherical shape ($\lambda_1 = \lambda_2 = \lambda_3$), and back to linear shape.

The orientation experiments were done with a linear tensor ($\lambda_1 = 1.0$, $\lambda_2 = \lambda_3 = 0.1$) that is rotated. The size experiments use the same linear tensor, which is enlarged by multiplying all components of the tensor with values from 0 to 60. The noise is added to the tensors after the changes in shape, orientation, and size were done.

In Table 2 the robustness of the measures is summarized. Some measures prove to be robust within only one or two of the invariant properties (shape,

Table 3. Root mean square difference (RMSD) between the sets of tensors with and without small variations

	Equation	Shape	Orientation	Size
ds_{FA}	(8)	0.007	0.372	0.279
ds_{MD}	(8)	0.316	0.332	0.002
d_{ang_1}	(9)	0.409	0.006	0.338
d_{L2}	(10)	0.005	0.013	0.002
s_{sp}	(11)	0.009	0.024	0.002
s_{tsp}	(12)	0.009	0.024	0.002
s_{ntsp}	(13)	0.003	0.024	0.246
s_{pnl}	(21)	0.008	0.018	0.237
d_g	(15)	0.012	0.044	0.007
d_{LE}	(17)	0.012	0.042	0.007
d_{KL}	(18)	0.013	0.056	0.007
s_{Bhat}	(19)	0.014	0.048	0.006

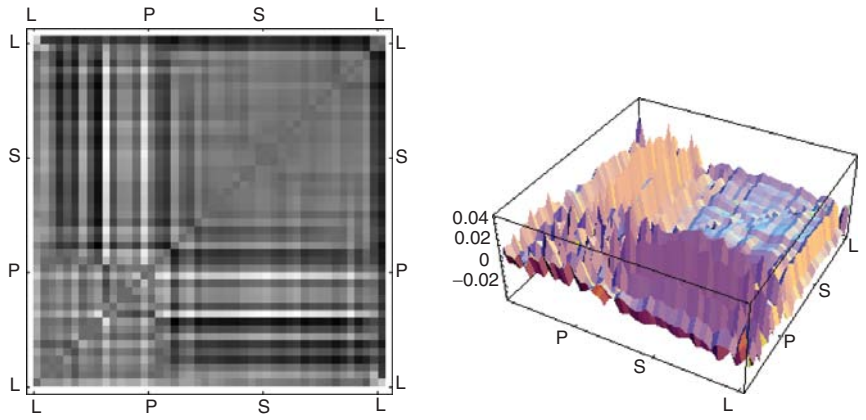


Fig. 11. Comparison plot showing the difference between the response of d_{KL} to a set with and without random noise. $RMSD = 0.013$

orientation, and size), that is, they have one or two relatively low values in Table 3. We classify them as such. For example, measure ds_{FA} is robust to changes in *shape only*. If the plots do not have steep parts, that is, high discontinuities, thus the values in Table 3 are small, we consider the noise robustness of the measures to be good for *all*.

Measure d_{L2} proves to be the most robust measure. Measure d_{ang_1} only takes the main diffusion direction into account. If the shape of the tensor is not linear, this direction can change randomly when small changes are made to the tensors. Thus, d_{ang_1} does not behave well under noise.

From the other measures, only the shape plots for d_g, d_{LE}, d_{KL}, and s_{Bhat} show steep edges. These edges appear where the shapes of the tensors are very linear or planar. Thus, in these areas the measures are very sensitive to noise. Figure 11 shows this behavior for d_{KL}.

6.5 Metric

All metrics are symmetric, this can be also seen in the plots, since they are symmetric by the diagonal.

Similarity measures have to be transformed into distance measures before we can evaluate if they are metrics. Similarity measures with *increases* as size change behavior or *not self-similar* as shape change behavior have a similarity $s(\mathbb{A}, \mathbb{A})$ that depends on the size or shape of tensor $\mathbb{A}$. Thus they cannot directly be translated into a distance measure that always fulfills metric condition (5). Measures ds, d_{ang_i}, and s_{ntsp} are invariant to one or more of the properties of Sect. 3. Thus, there are many tensors $\mathbb{A} \neq \mathbb{B}$ for which $d(\mathbb{A}, \mathbb{B}) = 0$, which invalidates metric condition (5).

It is clear that distance measures d_{L2}, d_g, d_{LE}, d_{KL} fulfill metric conditions (5) and (6). They also fulfill the triangle inequality (7) if the tensors $\mathbb{A}$ and $\mathbb{B}$

are infinitesimally close [2, 12, 16]. Therefore, they are *Riemannian metrics*. Because this is sufficient for the applications that need the distances to be a metric, we list them as *yes* for the metric property in Table 2. s_{Bhat} is a similarity measure, and so it cannot be a metric. However, it can be turned into one as is shown in [5]. Thus, we list *yes* for the metric property in Table 2.

7 Conclusions

Depending on the application, different distance or similarity measures can be used. Using the previous analysis of properties we can identify from a practical point of view the differences and similarities between the different measures.

It turns out that the behavior of s_{sp} and s_{tsp} is similar, even though s_{sp} deals with the tensor as if it is a vector. The L^2 distance d_{L2} is relatively simple, but shows good behavior. Also, all measures listed in Sects. 4.4 and 4.5 give practically the same results.

Except for s_{Bhat}, the similarity measures S cannot easily be converted into metrics, thus if that is a requirement for the application (e.g., for calculating geodesics), those measures are ruled out. This also rules out the ds measures and d_{ang_i}. Measure d_{L2} can be a good measure in that case. When using measures d_g, d_{LE}, d_{KL}, and s_{Bhat}, one has to be careful with the sensitivity to small shape changes close to the degenerate cases.

Throughout a complete brain, all diffusion properties vary. To take all properties into account when registering brains, no measure should be chosen that is invariant to any of them. Also, if the weighting for all DTs used in the registration must be the same, the similarity measures that list *increases* for size or *not self-similar* for shape should not be used because even for equal DTs, the computed similarities can vary depending on size and shape.

For interpolation of DTs the triangle inequality condition must be satisfied; therefore, only measures that are metrics can be used. Work has been done in the comparison of the different interpolation methods as in Arsigny et al. [2], Pennec et al. [12], and Kindlmann et al. [9].

We created an overview of existing distance and similarity measures for matching diffusion tensors and classified the measures. Such an overview, including recently introduced measures, was not previously available. We evaluated the properties of these measures and summed them up in Table 2. When researchers want to use a similarity or distance measure for their concrete application, they can define which properties their measure should have to and then study the measures that fulfill their requirements.

When new measures are introduced, it will be beneficial to classify them and see for which properties they differ from already existing measures, and how they differ. So in which sense they improve existing measures.

This chapter aims to help in the first selection of these measures, the next step is to test what measure performs better in a concrete application, for

example, white matter segmentation. If the goal is to segment the brain using DTs, the choice of measure depends on which properties are of importance for a given area. For example, segmenting the thalamic nuclei requires dependency of orientation for the measure used [21], while white and grey matter can be distinguished using the tensor shape.

For future work, it would be necessary to study the behavior of these measures under a more realistic model of noise. Furthermore, it would be useful to compare the measures in a concrete practical set up and see whether they behave according to our expectations. Also for some applications, such as segmentation and registration, it can prove useful to compare methods that apply the measures to the DTs directly with methods that segment or register derived data, such as fibers (the output of a fiber tracking algorithm). We also plan to use the measures for quantitative analysis and visualization of differences between tensors in small areas of the heart and brain DTI datasets. We expect that the overview of properties, that we presented in this work, will simplify the analysis of the results for these applications. It can also give indications for which properties a new measure for DTs might be useful. For example, having a measure of which the result for size changes is fractional (listed as *ratio* in Table 2), where the shape-dependency is not *sensitive*, can prove useful.

Acknowledgments

We thank Laura Astola for her insights in statistics and Riemannian geometry. This work was supported by the Dutch BSIK program entitled Molecular Imaging of Ischemic heart disease (project number BSIK 03033), Fundação para a Ciência e a Tecnologia (FCT, Portugal) under grant SFRH/BD/24467/2005, and the Netherlands Organization for Scientific Research (NWO-VENI grant 639.021.407).

References

[1] D. Alexander, J. Gee, and R. Bajcsy. Similarity measures for matching diffusion tensor images. In *Proceedings of the British Machine Vision Conference (BMVC)*, 93–102, 1999.

[2] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, 2006.

[3] P.J. Basser and C. Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of Magnetic Resonance*, 111(3):209–219, 1996.

[4] P.G. Batchelor, M. Moakher, D. Atkinson, F. Clamante, and A. Connelly. A rigorous framework for diffusion tensor calculus. *Magnetic Resonance in Medicine*, 53:221–225, 2005.

[5] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:564–577, 2003.

[6] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification (2nd Edition)*. Wiley, New York, 2000.

[7] L. Jonasson, X. Bresson, P. Hagmann, O. Cuisenaire, R. Meuli, and J. Thiran. White matter fiber tract segmentation in DT-MRI using geometric flows. *Medical Image Analysis*, 9(3):223–236, 2005.

[8] G. Kindlmann. *Visualization and Analysis of Diffusion Tensor Fields*. PhD thesis, School of Computing, University of Utah, 2004.

[9] G.L. Kindlmann, R.S.J. Estpar, M. Niethammer, S. Haker, and C.-F. Westin. Geodesic-loxodromes for diffusion tensor interpolation and difference measurement. In N. Ayache, S. Ourselin, and A. Maeder, editors, *MICCAI (1)*, volume 4791 of *Lecture Notes in Computer Science*, pages 1–9. Springer, 2007.

[10] J. Lötjönen, M. Pollari, T. Neuvonen. Affine registration of diffusion tensor MR images. In *MICCAI 2006*, Springer, Berlin Heidelberg, pp. 629–636, 2006.

[11] E.R. Melhem, S. Mori, G. Mukundan, M.A. Kraut, M.G. Pomper, and P.C.M. van Zijl. Diffusion tensor MR imaging of the brain and white matter tractography. *American Journal of Roentgenology*, 178(1):3–16, 2002.

[12] X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

[13] C. Pierpaoli and P.J. Basser. Toward a quantitative assessment of diffusion anisotropy. *Magnetic Resonance in Medicine*, 36:893–906, 1996.

[14] M. Rousson, C. Lenglet, and R. Deriche. Level set and region based surface propagation for diffusion tensor MRI segmentation. In *Computer Vision Approaches to Medical Image Analysis (CVAMIA) and Mathematical Methods in Biomedical Image Analysis (MMBIA) Workshop*, Prague, May 2004.

[15] A. Vilanova, S. Zhang, G. Kindlmann, and D. Laidlaw. An introduction to visualization of diffusion tensor imaging and its applications. In J. Weickert and H. Hagen, editors, *Visualization and Processing of Tensor Fields*, Mathematics and Visualization, chapter 7, Springer, Berlin Heidelberg, pp. 121–153, 2005.

[16] Z. Wang and B.C. Vemuri. DTI segmentation using an information theoretic tensor dissimilarity measure. *IEEE Transactions on Medical Imaging*, 24(10):1267–1277, 2005.

[17] C.-F. Westin, S. Peled, H. Gudbjartsson, R. Kikinis, and F.A. Jolesz. Geometrical diffusion measures for MRI from tensor basis analysis. In *ISMRM '97*, pp. 1742, 1997.

[18] H. Zhang, P.A. Yushkevich, D.C. Alexander, and J.C. Gee. Deformable registration of diffusion tensor MR images with explicit orientation optimization. *Medical Image Analysis – Special Issue: The Eighth*

International Conference on Medical Imaging and Computer Assisted intervention – MICCAI 2005, 10(5):764–785, October 2006. Invited submission. PMID: 16899392.

[19] L. Zhukov and A.H. Barr. Heart-muscle fiber reconstruction from diffusion tensor MRI. In *Proceedings of IEEE Visualization 2003*, IEEE Computer Society, pp. 597–602, 2003.

[20] L. Zhukov, K. Museth, D. Breen, R. Whitaker, and A. Barr. Level set modeling and segmentation of DT-MRI brain data, *Journal of Electronic Imaging*, 12:125–133, 2003.

[21] U. Ziyan, D. Tuch, and C. Westin. Segmentation of thalamic nuclei from DTI using spectral clustering. In *Ninth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'06)*, Lecture Notes in Computer Science 4191, Copenhagen, Denmark, pp. 807–814, 2006.

Tensor Glyph Warping: Visualizing Metric Tensor Fields using Riemannian Exponential Maps

Anders Brun[1,2,3] and Hans Knutsson[1,2]

[1] Department of Biomedical Engineering, Linköping University, Sweden
[2] Center for Medical Image Science and Visualization (CMIV), Linköping, Sweden
[3] Centre for Image Analysis, Uppsala, Sweden
anders@cb.uu.se, knutte@imt.liu.se

Summary. The Riemannian exponential map, and its inverse the Riemannian logarithm map, can be used to visualize metric tensor fields. In this chapter we first derive the well-known metric sphere glyph from the geodesic equation, where the tensor field to be visualized is regarded as the metric of a manifold. These glyphs capture the appearance of the tensors relative to the coordinate system of the human observer. We then introduce two new concepts for metric tensor field visualization: geodesic spheres and geodesically warped glyphs. These extensions make it possible not only to visualize tensor anisotropy, but also the curvature and change in tensor-shape in a local neighborhood. The framework is based on the $\exp_p(v^i)$ and $\log_p(q)$ maps, which can be computed by solving a second-order ordinary differential equation (ODE) or by manipulating the geodesic distance function. The latter can be found by solving the eikonal equation, a nonlinear partial differential equation (PDE), or it can be derived analytically for some manifolds. To avoid heavy calculations, we also include first- and second-order Taylor approximations to exp and log. In our experiments, these are shown to be sufficiently accurate to produce glyphs that visually characterize anisotropy, curvature, and shape-derivatives in sufficiently smooth tensor fields where most glyphs are relatively similar in size.

1 Introduction

The need for tensor visualization has grown over the past 20 years along with the advancement of image analysis, computer graphics, and visualization techniques. From being an abstract mathematical entity known mostly by experts in continuum mechanics and general relativity, tensors are now widely used and visualized in applied fields such as image analysis and geology. In particular, there has been an expansion over the years, from using tensors mostly in mathematical theories of the world, toward estimating tensor quantities from experimental data. See for instance [26] for a recent survey of both techniques and applications.

We propose a technique to paint tensor glyphs in a special "warped" or deformed coordinate system to enhance the visualization of curvature and partial derivatives of a metric tensor field. Glyphs are commonly used to represent the state of a tensor field pointwise and their collective behavior, when, for example, arranged in a grid, help to perceptualize the change of shape and orientation of the tensors. Inspired by others who have visualized tensor fields by regarding the tensor field as the metric of a manifold, see for instance [8, 11, 15, 16, 19, 25], we propose a method to warp the glyphs in accordance with the metric.

One of the most exciting areas where tensor data is derived from experiments is the medical imaging modality called diffusion tensor MRI (DT-MRI). It is now becoming so central that clinical radiologists in general need to understand and visualize tensor fields representing in vivo water diffusion in the human brain. Fortunately, the positive definite matrices found in DT-MRI data can be visualized using ellipses (2D) or ellipsoids (3D), making the data understandable without knowing the details of tensor algebra. In DT-MRI, the ellipsoids are elongated along the directions of maximum water diffusion, and it turns out that their shape is tightly connected to anatomical properties of the tissue being studied. In the human brain, for instance, they are elongated in the directions of nerve fiber bundles in white matter, because water diffusion is restricted in the directions perpendicular to the fibers. In the ventricles on the other hand, where the water molecules in the cerebrospinal fluid (CSF) diffuse freely in all three directions, the ellipsoids are large and spherical. These properties of ellipsoid glyphs make DT-MRI datasets easier to comprehend for a medical expert. Similar tensor fields are found in other medical settings, for example, in elastography analysis that is reviewed later in the chapter "A Tensor Approach to Elastography Analysis and Visualization."

Tensors are mathematical objects with special geometrical properties. Most of the research in tensor visualization has focused on the most commonly used low order tensors, in particular vectors (first order, 1D arrays) and matrices (second order, 2D arrays). In this chapter, we study the visualization of metric tensor fields in $\mathbb{R}^n$, where each tensor is a second-order tensor. These can be represented by $n \times n$ matrices, elements of $\mathbb{R}^n \otimes \mathbb{R}^n$, which are symmetric and positive definite, that is, they have positive eigenvalues. We call these tensor fields "metric tensor fields," since they may be interpreted as the metric of a Riemannian manifold. In DT-MRI, the pointwise inverse of the tensor field is the most natural metric to consider [19]. Structure tensor fields in image analysis, see for instance [10], is another example of tensor fields that can be interpreted as a metric.

2 Glyphs and Glyph Warping

A tensor glyph is a geometric object that graphically represents the local characteristics of the tensor field in a point. If a prefiltering is applied to the field, in which the field is averaged over a small neighborhood to avoid

aliasing effects or noise, one can argue that the glyph, in many applications, actually represents the average characteristics of the tensor field over a small area. The topic of this chapter is "glyph warping," which refers to a process of deforming glyphs according to the local characteristics of a tensor field or manifold. By construction, most kinds of tensor glyphs are scaled according to the tensor in a point, regarding this tensor as the local metric. The eigenvalues and eigenvectors of the tensor determine the stretching and rotation of the different axes of a glyph template, or determine the shape of the glyph in other ways. Ellipses (2D) and ellipsoids (3D) can be constructed from a circular or spherical template. Other glyphs, such as "space-ship glyphs" found in [28] or superquadric tensor glyphs [17], are not rotationally symmetric, but have three or more symmetry axes.

The curvature and the partial derivatives of the metric tensor field are usually not represented by individual glyphs. These pointwise properties can instead be inferred from the collective appearance of the glyphs in a neighborhood, that is, if the glyphs tend to rotate or change in size locally. However, these properties could also be visualized by warping the glyphs, for example, bending them like bananas to illustrate curvature. If we regard the tensor field as the metric of a manifold, that is, a curved geometric space, one natural way to think of this deformation process is to imagine that the glyph is placed *inside* this manifold. However, while the glyph template is defined in a vector space, the manifold is curved and there is in general no way to fit the flat template inside the manifold, while at the same time preserving correct distances in the glyph, if we compare Euclidean distances in the glyph template with geodesic distances in the manifold. The situation is similar to map making, that is, there is no way to map the surface of the Earth to a flat vector space, so that all geodesic distances are mapped to Euclidean distances in the new space.

Using the exp and log maps to deform the glyph, a procedure described in this chapter, is a compromise. These mappings will at least preserve radial distances and radial angles from the center of the glyph template to all other points.

3 Related Work

In 1881, the French cartographer Nicolas Auguste Tissot published ideas on using circles and ellipses to visualize the deformation of map projections. Mapping the Earth to a flat surface is not possible without introducing some kind of angular or area distortion in the process. The Tissot indicatrix, see Fig. 1, is a small circle or ellipse painted in a map projection. It represents the deformation of an infinitely small circle on the Earth after being deformed by the map projection. If the Tissot indicatrix is a perfect circle, and not an ellipse, then the projection is angle preserving (conformal), and if the area of Tissot indicatrices does not change across the map projection, the map

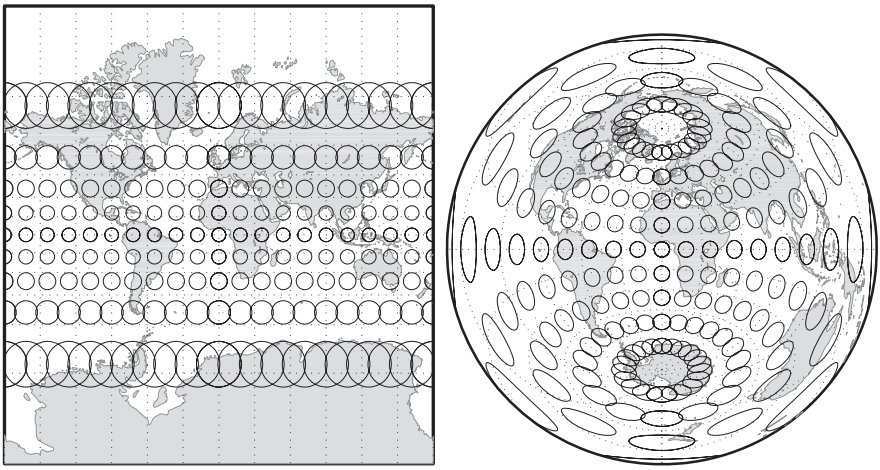


Fig. 1. The area and angle distortion of map projections visualized using Tissot indicatrices. *Left*: The Mercator projection, used in, for example, Google Maps. It is conformal. *Right*: The equidistant azimuthal projection. It is neither conformal nor authalic

projection is area preserving (authalic). A natural extension of the Tissot indicatrix is to use geodesic distances on the Earth to define the circle, in general resulting in a distorted ellipse. For this reason, the geodesic sphere glyph we propose in this chapter, for the visualization of arbitrary metric tensor fields, can be seen as a generalization of the original Tissot indicatrix. In Fig. 2 we show how the geodesic variant of the Tissot indicatrix may be used to visualize the deformation of the metric in a projection of two mathematical surfaces, a half-sphere, and a cone.

Later work in computer graphics has also described methods to visualize the distortion of a projected surface, or manifold in general, from the information contained in a metric tensor field. In spot noise [25], a small image or spot is pasted stochastically in multiple copies over a parametric surface to create different textures. The original paper on spot noise also demonstrates how anisotropic spot noise, in the 2D texture coordinate system of a curved surface embedded in 3D, results in isotropic patterns in object space. This is in fact a way to visualize the metric tensor of the surface. Textures have also been used to visualize vector fields. In line integral convolution (LIC) [6], vector fields are visualized by convolution (integration) of a random texture with streamlines created from the vector field. This yields a low-frequency response along the streamlines. In a method similar to spot noise and LIC, noise is filtered by anisotropic filters steered by second order tensors to visualize the tensor field, see for instance [17] for an early example or [18, 23]. Another example of second order tensor field visualization include the Hyper-LIC [30], an extension of the LIC method where the convolution proceeds not

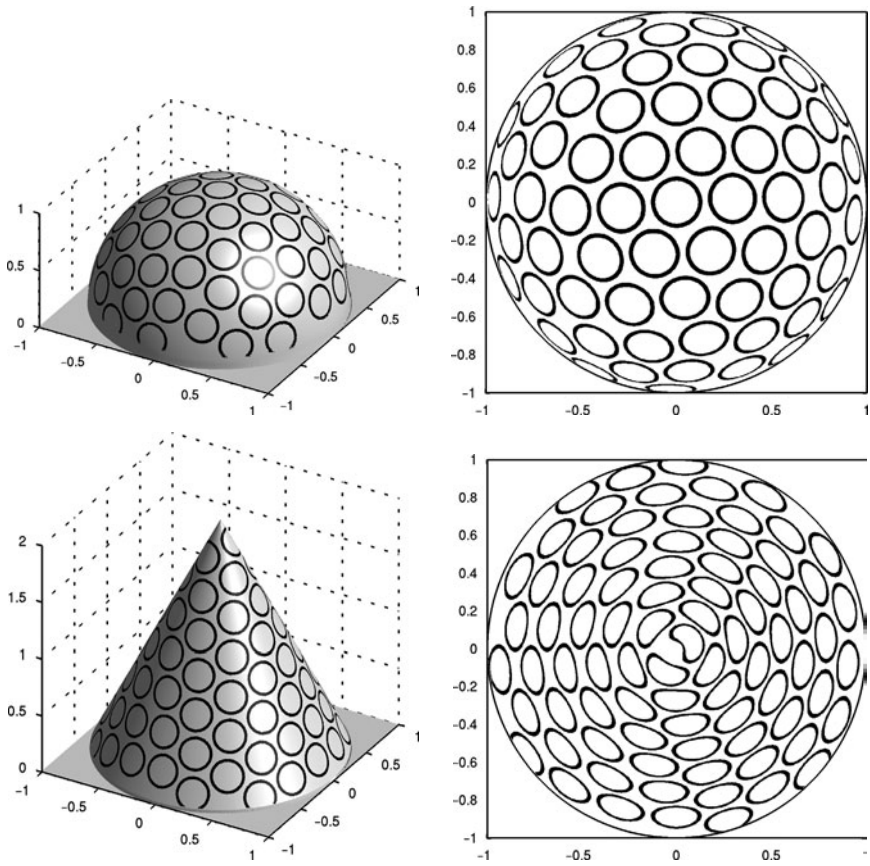


Fig. 2. *Top left*: A half sphere in $\mathbb{R}^3$ painted with geodesic spheres (i.e., circles in 2D). *Top right*: A 2D chart describing the half-sphere, that is, the z-direction has been removed. The same geodesic spheres efficiently visualize the space-variant metric. *Bottom left*: A cone in $\mathbb{R}^3$ painted with geodesic spheres. *Bottom right*: A 2D chart describing the cone, that is, the z-direction has been removed. Note in particular the banana-shaped glyphs in the center and the more ellipse-shaped glyphs close to the perimeter

only along a single streamline, but along a nonlinear patch that is aligned to streamlines derived from both the first and second eigenvectors of the tensor field. This method is somewhat similar to the approach taken in this chapter, since a warped coordinate system is created, which can be used for glyph warping. In [11] an approach is presented based on a physical interpretation of the tensor field that is also able to, in contrast to many other methods, visualize second order tensors with negative eigenvalues. Finally, a procedural generation of textures from tensor fields have been investigated in [15], where reaction-diffusion patterns are steered by a metric tensor field. This

yields a pattern that seems to be composed by separate glyphs, ellipses in 2D, which are adaptively placed, scaled, and deformed by the tensor field. For a successful implementation of this method, however, one has to overcome the numerical problems of simulating a highly nonlinear PDE.

In the medical community, there has been a special need to extract information from tensor fields that goes beyond the visualization of local properties of the field. In "tractography," entire tracts are visualized by performing streamline tracking along the main eigenvector field of a second order tensor field. This procedure, called "fiber tracking," helps radiologists to locate fiber bundles in the human brain and find out about long range white matter fiber connectivity. Fiber tracking shares many similarities with the LIC, Hyper-LIC, and Hyper-streamlines [7], but it is also a research topic in its own right since it is heavily biased by clinical needs and the quest for anatomical understanding of the human brain.

Two properties of spot noise and reaction-diffusion visualization seem to be important for the quality and perception of the tensor visualization. First, both of these methods spread the glyph-like spots in a uniform way according to the tensor field regarded as a metric. The latter of these methods not only scale but also bend the glyph-like structures according to the curvature of the tensor field. In recent work on glyph packing [16] and anisotropic noise sampling [8], also presented in the chapter "Dense Glyph Sampling for Visualization," the first of these behaviors is mimicked and glyphs are placed uniformly over the field. However, the glyphs themselves are still based on the value of the tensor field in each point and do not consider curvature. In this chapter, we present glyphs that do exactly that: they bend, expand, and contract according to the derivative of the tensor field. In combination with a glyph-packing procedure, this technique has the potential to mimic the two most desirable properties of the reaction–diffusion method, in a framework that is numerically stable and relatively fast to compute.

The work presented here is also related to work on texture mapping in computer graphics, in particular the decal compositing with discrete exponential maps [21]. Decal compositing refers to the mapping of small texture maps, decals, onto surface models embedded in $\mathbb{R}^3$. It has been used mainly for artistic purposes and it is defined only for 2D surfaces embedded in 3D. Other methods for the calculation of exponential maps on general manifolds have also been presented. In [22] fast marching is presented as a means to calculate geodesics emanating from a point, that is, indirectly the calculation of exponential maps. In [29] fast methods are presented to calculate all geodesics in a manifold, starting from any point in any direction and traveling any distance. In [5] and [4], the LogMap method is presented as a means of calculating the inverse of the Riemannian exponential map. It is a method that may be particularly well suited for glyph rendering using texture mapping and it is reviewed later in this chapter.

4 Tensors and Index Notation

Tensors generalize scalars, vectors, and matrices to higher dimensions. Sometimes the word "tensor" is used for any multidimensional array with more indices than a matrix, that is, more than two. We use the term in a more precise manner that is in agreement with the notation in physics and differential geometry. In these fields of research, tensors are geometric objects that are invariant under coordinate changes. A vector is a tensor, that is, it is a geometric object that remains the same regardless of the choice of basis or coordinate system that is used to describe it. In physics, the word "tensor" usually refers to what in mathematics would be called a "tensor field" but in both domains it is meaningful to think of tensors as objects defined pointwise in a vector space V.

Many spatial quantities in physics are tensors, for instance; velocity ($\mathrm{m\,s^{-1}}$), diffusion ($\mathrm{m^2\,s^{-1}}$), and electric field strength ($\mathrm{V\,m^{-1}}$). In mathematics and physics, contravariant vectors are those that transform like velocity and position vectors, while the covariant vectors transform like gradients under a change of coordinate system. An example of a higher-order tensor in physics is for instance the stiffness tensor, which is an object with four indices, that is, a mathematical object represented by a 4D array of components. For a general definition, a tensor F is a multilinear map,

$$F : \underbrace{V^* \times \ldots \times V^*}_{r} \times \underbrace{V \times \ldots \times V}_{s} \to \mathbb{R}, \tag{1}$$

that is, a map that is linear in each of its arguments. Its order is $r+s$ and it has type (r, s), meaning that it operates on r covariant vectors and s contravariant vectors. In some contexts, order is called rank and type is called valence, which can be confusing since rank is also used to describe the rank of matrices. Similar to vectors and quadratic forms, the action of tensors can be defined by components that are derived from the action on all combinations of basis vectors $\{\mathbf{w}^i\}$ in the dual space V^* and $\{\mathbf{b}_j\}$ in V,

$$F^{i_1,i_2,\ldots,i_r}_{j_1,j_2,\ldots,j_s} = T(\mathbf{w}^{i_1}, \ldots, \mathbf{w}^{i_r}, \mathbf{b}^{j_1}, \ldots, \mathbf{b}^{j_s}). \tag{2}$$

The number of components is n^{r+s}. If the coordinates are changed, $\tilde{x}^i = \mathrm{T}_k{}^i x^k$, then each contravariant index is transformed as a vector and each covariant index is transformed as a dual vector,

$$\tilde{F}^{i_1,i_2,\ldots,i_r}_{j_1,j_2,\ldots,j_s} = \sum_{a_1\ldots r, b_1\ldots s=1}^{n} F^{a_1,a_2,\ldots,a_r}_{b_1,b_2,\ldots,b_s} \mathbf{T}_{a_1}{}^{i_1} \mathbf{T}_{a_2}{}^{i_2} \ldots \mathbf{T}_{a_r}{}^{i_r} (\mathbf{T}^{-1})^{b_1}{}_{j_1} (\mathbf{T}^{-1})^{b_2}{}_{j_2} \ldots (\mathbf{T}^{-1})^{b_s}{}_{j_s} \tag{3}$$

In physics, this is sometimes how tensors are defined, that is, as objects that transform according to certain transformation laws.

In the above definition, we have used index notation, which is commonly used in differential geometry to denote tensors and differentiate between covariant (lowered) and contravariant (raised) indices. For an introduction to tensors and tensor notation, see for instance [12, 27]. To make the interpretation accessible to a broader audience, we will not use the customary Einstein summation convention, meaning that all sums will instead be written out explicitly. In index notation a (contravariant) vector is identified with its coordinates, meaning that a vector $\mathbf{v}$ in Euclidean space $\mathbb{R}^n$ is written using its coordinates v^i in some basis,

$$\mathbf{v} = v^i = \sum_{i=1}^{n} v^i \mathbf{b}_i. \tag{4}$$

Note in particular that the basis vectors have been dropped and are assumed implicitly in the short form v^i. The index variable i is an integer in the range $1 \ldots n$ and it is typeset in superscript to indicate that this index, and this vector, is contravariant. To further increase readability we also write equations in ordinary linear algebra notation when possible, that is, bold face lower case letters for both contravariant and covariant vectors $(\mathbf{v}, \mathbf{x}, \ldots)$ and upper case bold letters for matrices $(\mathbf{A}, \mathbf{G}, \ldots)$. In some expressions, we use $\dot{x}^i$ and $\ddot{x}^i$ to denote first- and second-order time derivatives.

In addition to vectors, we consider higher-order tensors in this chapter, in particular the metric tensor. The metric tensor is a mathematical object, defining the scalar product between (contravariant) vectors, which in turn can be used to measure important properties in space such as lengths, angles, and area. In vector algebra the scalar product is often implicitly defined simply by

$$\langle \mathbf{v}, \mathbf{u} \rangle = \mathbf{v}^T \mathbf{u} = \sum_{i=1}^{n} v^i u^i, \tag{5}$$

but in general any symmetric positive definite $n \times n$-matrix $\mathbf{G}$ can be used to define a metric,

$$\langle \mathbf{v}, \mathbf{u} \rangle_G = \mathbf{v}^T \mathbf{G} \mathbf{u} = \sum_{i=1}^{n} \sum_{j=1}^{n} v^i g_{ij} u^j. \tag{6}$$

The latter also introduces the commonly used tensor notation for the metric, that is, lowercase with indices written in subscript g_{ij}. In index notation, upper- and lower case letters have less meaning and to comply with standard notation in both linear algebra and differential geometry, we denote the metric by either g_{ij} or $\mathbf{G}$. Subscript indices indicate that the metric is a covariant tensor. In tensor algebra it is natural to pair contravariant indices with covariant ditto, and so the previous expression in (5) for a scalar product is somewhat odd. Instead, it is better to write out the metric explicitly,

$$\langle \mathbf{v}, \mathbf{u} \rangle = \mathbf{v}^T \mathbf{u} = \sum_{i=1}^{n} v^i \delta_{ij} u^j, \tag{7}$$

where δ_{ij} is the Kronecker delta symbol, defined as $\delta_{ij} = 1$ when $i = j$ and 0 elsewhere. It can be regarded as the unit-metric. The number of contravariant (upper) and covariant (lower) indices now match, meaning that the result of the calculation is a scalar (no index).

In summary, the index notation is a handy way to denote vectors and matrices, which easily extends to higher dimensions by adding more indices. At a first glance, the common notation for vectors and matrices may seem more intuitive and easy to use, but there are at least four reasons for sometimes using the index notation: First, index notation extends naturally to higher-order tensors, that is, objects with three or more indices. Second, index notation can differentiate between covariance and contravariance by the use of upper- and lower indices. Third, index notation is particularly efficient when used in combination with the Einstein summation convention, meaning that the summation symbol $\sum_{i=1}^{n}$ is omitted from all expressions and instead it is assumed that indices i, j, etc. appearing more than one time in an expression is summed over, from $1 \ldots n$. In this notation the above scalar product is simply

$$\langle \mathbf{v}, \mathbf{u} \rangle_g = v^i g_{ij} u^j = g_{ij} v^i u^j = g_{ij} u^j v^i. \tag{8}$$

Finally, an advantage with the index notation is that the ordering of the tensors, as factors in a product, becomes irrelevant, in contrast to the usual notation of matrices and vectors in linear algebra.

5 The Metric and Metric Spheres

We now take a closer look at the metric, or metric tensor, and see how it can be visualized. We also introduce a particular orthonormal (ON) coordinate system.

The metric encodes how to measure lengths, angles, and area in a particular point on the manifold by specifying the scalar product between tangent vectors in this particular point. A natural way to visualize the metric is to visualize a "unit sphere," that is, a sphere with radius equal to 1. By "natural" we do not necessarily mean the most suitable way to visualize a metric from a human perception point of view, but rather a straightforward way to visualize the metric using simple mathematics. In Euclidean space the unit sphere is the set of points, $\mathbf{x} \in \mathbb{R}^n$, satisfying $||\mathbf{x}|| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = 1$. In tensor notation and with an arbitrary metric g_{ij} this translates to

$$\sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} x^i x^j = 1. \tag{9}$$

While the metric $g_{ij} = \mathbf{G}$ may be interpreted as a symmetric positive definite matrix, it can be spectrally decomposed,

$$\mathbf{G} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^*, \tag{10}$$

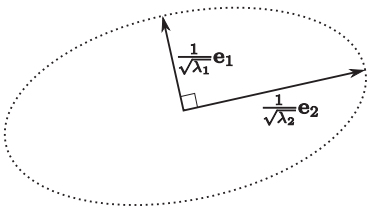


Fig. 3. Coordinate basis vectors in $\mathbb{R}^2$ derived from some metric g_{ij}. This coordinate basis is orthonormal in g_{ij} and orthogonal in the default metric δ_{ij} of the chosen coordinate system, that is, the coordinate system of the observer

where $\mathbf{U}$ is a unitary matrix, $\mathbf{U}\mathbf{U}^* = \mathbf{I}$, and $\mathbf{\Lambda}$ is a diagonal matrix with the eigenvalues of $\mathbf{G}$ ordered in descending order, $\mathbf{\Lambda}_{ii} = \lambda_i$. The eigenvectors to $\mathbf{G}$, found in the columns of $\mathbf{U}$, form an orthogonal basis in $\mathbb{R}^n$ for both the unit metric δ_{ij} and the arbitrary metric g_{ij}. For instance, in $\mathbb{R}^2$ the first eigenvector, corresponding to the eigenvalue λ_1, point along the major axis and the last eigenvector, corresponding to λ_2, point along the minor axis of the ellipse-shaped metric sphere (Fig. 3). In the general case, $\mathbb{R}^n$, the metric sphere will be a hyper-ellipsoid. Using this knowledge we may design a special coordinate system, which is aligned with the axes of the hyper-ellipsoid. If $\mathbf{U} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ and coordinates are denoted by c^i, a vector $\mathbf{v} \in \mathbb{R}^n$ is decomposed by

$$\mathbf{v} = v^i = \frac{1}{\sqrt{\lambda_1}}\mathbf{e}_1 c^1 + \frac{1}{\sqrt{\lambda_2}}\mathbf{e}_2 c^2 + \ldots + \frac{1}{\sqrt{\lambda_n}}\mathbf{e}_n c^n. \tag{11}$$

This coordinate system has many advantages, in $\mathbb{R}^2$ for instance we may now easily parameterize the surface of the metric sphere by painting an isotropic sphere in the c^i coordinates, $c^1 = \cos(t)$ and $c^2 = \sin(t)$, $0 \leq t < 2\pi$ (Fig. 3). An alternative approach to visualize the metric, and emphasize the direction on the eigenvectors, is to paint a unit box, $c^i : \max(c^1, c^2) = 1$. In fact, we may paint any tensor glyph in this coordinate system, for instance superquadric tensor glyphs [17] or even the "space ship" glyph in [28].

We call the map from this coordinate system to the vector space E, $E : \mathbb{R}^n \rightarrow V$. It is an isomorphism from the Euclidean space $\mathbb{R}^n$ (and the unit metric) to a new vector space V equipped with the metric $\mathbf{G} = g_{ij}$. Of many such isomorphisms, it has the special property that it is aligned with the axes of the hyper ellipsoid describing g_{ij} in V, in a particular basis.

6 The Geodesic Equation and Geodesic Spheres

In most applications where metric tensor fields are visualized, the metric is not constant but changes from point to point. A natural theory for space-variant metrics is the non-Euclidean geometry of Riemannian manifolds, a fact that has been pointed out by several authors, see for instance [19]. In

Riemannian geometry, the distance between two points in space is defined by the length of the shortest curve between them, where the length of this curve is obtained from the integral over the tangent vectors to a curve, measured using a space-variant metric $g_{ij}(x)$,

$$d(a,b) = \min_{\gamma:\gamma(0)=a,\gamma(1)=b} \int_0^1 \sqrt{\dot{\gamma}(t)^i g_{ij}(\gamma(t))\dot{\gamma}(t)^j}\, dt. \tag{12}$$

Similar to the case of a constant metric, we may now define geodesic spheres in this Riemannian manifold. For a sphere centered in a point p in the manifold, the following relation hold for points x on the geodesic sphere,

$$d(p,x) = 1. \tag{13}$$

The problem with this metric, from an application point of view, is that the space-variant metric makes it more difficult to evaluate the distance between two points since the minimization is performed over an infinite set of curves γ. One way to approach this problem is to derive a parametric function for points on the sphere, without measuring distances explicitly. Using the geodesic equation, defined in (14) below, geodesics emanating from a point p starting off in a specific direction and traveling a specific distance (in this case 1) may be generated. These solutions correspond to paths of free particles moving in the manifold, without any forces acting on them, and in this sense they generalize the notion of straight lines in Euclidean geometry. Without going into details, geodesics can be described and calculated using the geodesic equation. It is a second order ODE, which expresses that the second derivative of the position, the acceleration, is zero. Because of the space variant metric, a special term involving the Christoffel symbol ${\Gamma^i}_{jk}$ needs to be included,

$$\frac{d^2x^i}{dt^2} + \sum_{j=1}^{n}\sum_{k=1}^{n} {\Gamma^i}_{jk} \frac{dx^j}{dt}\frac{dx^k}{dt} = 0, \tag{14}$$

where $1 \leq i,j,k \leq n$. The Christoffel symbol is actually a function of the coordinate system, that is, ${\Gamma^i}_{jk}(x^i)$. It is not a tensor in a strict sense, it does not transform as a tensor when the coordinate system is changed, but it benefits greatly from the index notation since it has three indices. It is derived from the metric tensor,

$${\Gamma^i}_{jk} = \frac{1}{2}\sum_{m=1}^{n} g^{im}\left(\frac{\partial g_{mj}}{\partial x^k} + \frac{\partial g_{mk}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^m}\right), \tag{15}$$

where g^{ij} is the inverse of the metric g_{ij}, that is, $g^{ij} = \mathbf{G}^{-1}$. A geodesic starting at $\gamma(0) = p$, where p is a point on the manifold, with a velocity $\dot{\gamma}(0) = v^i$ will have a geodesic length $||v^i|| = \sqrt{\sum_{i,j} v^i g_{ij} v^j}$ at $t = 1$, and thus $d(p,\gamma(1)) = ||v^i||$. In this way, by following geodesics starting at p with different unit speed tangent vectors, we obtain a polar representation of a geodesic sphere. We return to how this is solved in practice in a later section dealing specifically with the implementation of this.

7 The Exponential Map and Riemannian Normal Coordinates

With the introduction of geodesic distance and geodesics, we now have a way to paint geodesic spheres to visualize some of the characteristics of a space-variant metric tensor field. However, we have not yet introduced a coordinate system similar to the coordinates c^i introduced for a constant metric. A first step toward the introduction of such a coordinate system is to define the Riemannian exponential map, known from differential geometry. This map, and its inverse the Riemannian logarithm, is depicted in Fig. 4.

Let T_pM denote the tangent space to a manifold M at a point $p \in M$. In our space-variant metric, this is simply the space of all tangent vectors of curves through a point p, which is a vector space. In particular, this is the space of all possible tangent vectors to geodesics emanating from p. The map $\exp_p : T_pM \to M$ is defined by

$$\exp_p(v^i) = \gamma(1), \tag{16}$$

where γ is the geodesic for which $\gamma(0) = p$ and $\dot{\gamma}(0) = v^i$. It is appropriate to use a "shooting" analogy here, $\exp_p(v^i)$ is where a particle ends up after one time unit, if it is shot from a point p with velocity v^i.

The introduction of the exponential map can be done without any reference to coordinates in a specific basis, it is simply a map from vectors v^i seen as geometric objects in the tangent vector space of a point p, T_pM, to other points in the manifold. By choosing an ON coordinate system for T_pM, we obtain what is called Riemannian normal coordinates (RNC), Geodesic normal coordinates, or normal coordinates for short. In cartography, where M is the spherical surface of the Earth, this coordinate system is called equidistant azimuthal projection, which is depicted in Fig. 1. In the general case, this ON basis can be seen as an isomorphism $E : \mathbb{R}^n \to T_pM$. Joining it with the exponential map, we have a map from $\mathbb{R}^n \to M$, and the inverse of this map

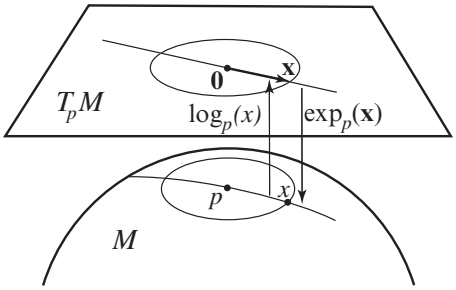


Fig. 4. A schematic view of the $\exp_p$ and $\log_p$ maps, connecting vectors **x** (boldface) in the tangent plane, T_pM, with points x (italic) in the manifold M

gives us the coordinate of a point q on the manifold by $\varphi(q) = E^{-1} \exp_p^{-1}(q)$, which is a well defined inverse in a neighborhood U around p.

8 Solving the Geodesic Equation

Before we actually use the geodesic equation to paint glyphs, we briefly touch upon how to solve it, both accurately using ODE solvers and approximately using a Taylor approximation. Like any second- or higher-order ODE, it can be reformulated by a system of first order ODEs, $\frac{\partial s}{\partial t} = f(s,t)$, for a vector valued state s. The two variables x^i and $\dot{x}^i$ evolve in time according to

$$\begin{bmatrix} \frac{\partial x^i}{\partial t} \\ \frac{\partial \dot{x}^i}{\partial t} \end{bmatrix} = \begin{bmatrix} \dot{x}^i \\ -\sum_{j=1}^n \sum_{k=1}^n \Gamma^i{}_{jk} \dot{x}^j \dot{x}^k \end{bmatrix}, \tag{17}$$

where $\Gamma^i{}_{jk}$ is spatially varying depending on x^i. The right hand side is independent of t, that is, it is an autonomous ODE. Given that initial conditions are known, for example, $x(0) = p$ and $\dot{x}(0) = v^i$, this system of ODEs has exactly one unique solution under very general conditions according to the Picard–Lindelöf theorem. We identify a particular solution with a geodesic curve $x^i(t) = \gamma(t)$. While the Christoffel symbol might be difficult to comprehend at first, it is worth noting that the contribution by $\Gamma^i{}_{jk}$ is symmetric with respect to a flip of sign in $\dot{x}^i$.

Implementation of a numerical solution to this ODE in for example, MATLAB is straightforward using standard ODE solvers, such as `ode45`, based on the Runge–Kutta (4,5) formula, which was used for the experiments in this chapter. The only reservation is that even a third order tensor-like object, like the Christoffel symbol, generates a notation that is quite involved when implemented in a vector- and matrix-oriented language like MATLAB.

It is important to use a proper interpolation scheme in the calculation of the derivatives of g_{ij}, if the tensor field is known only in a discrete set of samples as in DT-MRI. We found that bilinear or cubic interpolation gave a tensor field that was smooth enough to use as input to the ODE and to compute Taylor approximations to the geodesic equation. To ensure positive definiteness we computed the interpolation in the Log-Euclidean domain [1], which can be seen as an approximation to an affine invariant interpolation, see, for example, [2] for a complete explanation. It is, however, important to point out that the proper choice of interpolation method highly depends on the application, and the glyph warping framework we present is applied after the user has decided upon the right choice of interpolation. In some applications a continuous tensor field will be known in every point. In other applications, the user may choose a model where the tensor field is constant inside each pixel or voxel. In the latter case, the solution to the geodesic equation must be generalized to noncontinuous tensor fields, which analogous

to optics and acoustics corresponds to a generalized law of refraction on the boundary between two anisotropic media on the edge of every pixel or voxel.

For many applications in computer graphics, speed and ease of implementation is an issue. For this reason, we also derive Taylor approximations of the exponential map. Directly from the geodesic equation, we have the second-order derivative of our geodesic curve. Given the initial value of the position and derivative, $x(0)$ and $\dot{x}(0)$, we have everything needed to make a second-order Taylor approximation of a geodesic, valid for small values of t:

$$\tilde{x}^i(t) = x^i(0) + \dot{x}^i(0)t - \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} \Gamma^i{}_{jk}\dot{x}(0)^j\dot{x}(0)^k t^2 + O(||\dot{x}(0)^i t||^3), \quad (18)$$

which for $t = 1$, according to (16), yields for a coordinate system in which $p^i = 0$,

$$\exp_p(v^i) = 0 + v^i - \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} \Gamma^i{}_{jk}v^j v^k + O(||v^i||^3). \quad (19)$$

This approximation will only be valid around a small neighborhood to p. As of today, it is not entirely clear how good this approximation is and more research is needed to find bounds on the approximation error and perhaps also derive higher-order Taylor approximations for geodesics. As will be shown in the experimental section, this approximation is, however, good enough to be useful for slowly varying tensor fields.

9 Geodesic Spheres and Warped Coordinate Systems

Using (16), (17) or the approximation (19), we are able to explicitly map unit vectors in T_pM to coordinates on the manifold and thereby paint unit spheres. By choosing the special coordinate system derived in Sect. 5, c^i, in combination with these formulas, we may also navigate on the manifold using a Riemannian normal coordinate system that is aligned with the major and minor axes of the metric ellipse or ellipsoid. This allows us to map not only spheres, but in fact any glyph that is naturally defined in the ellipse- or ellipsoid aligned coordinate system. Later in this chapter, we demonstrate this by mapping the aligned unit box by using RNC. This procedure will result in a box glyph with approximately unit length sides, which has its major axis along the main eigenvector of the local metric, but on a larger scale has its shape deformed according to geodesics emanating from its center point. We refer to this process as glyph warping.

10 The Logarithmic Map

The function $\log_p(q)$ is a function that maps points q on the manifold to the tangent space in p, T_pM, and it is the inverse of $\exp_p(v^i)$. While the exponential function is fairly easy to calculate numerically by solving a second

order ODE, the estimation of the $\log_p(q)$ mapping has attracted less attention in the literature, perhaps due to the lack of fast and accurate solutions. From the Taylor approximation in (19), it is, however, straightforward to derive the second-order Taylor approximation for this inverse,

$$\log_p(q^i) = 0 + q^i + \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} \Gamma^i{}_{jk} q^j q^k + O(||q^i||^3). \tag{20}$$

In our experience this approximation is less stable than the Taylor approximation of $\exp_p(v^i)$ in (19), that is, it is only valid in a very small neighborhood around p, and for this reason, we have not used this second-order Taylor approximation of this mapping in our experiments.

A recently proposed method to calculate the $\log_p(q)$ map is the LogMap method [4, 5]. One way to explain this method is to study how the intrinsic mean is computed [9, 13]. Let $\{x_i\}$ be N data points in a manifold M and seek the minimizer to the function

$$f(p) = \frac{1}{2N}\sum_{i=1}^{N} d^2(p, x_i), \tag{21}$$

where $d^2(p, x_i)$ is the squared geodesic distance between points p and x_i. Then, the gradient of f is [13, 20]

$$\nabla f(p) = -g_{st}\frac{1}{N}\sum_{i=1}^{N} \log_p x_i. \tag{22}$$

Setting $N = 1$ and $x_1 = x$ gives the following formula for $\log_p$,

$$\log_p(x) = -g^{st}\frac{1}{2}\nabla_y d^2(y, x)\Big|_{y=p}. \tag{23}$$

The metric g_{st} and the inverse metric $g^{st} = (g^{-1})^{st}$ have been added here to handle the general case, but choosing an ON-basis for T_pM yield $g_{st} = g^{st} = \delta_{st}$ and allow us to identify co- and contravariant vectors. With the formula above, estimating $\log_p(q)$ becomes a matter of estimating geodesic distances on M. If distances $d(x, y)$ are known for all $x \in N(p)$, where $N(p)$ is some small neighborhood of p, and for all $y \in M$, then the gradient of the squared distance function can be easily estimated numerically by fitting a second-order polynomial, which is then differentiated analytically. (The reason for using the squared distance function is simply that it is much easier to approximate using a finite set of basis functions, compared to the plain distance function, which has a discontinuity in its origin.) Distance functions in turn can be estimated numerically for manifolds by solving the eikonal equation, for example, by using level-set methods for front propagation [19, 22], or ordered upwind methods such as fast marching [22, 24] and the Dijkstra

algorithm [3]. In some special cases such as the sphere, the cone, and the Poisson disk model of the hyperbolic plane, the distance function can also be derived analytically.

In this chapter, we focus mainly on the $\exp_p(v^i)$ map, since it is the most convenient mapping to use if one has a glyph that is described by a set of connected vertices. We note, however, that if the glyph is given by a texture, the LogMap method might be convenient since it yields a mapping from points q on the manifold directly to texture coordinates v^i. It also has the computational advantage that it calculates the mapping for all points in the glyph in one step, given only a few localized distance functions from points around p.

11 Experiments

In this section, we first describe some experiments performed on a simulated synthetic two-dimensional DT-MRI dataset, where Rician noise and partial volume effects have been introduced using realistic parameter settings representative for common clinical protocols. This dataset consists of a 2D tensor field with 2×2 symmetric positive definite tensors. We have chosen a 2D dataset because it demonstrates several features of glyph warping and yet it is easy to visualize in print. Glyph warping using exponential maps is, however, not restricted to 2D, but works in any dimensions. In Fig. 5, we show a close-up of the tensor field displayed using three variants of sphere-glyphs. The first variant is the metric sphere, which may be seen as a first-order approximation to the geodesic equations. The second and third image shows the second-order approximation and the Runge–Kutta (R-K) solution to the geodesic ODE. In Fig. 6, we demonstrate the effect on a global scale, once again we use the sphere-glyph. The difference is subtler now, but experts in tensor image processing still agree that the two rightmost images have a softer appearance. In a third experiment, see Fig. 7, we once again tried the three variants of glyph warping, but this time we used the box glyph instead of an ellipse. We note that both curvature and changes in tensor shape may be seen in the two rightmost visualizations. Again there is little difference between the second-order Taylor approximation and the R-K solution. Compared to the sphere glyph, the box glyph contains straight lines and sharp corners, which appears to be the main reason why it is easier to see the effect of the nonlinear mapping. In a fourth experiment, see Fig. 8, we tried glyph warping on somewhat more exotic glyphs. In the image to the left, we have used texture maps of soda cans as tensor glyphs. In the middle image, we used a single template inspired by superquadric glyphs. In the right image we have used glyph warping on anisotropy-adaptive superquadric glyphs as defined in [17], where isotropic tensors have been assigned round glyphs and anisotropic glyphs have a more box-shaped appearance

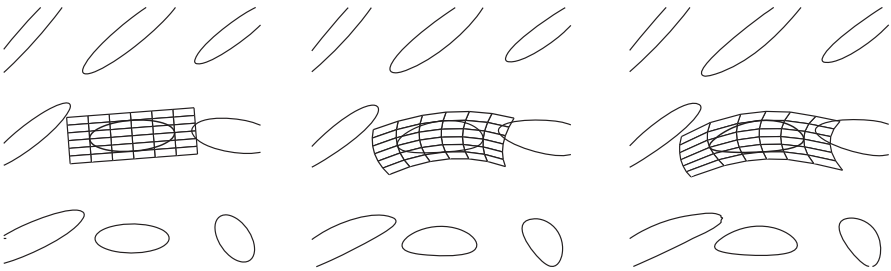


Fig. 5. *Left*: In a first order approximation of the RNC, the unit sphere (circle) is equivalent to the well-known metric ellipse. *Middle*: In a second order approximation of RNC, similar to exact RNC, the unit sphere might be bent. *Right*: Solving the geodesic ODE using R-K results in almost exact RNC. Despite the visible deformation of the Riemannian normal coordinate system attached to the center point, the geodesic sphere glyph looks almost the same in all three examples. For this reason, geodesic spheres may not always be the best choice from a human perceptual point of view. The box glyph overcomes some of these limitations

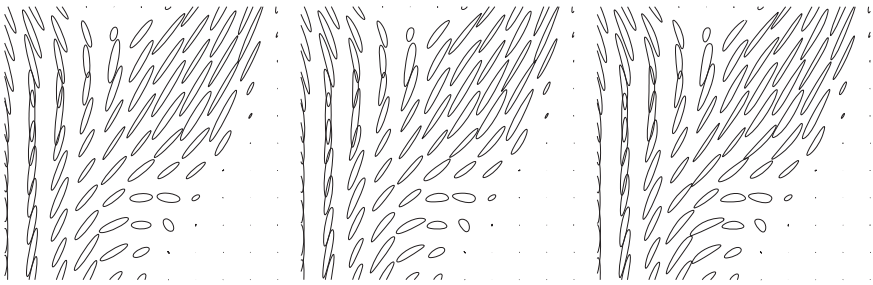


Fig. 6. *Left*: Metric sphere glyphs painted in a first order approximation of the Riemannian normal coordinate system, equivalent to metric ellipses. *Middle*: Metric sphere glyphs painted in a second order approximation of RNC. *Right*: Metric sphere glyphs painted in true RNC found by numerically solving the geodesic equation

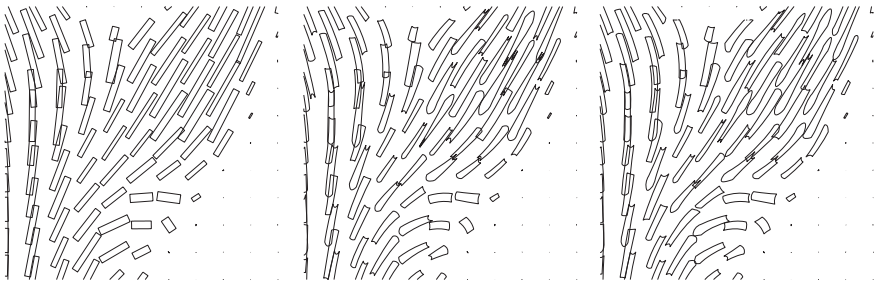


Fig. 7. *Left*: Tensor box glyphs painted using a first order approximation of RNC. *Middle*: Tensor box glyphs painted using a second order approximation of the box glyph. Note that glyphs are not only bent, they also vary in thickness. *Right*: Box glyphs painted in true RNC

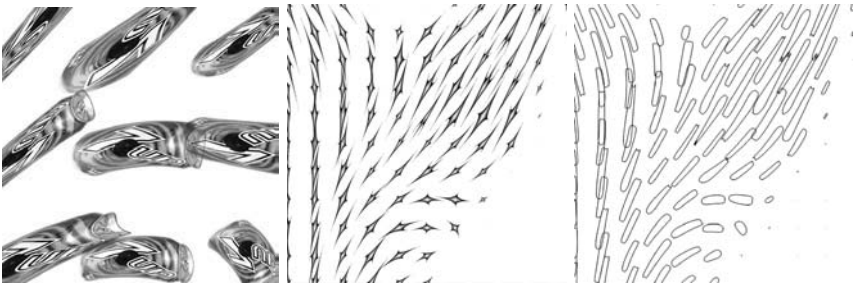


Fig. 8. Warping various glyph templates. *Left*: Glyphs based on a single soda can template. *Middle*: Glyphs derived from a single superquadric-inspired template. *Right*: Glyphs based on anisotropy-adaptive superquadric templates

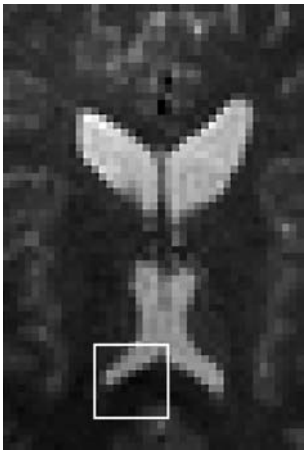


Fig. 9. An axial slice of a human brain, giving an anatomical overview of the experiment. The square marks the area studied in detail in Fig. 10

Glyph warping was also tested on real DT-MRI data, acquired at CMIV. In Fig. 9, an anatomical overview is given of the experiment presented in detail in Fig. 10. In this experiment first- and second-order Taylor approximations are compared to the more exact R-K solution, and anisotropy-adaptive superquadric glyphs were used as templates for the glyph warping. The results are also presented numerically in Table 1. Apparently the most visible effect of the new glyphs is obtained for large tensor glyphs in areas where the shape or orientation change rapidly enough. Numerically, the second-order approximation is always better than the first-order approximation, but the conclusion from a visual inspection is that for large glyphs (that cover several pixels) it is better to solve the ODE using R-K to avoid inaccurate results.

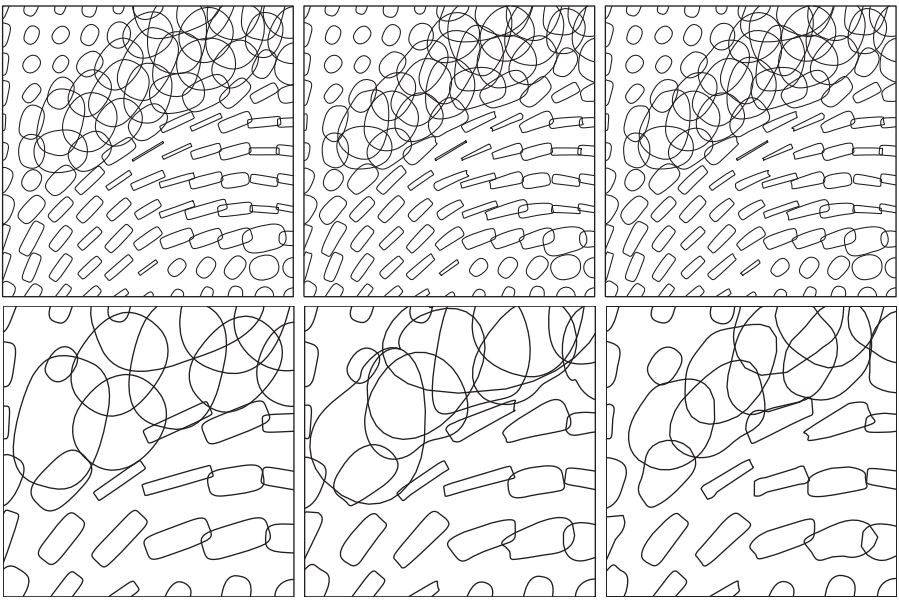


Fig. 10. Tensor glyphs of the planar components of the diffusion tensor field in a human brain, shown in high-resolution (*Top*) and in down-sampled resolution (*Bottom*). *Left*: First order glyph approximation. *Middle*: Second order glyph approximation. *Right*: RNC from solving geodesic ODEs with R-K. Apparently, even the second order approximation introduces visible errors, compared to the true RNC, for the largest glyphs. For small glyphs, the second order approximation is qualitatively similar to the true RNC

Table 1. The standard deviation of the error for the Taylor approximated glyphs in Fig. 10, when the R-K solution is regarded as ground truth. Apparently the second order approximation has the smallest error

	Taylor-1 vs. R-K	Taylor-2 vs. R-K
High-resolution	0.0694	0.0475
Low-resolution	0.1988	0.1738

12 Conclusion

We have presented a framework for visualization of metric tensor fields based on the Riemannian exponential map and its inverse the Riemannian logarithm map. This framework extends some of the previous methods for painting glyphs based on tensor eigen decomposition.

The glyph warping framework is different from other proposed visualizations of tensor fields using glyphs by the fact that a warped glyph is not a

function of the tensor field in a single point, but rather the result from an integration or averaging around this point. The proposed method for warping glyphs works not only in $\mathbb{R}^2$, seen in the experiments, but can be easily generalized to $\mathbb{R}^3$. By changing the glyph or modifying the tensor field, for example, by exponentiation of the tensors, we may obtain visualizations emphasizing different characteristics in the tensor field. Depending on the need for accuracy or speed, one may choose either numerically accurate geodesic warping by solving the ODE using, for example, the Runge–Kutta (4, 5) formula or alternatively, choose the faster version where the bending of the glyphs is calculated using a second-order Taylor approximation of the geodesic.

In summary, the Riemannian exponential map, and its inverse the logarithm map, provides a framework for warping glyphs and visualizing geodesics on a manifold defined by a space-variant metric tensor field in $\mathbb{R}^n$. In particular, it allows for visualization of curvature and derivatives of a tensor field, by bending and deforming glyphs according to the local Riemannian geometry.

Acknowledgments

We thank Magnus Herberthson for valuable discussions on tensors and manifolds and Carl-Fredrik Westin for discussions on the application of these glyphs to Diffusion Tensor MRI data. We are also grateful for the financial support from The Manifold Valued Signal Processing project, Swedish Research Council (Vetenskapsrådet, grant 2004-4721), CMIV (http://www.cmiv.liu.se), the Center for Medical Image Science and Visualization and MOVIII (http://www.moviii.isy.liu.se/), the center for Modeling, Visualization and Information Integration funded by the Swedish Foundation for Strategic Research, SSF.

References

[1] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, 2006.

[2] A. Barmpoutis, B. C. Vemuri, T. M. Shepherd, and J. R. Forder. Tensor splines for interpolation and approximation of dt-mri with applications to segmentation of isolated rat hippocampi. *IEEE Transactions on Medical Imaging*, 26(11):1537–1546, 2007.

[3] M. Bernstein, V. de Silva, J. Langford, and J. Tenenbaum. Graph approximations to geodesics on embedded manifolds. Technical report, Department of Psychology, Stanford University, 2000.

[4] A. Brun. *Manifolds in Image Science and Visualization*. PhD thesis, Linkping University, 2007. Linkping Studies in Science and Technology Dissertations, No. 1157, ISBN 978-91-85715-02-2.

[5] A. Brun, C.-F. Westin, M. Herberthson, and H. Knutsson. Fast manifold learning based on Riemannian normal coordinates. In *Proceedings of the 14th Scandinavian Conference on Image Analysis (SCIA'05)*, Joensuu, Finland, June 2005.

[6] B. Cabral and L. Casey Leedom. Imaging Vector Fields Using Line Integral Convolution. In J. T. Kajiya, editor, *SIGGRAPH93*, CGPACS, pp. 263–270, New York, 1993. ACM Press/ACM SIGGRAPH.

[7] T. Delmarcelle and L. Hesselink. Visualizing second-order tensor fields with hyper streamlines. *IEEE Computer Graphics and Applications*, 13(4):25–33, 1993.

[8] L. Feng, I. Hotz, B. Hamann, and K. Joy. Anisotropic noise samples. *Transactions on Visualization and Computer Graphics*, 2008.

[9] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, 2004.

[10] G. H. Granlund and H. Knutsson. *Signal Processing for Computer Vision.* Kluwer Academic Publishers, 1995. ISBN 0-7923-9530-1.

[11] I. Hotz, L. Feng, H. Hagen, B. Hamann, K. Joy, and B. Jeremic. Physically based methods for tensor field visualization. In *Proceedings of IEEE Visualization 2004*, pp. 123–130, 2004.

[12] C. J. Isham. *Modern Differential Geometry for Physicists (World Scientific Lecture Notes in Physics).* World Scientific Publishing Company, 1989.

[13] H. Karcher. Riemannian center of mass and millifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, 1977.

[17] G. Kindlmann. Superquadric tensor glyphs. In *Proceedings of IEEE TVCG/EG Symposium on Visualization 2004*, pp. 147–154, May 2004.

[15] G. Kindlmann, D. Weinstein, and D. Hart. Strategies for direct volume rendering of diffusion tensor fields. *IEEE Transactions on Visualization and Computer Graphics*, 6(2):124–138, 2000.

[16] G. Kindlmann and C.-F. Westin. Diffusion tensor visualization with glyph packing. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 2006.

[17] H. Knutsson, R. Wilson, and G. H. Granlund. Anisotropic non-stationary image estimation and its applications — Part I: Restoration of noisy images. *IEEE Transactions on Communications*, 31(3):388–397, 1983.

[18] T. McGraw, Baba C. Vemuri, Z. Wang, Yun Chen, M. Rao, and T. Mareci. Line integral convolution for visualization of fiber tract maps from dti. In *MICCAI '02: Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part II*, London, UK. Springer, Berlin Heidelberg New York, pp. 615–622, 2002.

[19] L. O'Donnell, S. Haker, and C.-F. Westin. New approaches to estimation of white matter connectivity in diffusion tensor MRI: Elliptic pdes and geodesics in a tensor-warped space. In *Fifth International Conference on*

Medical Image Computing and Computer-Assisted Intervention (MICCAI'02), Tokyo, Japan, pp. 459–466, 2002.

[20] X. Pennec. Probabilities and statistics on riemannian manifolds: A geometric approach. Research Report 5093, INRIA, January 2004. An extended version will appear in the International Journal of Mathematical Imaging and Vision.

[21] R. Schmidt, C. Grimm, and B. Wyvill. Interactive decal compositing with discrete exponential maps. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, ACM, New York, NY, USA, pp. 605–613, 2006.

[22] J. A. Sethian. *Level Set Methods and Fast Marching Methods*. Cambridge, 2001. ISBN 0 521 64204 3.

[23] A. Sigfridsson, T. Ebbers, E. Heiberg, and L. Wigström. Tensor field visualization using adaptive filtering of noise fields combined with glyph rendering. In *Proceedings of IEEE Visualization 2002*, Boston, Massachusetts, pp. 371–378, 2002.

[24] J. N. Tsitsiklis. Efficient algorithms for globally optimal trajectories. *IEEE Transactions on Automatic Control*, 40(9):1528–1538, 1995.

[25] J. van Wijk. Spot noise: Texture synthesis for data visualization. In *Proceedings of ACM SIGGRAPH 1991*, volume 25, Addison Wesley, Reading, MA, pp. 309–318, 1991.

[26] A. Vilanova, S. Zhang, G. Kindlmann, and D. H. Laidlaw. An introduction to visualization of diffusion tensor imaging and its applications. In *Visualization and Image Processing of Tensor Fields*. Springer, Berlin Heidelberg New York, 2005. In Press.

[27] R. M. Wald. *General Relativity*. University Of Chicago Press, Chicago, 1984.

[28] C-F. Westin. *A Tensor Framework for Multidimensional Signal Processing*. PhD thesis, Linköping University, Sweden, SE-581 83 Linköping, Sweden, 1994. Dissertation No 348, ISBN 91-7871-421-4.

[29] L. Ying and E. J Candès. Fast geodesics computation with the phase flow method. *Journal of computational physics*, 220(1):6–18, 2006.

[30] X. Zheng and A. Pang. Hyperlic. In *Proceedings of IEEE Visualization 20003*, Seattle, Washington, pp. 249–256, 2003.

Interactive Volume Rendering of Diffusion Tensor Data

Mario Hlawitschka[1,5], Gunther H. Weber[2], Alfred Anwander[3], Owen T. Carmichael[4], Bernd Hamann[5], and Gerik Scheuermann[1]

[1] University of Leipzig, Germany
{hlawitschka, scheuermann}@informatik.uni-leipzig.de
[2] Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720-8139, USA
ghweber@lbl.gov
[3] Max Planck Institute for Human Cognitive and Brain Science, Leipzig, Germany
anwander@cbs.mpg.de
[4] Department of Neurology, University of California, Davis, CA 95616-4859, USA
ocarmichael@ucdavis.edu
[5] Institute for Data Analysis and Visualization (IDAV) and Department of Computer Science, University of California at Davis, CA 95616-8562, USA
hamann@cs.ucdavis.edu

Summary. As 3D volumetric images of the human body become an increasingly crucial source of information for the diagnosis and treatment of a broad variety of medical conditions, advanced techniques that allow clinicians to efficiently and clearly visualize volumetric images become increasingly important. Interaction has proven to be a key concept in analysis of medical images because static images of 3D data are prone to artifacts and misunderstanding of depth. Furthermore, fading out clinically irrelevant aspects of the image while preserving contextual anatomical landmarks helps medical doctors to focus on important parts of the images without becoming disoriented. Therefore, we present techniques for multimodal volume rendering of medical data sets with a focus on visualization of diffusion tensor images. The techniques presented allow interactive filtering of information based of importance, directional information, and user-defined areas. By influencing the blending between the data sets, contextual information around the selected structures is preserved.

1 Introduction

As 3D volumetric images of the human body become an increasingly crucial source of information for the diagnosis and treatment of a broad variety of medical conditions, advanced techniques that allow clinicians to efficiently and clearly visualize volumetric images become increasingly important. Interaction has proven to be a key concept in analysis of medical images because

static images of 3D data are prone to artifacts and misunderstanding of depth. Furthermore, fading out clinically irrelevant aspects of the image while preserving contextual anatomical landmarks helps medical doctors to focus on important parts of the images without becoming disoriented. Our goal was to develop a tool that unifies interactive manipulation and context preserving visualization of medical images with a special focus on diffusion tensor imaging (DTI) data.

At each image voxel, DTI provides a 3×3 tensor whose entries represent the 3D statistical properties of water diffusion locally. Water motion that is preferential to specific spatial directions suggests structural organization of the underlying biological tissue; in particular, in the human brain, the naturally occurring diffusion of water in the axon portion of neurons is predominantly anisotropic along the longitudinal direction of the elongated, fiber-like axons [MMM*02]. This property has made DTI an emerging source of information about the structural integrity of axons and axonal connectivity between brain regions, both of which are thought to be disrupted in a broad range of medical disorders, including multiple sclerosis, cerebrovascular disease, and autism [Mos02, FCI*01, JLH*99, BGKM*04, BJB*03].

To date, the predominant visualization techniques for tensor images are based on precalculation of geometric primitives that represent tensor properties. Glyph rendering, for example, converts each tensor into an iconic shape that is elongated to match the water diffusion properties implied by the tensor entries, whereas line rendering reduces the tensor data to a discrete set of contours that trace paths along the presumed locations of axon bundles, that is, along directions of highly anisotropic water diffusion (see Delmarcelle et al. [DH92, BL92].) However, computation of these geometric primitives can be computationally expensive, renderings of them can be prone to visual clutter, and they are difficult to interactively modify to accentuate varying aspects of the data in a session-specific way to meet the goals of specific users. To circumvent the cluttering problem, clustering of streamlines has been introduced, for example, by Enders et al. [ESM*05], but in the same year, a user study by Moberts et al. [MVvW05] revealed that clustering neural pathways is a highly user-dependent topic and the additional computation time cannot be neglected. Others presented techniques to interactively select lines depending on different features, for example, Blaas et al. [BBVP05], who try to maintain the context by using high quality shading of rendered of lines, still these techniques used precomputed lines. Kondratieva et al. [KKW05] reduced the computation time by integrating particle lines on the GPU where a simple Euler approach of order one is used for particle integration. Even on current GPUs the computational power is limited and therefore only a limited number of lines can be rendered, which restricts the number of features that can be visualized.

Volume rendering, in contrast, has the potential to present entire images without reducing them to static, discrete sets of geometric primitives. Recent volume rendering techniques utilizing commodity graphics hardware support

interactive frame rates and provide a high degree of interactive manipulation. However, the question of how to convert a volume of tensor-valued voxels into color and opacity maps for volume rendering is not trivial. Previously, univariate or tri-variate summary measures of the anisotropy of water diffusion at each voxel were computed based on the tensor eigenvalues; the summary measures and direction of the major eigenvector were mapped to alpha channels and color, respectively. Lit Tensors, introduced by Kindlmann [KW99], define another approach that employs shading to emphasize anisotropy. Empirical evidence suggests that physicians become disoriented by direction-encoding schemes that deviate from the traditional approach of assigning red, green, and blue to three cardinal directions (compare Pajevic et al. [PP99]). Therefore, we focus on improving other visualization parameters while keeping this well known color coding. This facilitates fast switching between pseudo-colored section planes and our volume rendering approach. Previous approaches to volume rendering as presented by Vilanova et al. [VZKL06] focus on setting transparency depending on scalar values or modifying the shading using directional information as presented by Kindlmann et al. in [KW99]. Both approaches fail to fade out information that is currently not important to the user; thus, we employ directional parameters in addition to those presented previously to modify the opacity value of the transfer function.

Other approaches to the occlusion problem in volume rendering strive to change the location or size of features in the volume to make them visible to the user, see Correa et al. [CSC06] and Bruckner et al. [BG06]. While these techniques are capable of effectively uncovering important aspects of the data and provide good illustrations of the findings, they may be inappropriate in a clinical setting because the location and size of features in brain images often relate directly to the presence or severity of medical disorders (compare Simon et al. [SDB*05]) and, therefore, should be preserved. In contrast, our method is inspired by Importance-Driven Volume Rendering by Viola et al. [VKG04] in that we improve visibility of important sections of the image while fading out irrelevant parts and preserving voxel location and size. Our approach differs, however, in that it does not require presegmented models and therefore can be applied to raw DTI and structural magnetic resonance (MR) data without requiring a preprocessing step.

While previous visualization techniques for tensor data are based on manipulation of simple scalar value-based alpha functions either using fractional anisotropy (FA) as single scalar value or Westin's barycentric spherical, planar, and linear anisotropy measures a_s, a_p, a_l [WPG*97], we introduce transfer functions that depend on more information of the tensor, especially its eigenvector directions. While local direction information does not have medical importance as, for example, neural fibers are defined as integral curves in a certain volume, in many areas of the brain, neighboring fiber bundles have different directions but their anisotropy values only change marginally. Kindlmann et al. [KTW06] showed that FA ceases can be found after preprocessing

the data set, but these methods cannot be easily transferred to volume rendering. Nevertheless, these areas can be found easily by looking at the local change of direction.

2 Volume Rendering

Simple volume rendering of DTI data is based on mapping the normalized major eigenvector (x,y,z) to a color by setting $RGB = (|x|, |y|, |z|)$ and using FA or the barycentric system of linear, planar, and spherical anisotropy to define 1D or 2D opacity transfer functions [EHK*06, VZKL06]. Kindlmann described another approach for volume rendering of diffusion tensor data. He defined transparency based on FA values and added color and shading depending on other tensor properties such as direction and shape. Being inspired by these basic approaches, our method emphasizes the directional features of the data set by using the default color mapping as it is used in the simple approach and provides additional tools for selecting important features.

The major task for our application is reducing the occlusion problem. We do this by allowing the user to interactively manipulate the transfer function depending on the full tensor information. Previous methods represent the isosurfaces of anisotropy values and do not take into account that medical doctors are mostly interested in fiber bundles. Therefore, it should be possible to select distinct bundles and hide others that occlude important parts of the data set. While anisotropy measures and other tensor invariants have proven to provide a good estimate of white matter boundaries, they do not provide information concerning fiber bundle boundaries. However, in most cases, changes of the directional tensor information provide this information. Therefore, we propose to make use of this directional information in direct volume rendering.

2.1 The Occlusion Problem

Since visualization maps information into two spatial dimensions, occlusion is one of the major problems when handling volumetric data sets. Making surfaces more transparent is one of the basic methods that helps provide additional information. The selection of good occlusion coefficients is a difficult problem and painting too many semitransparent surfaces leads to visual clutter produced by both mixture of color and additional geometry. We propose an interactive selection of regions of interest by considering local fiber direction. In addition to the transfer function, one or more direction vectors are used to select areas of interest – or areas that should not be shown at all.

Although the color value of the RGB color mapping scheme provides directional information, the map is not invertible, that is, even neglecting orientation, directions cannot be recomputed from the color value. In addition to that, interpolation of eigenvalues and eigenvectors directions is not the

same as interpolating tensor values and recomputing eigenvalues and eigenvectors afterwards [ZB02, Kin04]. Therefore, we compute the eigenvalues and the required major eigenvector on the GPU using the non-iterative approach by Hasan et al. [HBPA01]. Different criteria are used to select transparency values including the following:

- Anisotropy values such as fractional anisotropy (FA), relative anisotropy (RA), linear, planar, and spherical anisotropy [WPG*97],
- Local direction of major eigenvector, and
- A magic lens[6] that modifies the local transfer function [BSP*93].

We experimented with several functions for directional filtering and determined that the type of function itself is not critical. While hard drop-offs in transparency should be avoided, it is possible to use linear step functions as well as smooth step functions, either the one provided by GLSL defined as

$$f(x) = \begin{cases} 0 & \text{when } x < 0 \\ 3x^2 - 2x^3 & \text{when } 0 \leq x \leq 1 \\ 1 & \text{when } x > 1 \end{cases} \tag{1}$$

or higher-order ones like

$$f(x) \begin{cases} 0 & \text{when } x < 0 \\ (126 + (-420 + (540 + (-315 + 70x) & \\ \quad *x) * x) * x) * x^5 & \text{when } 0 \leq x \leq 1 \\ 1 & \text{when } x > 1. \end{cases} \tag{2}$$

Filtering out single directions breaks the typical symmetry of DTI data sets that is provided by the standard color coding, which produces unfamiliar looking images and interferes with the common way of comparing both hemispheres of the brain. To maintain the same symmetric properties, as the color coding does, one can compare all directions that have the same color (which are up to eight vectors in general, four because of the independence of eigenvector orientation.) To provide more information about the local direction of fibers, the comparison has to be restricted to one direction on the hemisphere. This can be done by defining the vector in one hemisphere and mirroring it at a plane separating both hemispheres of the brain. Having computed the normalized direction vector v and the normalized eigenvector e in the hemisphere, the alpha value is modified by

$$\alpha \propto f(\|\langle e, v \rangle\|). \tag{3}$$

2.2 Preserving Context

It turns out that removing large parts of the image completely, that is, making it completely transparent, may confuse the user. Therefore, our system allows

[6] "Magic Lens" is a trademark of Xerox Cooperation

one to select an opacity for "hidden" areas making them translucent, which preserves the context by fading out uninteresting parts of the data only a bit while still allowing the user to see more important parts that may lie behind the uninteresting parts.

2.3 Combined Rendering of MRI and DT-MRI Data

The previous section focused on methods used for rendering DTI data. While many neuroscientists are primarily interested in this type of data, scalar-valued MRI images provide a higher resolution and make it easier to navigate in the brain. In addition, Gyral structure of brain gray matter and thin, soft structures like blood vessels can be seen in MRI scans while they do not show up in most DTI data sets. Therefore, a combined visualization of DTI and MRI data is useful, especially for surgical planning. We integrate visualization of conventional MRI and DTI data in our volume rendering approach in two different ways:

- We add it to a single-pass volume rendering and define more complex transfer functions based on both local DTI and MRI data and
- Implement a two-pass volume rendering approach to enforce importance-driven volume rendering.

As we aim to restrict the diffusion tensor color map to the RGB model and mapping, introduction of additional colors would be misleading. Most of the information present in MRI data can be seen in simple grayscale slices of the brain, which is the most common way of displaying them. Therefore, we use volume rendering of the MRI data only to highlight boundaries of objects by using a standard 2D grayscale transfer function for MRI data. Special material attributes and gradient-based shading as done in most volume rendering applications, see Bruckner et al. [BVG05], are applied to improve the 3D perception of the surfaces.

2.4 Implementation

Our implementation is based on a standard GPU-based ray-casting volume renderer with 2D transfer functions and the 3D texture based approach as described, for example, by Engel et al. [EHK*06]. Gradient computation can be done either on the GPU or on the CPU, depending on the available graphics board or by the choice of the user. The transfer functions have been extended to support either 2D transfer functions based on gradient magnitude and value or 2D functions based on two scalar values, for example, MRI signal and FA. To filter the data by direction, we need additional information for the graphics board. Modern graphics cards provide several ways for sending the data. For our approach, we use the light information present in OpenGL as it is global information for the entire data set, which can be easily manipulated

between frames. Furthermore, it provides additional parameters, like the cut-off-angle, and exponent that can be used for parameterization of the smooth step function. In medical imaging, usually, no more than two light sources are used (e.g., one static light and one headlight); therefore, at least five light sources are unused that are more than enough for definition of transfer functions.

While the 2D transfer function is based on a lookup table, we evaluate the tensor transfer function analytically on the GPU as it is not computationally extensive and can be manipulated without the need to exchange textures. The major three steps, computing the stencil, computing the tensor volume rendering, and computing the conventional MRI overlay, are shown in Figs. 1 and 2.

The magic lens selection provides an interactive way to select the weighting of MRI and DTI data in the final image. All computation is done on the GPU. We use the lens to emphasize areas of interest in the DTI image while providing the context of the Gyri using MRI. In contrast to Wang et al. [WZMK05], we do not apply different transfer functions to one data set but we select different transfer functions on different modalities to build a combined image that provides all information needed by the user while the geometry is not distorted. Application of this is shown in Figs. 7–9. By using these simplifications, the lens and the volume rendering of both data sets can

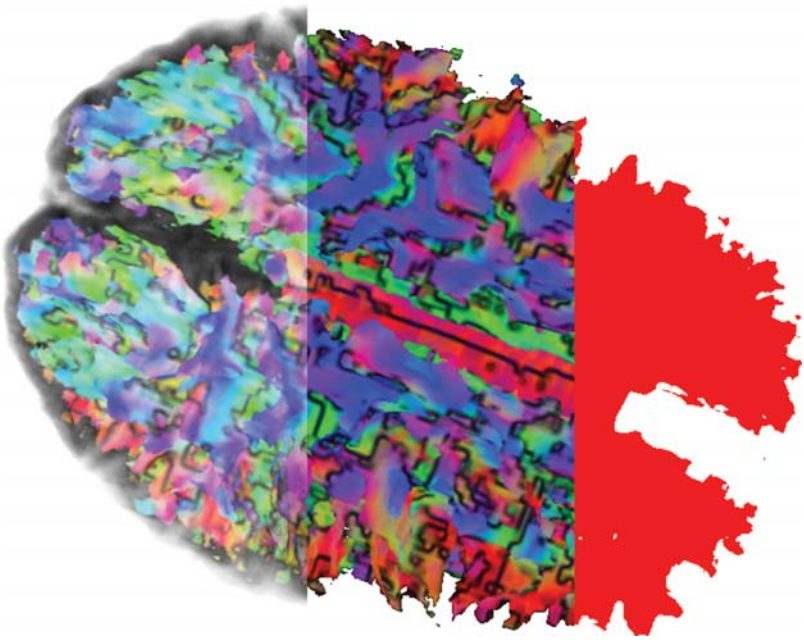


Fig. 1. Three components of the compositing step – from right to left: importance mask, directional-colored DTI data, final compositing with MRI data

Fig. 2. Another view of Fig. 1

be painted in a single rendering pass, which provides interactive frame rates on a full screen application. The speed of the algorithm is independent of the size of the lens. Furthermore, we experienced no change in the frame rate for different sizes of DTI and MRI data sets.

3 Application and Results

We have applied our method to multiple data sets. Our main focus was to determine what users can learn from interactive volume rendering that cannot be seen using pseudo-colored slices or fiber tracking algorithms.

3.1 Quality Analysis for Fiber Tracking in the Gyri

Analysis of connectivity in the human brain is one of the major fields of research related to DTI, where the main interest is in understanding how the brain works by analyzing the connectivity of different areas in the brain. This abstract type of research finds its application in neurosurgery. In tumor surgery, the basic understanding of where neural fibers are situated and which of them are more important than others is one of the most important fields of research having major impact on quality and safety of brain surgeries. Most of the time, neural fibers are visualized by displaying precomputed line structures that imply single lines with defined start and end points, which is

misleading since lines are often started and stopped based on other parameters, for example, by FA. By using volume rendering for DTI data combined with displaying the cortex extracted from MRI data, we can show how far a reasonable tensor line tracking algorithm can proceed into the outer gyri.

An example on this strategy can be seen in Figs. 3 and 4. It can be seen in the transparent area between the outer shell and the visualized DTI data that

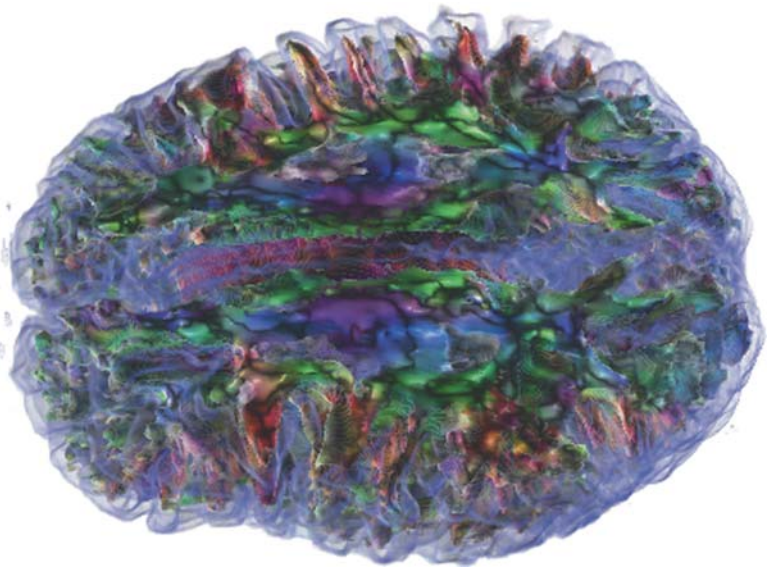


Fig. 3. Overview of gyri clipped by an almost axial clipping plane. (RGB color coding of DTI and clipping of DTI data based on FA value of 0.3 as used in many fiber-tracking algorithms.) A semi-transparent surface was extracted from the original MRI data set visualizing the boundary of the brain

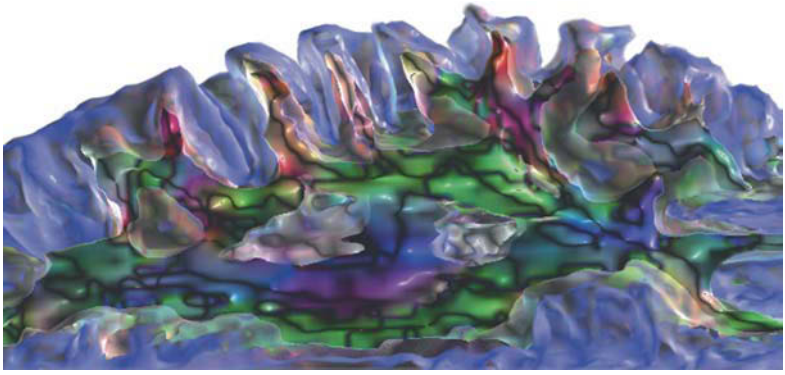


Fig. 4. Close-up view of Fig. 3. The white matter in the gyri can be clearly seen. Fiber tracking without additional information stops at the colored surface. The surface color encodes the direction of fibers directly below the surface

there is an area of about 2–3 mm in size where no tracking can be performed by streamline algorithms due to ambiguity of the tensor information at the interface between gray and white matter. Different transfer functions can be applied to determine which values of FA, apparent diffusion coefficient (ADC), and other tensor parameters can be used as threshold for tractography. Even though there is ambiguity of orientation, the color provides a hint of the behavior of fibers in the gyri. In this case, most fibers would be oriented outwards. The small fibers connecting to the side walls of the gyri cannot be found in the data set using reasonable thresholds.

We believe that it is important to transport that information showing the limits of tractography algorithms to the user, to make him understand the images better. Displaying the limits of algorithms leads to increased understanding and, therefore, better confidence in the information that is actually shown.

3.2 Quality Analysis in the Inner Brain

While simple volume rendering provides good hints on how far fiber tracking can advance into the gyri, we can use our algorithm to filter out parts of the data to see what phenomena are occurring inside the brain. It can be seen in the images shown in Figs. 5 and 6 that the pyramidal tract is "broken." This is due to a resolution problem leading to a strong influence of the corpus callosum and the superior longitudinal fasciculus connecting the frontal and occipito-temporal part of the brain in these voxels.

Tracking fiber bundles in those areas of the data set leads to false results, for example, they follow the wrong bundles in areas where too many fibers pass closely to each other.

It is an open problem whether there is a way to apply volume rendering to higher-order tensor approaches that are usually used to avoid some of these problems. Additional information of the tensor can be used to produce a better angular structure of the tensors that makes it possible to follow different fiber tracts in these regions [TRWW03, HS05].

3.3 Detection and Analysis of Diseases

Many diseases of the human brain, among them Alzheimer's disease and multiple sclerosis, are related to a change of connectivity in the brain. One special disease is the Deletion 22q Syndrome, which is characterized by increased FA and ADC in certain areas of the brain (cf. Simon et al. [SDB*05] and chapter by Cook et al.). These parameters can be easily displayed with our volume rendering approach, providing a context-preserving visualization that helps to easily locate the areas of change in the brain. Despite the fact that one can obtain a good overview of this kind of data sets, 2D transfer functions that base on ADC and FA and statistical values obtained from a larger pool of subjects have to be implemented and will be subject to future research.

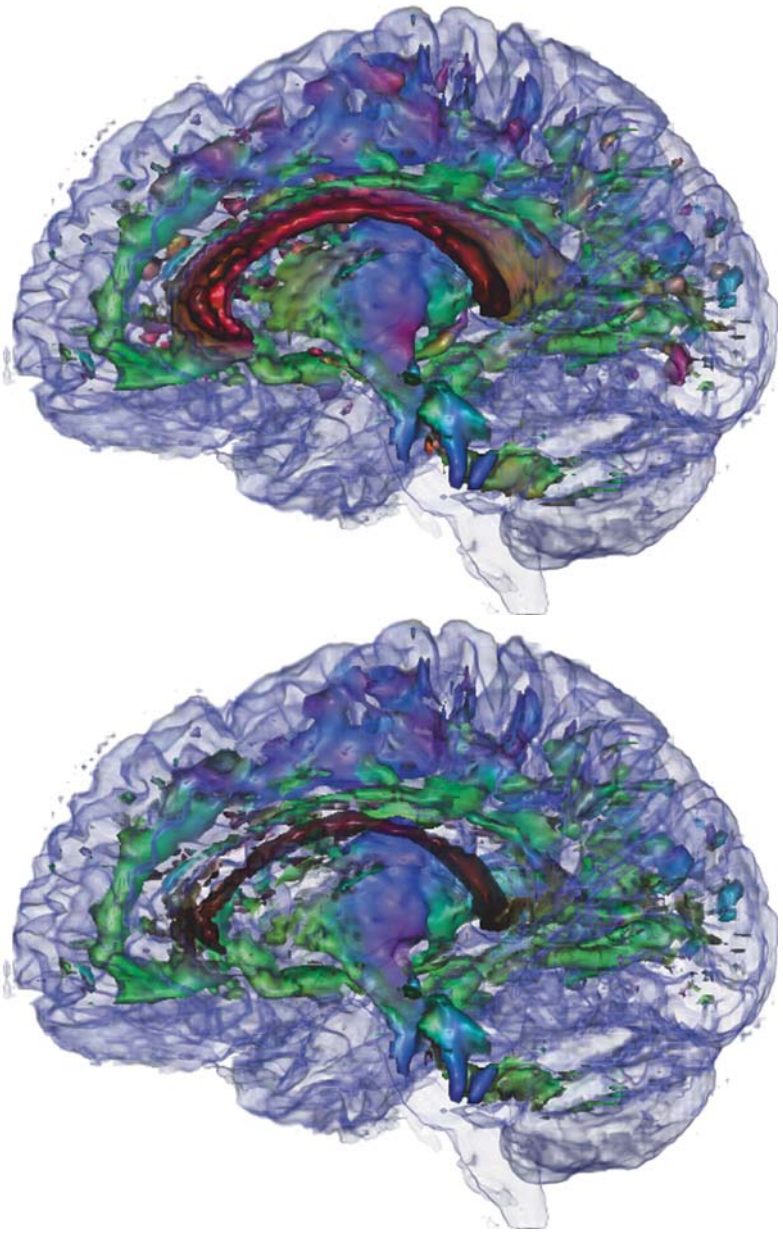


Fig. 5. Original image and x-direction (left–right) removed from DTI data to allow to produce a cleaner view of the singuli and the pyramidal tract. The lasting red component is the corpus callosum as seen in the MRI data and is shown here to provide context

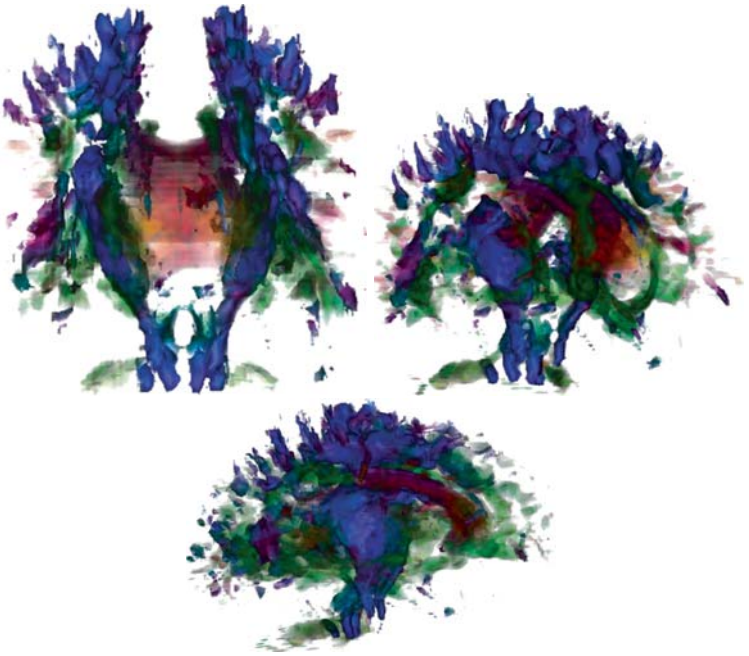


Fig. 6. Pyramidal tract (*blue*) highlighted – other directions are faded out. The pyramidal tract is broken in this data set by the influence of the corpus callosum (*red*) and the green bundle passing by

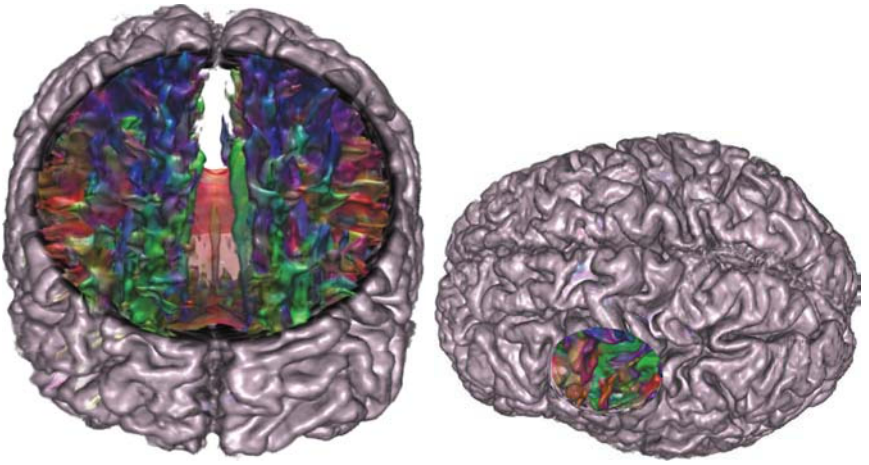


Fig. 7. *Left*: Top/frontal view on the prefrontal lobe of a healthy subject. The cingulum (*green*) and the corpus callosum (*red*) are clearly visible. *Right*: Top left view highlighting the pre- and postcentral sulcus. The underlaying superior longitudinal fasciculus (*green*) is also visible

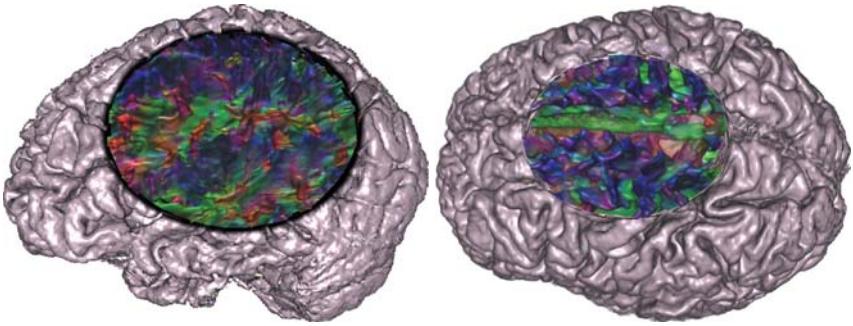


Fig. 8. *Left*: Side view on the parietal and superior temporal lobe showing the superior longitudinal fasciculus (green, middle of the loop) and the inferior fronto-occipital fasciculus (green, lower left of the lens). *Right*: Dorsal view on medial brain regions showing also the cingulum (green) and the corpus callosum with two different shadings of the magic lens. Two different shading models of the lens are shown

4 Conclusions and Future Work

We have described a novel method that allows us to interactively explore complex volumetric DTI data sets. The method has been applied to different data sets, and two major applications have been highlighted where patterns could be identified in less time than using conventional methods. We have demonstrated that all proposed methods can be easily implemented in existing medical visualization systems.

Acknowledgements

We thank the "German Academic Exchange Service" (DAAD) for partially funding this research and for making this collaboration possible (M. Hlawitschka was supported by a DAAD grant.) Furthermore, we want to thank the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany, and Cameron S. Carter, University of California, Davis, Imaging Research Center, for providing the data sets used in this research. We thank the members of the Visualization and Computer Graphics Research Group at the Institute for Data Analysis and Visualization (IDAV) at the University of California, Davis, USA, and the members of the FAnToM group at the University of Leipzig, Germany, and Xavier Tricoche at the University of Utah, Salt Lake City, USA, and Christoph Garth at the University of Kaiserslautern, Germany.

References

[BBVP05] Blaas J., Botha C. P., Vos F. M., Post F. H. Fast and reproducible fiber bundle selection in DTI visualization. In *Proceedings of IEEE Visualization 2005* (Los Alamitos, CA, USA, 2005), Silva C. T., Gröller E., Rushmeier H., (Eds.), IEEE Computer Society, IEEE Computer Society Press, pp. 59–64.

[BG06] Bruckner S., Gröller M. E. Exploded views for volume data. In *Proceedings of IEEE Visualization'06* (Los Alamitos, CA, USA, 2006), Gröller E., Pang A., Silva C. T., Stasko J., van Wikj J., (Eds.), IEEE Computer Society Press, pp. 1077–1084.

[BGKM*04] Barnea-Goraly N., Kwon H., Menon V., Eliez S., Lotspeich L., Reiss A. L. White matter structure in autism: Preliminary evidence from diffusion tensor imaging. *Biological Psychiatry 55* (2004), 323–326.

[BJB*03] Burns J., Job D., Bastin M., Whalley H., Macgillivray T., Johnstone E., Lawrie S. Structural disconnectivity in schizophrenia: a diffusion tensor magnetic resonance imaging study. *British Journal of Psychiatry 182* (2003), 439–443.

[BL92] Basser P. J., LeBihannis D. Fiber orientation mapping in an anisotropic medium with NMR diffusion spectroscopy. *11th Annual Meeting of the SMRM, Berlin* (1992), 1221.

[BSP*93] Bier E. A., Stone M. C., Pier K., William B., DeRose T. D. Toolglass and magic lenses: The see-through interface. In *Proceedings of Siggraph '93* (1993), ACM, pp. 73–80.

[BVG05] Bruckner S., Viola I., Gröller M. E. Volumeshop: Interactive direct volume illustration. In *ACM Siggraph 2005 DVD Proceedings (Technical Sketch)* (2005).

[CSC06] Correa C. D., Silver D., Chen M. Feature aligned volume manipulation for illustration and visualization. In *Proceedings of IEEE Visualization'06* (Los Alamitos, CA, USA, 2006), Gröller E., Pang A., Silva C. T., Stasko J., van Wikj J., (Eds.), IEEE Computer Society Press, pp. 1069–1067.

[DH92] Delmarcelle T., Hesselink L. Visualization of second order tensor fields and matrix data. In *Proceedings of IEEE Visualization 1992* (Los Alamitos, CA, USA, 1992), IEEE Computer Society Press, p. 316.

[EHK*06] Engel K., Hardwiger M., Kniss J. M., Rezk-Salama C., Weiskopf D. *Real-Time Volume Graphics*. A K Peters, Ltd, Wellesley, MA, 2006.

[ESM*05] Enders F., Sauber N., Merhof D., Hastreiter P., Nimsky C., Stamminger M. Visualization of white matter tracts with wrapped streamlines. In *Proceedings of IEEE Visualization*

2005 (Los Alamitos, CA, USA, 2005), Silva C. T., Gröller E., Rushmeier H., (Eds.), IEEE Computer Society, IEEE Computer Society Press, pp. 51–58.

[FCI*01] Filippi M., Cercignani M., Inglese M., Horsfield M., Comi G. Diffusion tensor magnetic resonance imaging in multiple sclerosis. *Neurology 56*, 3 (February 2001), 304–311.

[HBPA01] Hasan K. M., Basser P. J., Parker D. L., Alexander A. L. Analytical computation of the eigenvalues and eigenvectors in DT–MRI. *Journal of Magnetic Resonance 152* (2001), 41–47.

[HS05] Hlawitschka M., Scheuermann G. HOT-lines – tracking lines in higher order tensor fields. In *Proceedings of IEEE Visualization 2005* (Oct. 2005), Silva C. T., Gröller E., Rushmeier H., (Eds.), pp. 27–34.

[JLH*99] Jones D. K., Lythgoe D., Horsfield M. A., Simmons A., Williams S. C. R., Markus H. S. Characterization of white matter damage in ischemic leukoaraiosis with diffusion tensor MRI. *Stroke 30* (1999), 393–397.

[Kin04] Kindlmann G. *Visualization and Analysis of Diffusion Tensor Fields*. PhD thesis, School of Computing, University of Utah, Salt Lake City, UT, USA, 2004.

[KKW05] Kondratieva P., Krüger J., Westermann R. The application of GPU particle tracing to diffusion tensor field visualization. In *Proceedings of IEEE Visualization 2005* (Los Alamitos, CA, USA, 2005), Silva C. T., Gröller E., Rushmeier H., (Eds.), IEEE Computer Society, IEEE Computer Society Press, pp. 73–78.

[KTW06] Kindlmann G., Tricoche X., Westin C.-F. Anisotropy creases delineate white matter structure in diffusion tensor MRI. In *Ninth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'06)* (Copenhagen, Denmark, October 2006), Lecture Notes in Computer Science 4190, pp. 126–133.

[KW99] Kindlmann G., Weinstein D. Hue-balls and lit-tensors for direct volume rendering of diffusion tensor fields. In *VIS '99: Proceedings of the conference on Visualization '99* (Los Alamitos, CA, USA, 1999), IEEE Computer Society Press, pp. 183–189.

[MMM*02] Melhem E. R., Mori S., Mukundan G., Kraut M. A., Pomper M. G., van Zijl P. C. M. Diffusion tensor MR imaging of the brain and white matter tractography. *American Journal of Roentgenology 178*, 1 (January 2002), 3–16.

[Mos02] Moseley M. Diffusion tensor imaging and aging – a review. *NMR In Biomedicine 15* (2002), 553–560.

[MVvW05] Moberts B., Vilanova A., van Wijk J. J. Evaluation of fiber clustering methods for diffusion tensor imaging. In *Proceedings of IEEE Visualization 2005* (Los Alamitos, CA, USA, 2005), Silva

C. T., Gröller E., Rushmeier H., (Eds.), IEEE Computer Society, IEEE Computer Society Press, pp. 65–72.

[PP99] Pajevic S., Pierpaoli C. Color schemes to represent the orientation of anisotropic tissues from diffusion tensor data: Application to white matter fiber tract mapping in the human brain. *Magnetic Resonance in Medicine 42 (3)* (1999), 526–540.

[SDB*05] Simon T. J., Ding L., Bish J. P., McDonald-McGinn D. M., Zackai E. H., Gee J. Volumetric, connective, and morphologic changes in the brains of children with chromosome 22q11.2 deletion syndrome: an integrative study. *NeuroImage 25* (2005), 169–180.

[TRWW03] Tuch D. S., Reese T. G., Wiegell M. R., Wedeen V. J. Diffusion MRI of complex neural architecture. *Neuron 40* (December 2003), 885–895.

[VKG04] Viola I., Kanitsar A., Gröller M. E. Importance-driven volume rendering. In *Proceedings of IEEE Visualization'04* (Los Alamitos, CA, USA, 2004), Rushmeier H., Turk G., van Wijk J. J., (Eds.), IEEE Computer Society Press, pp. 139–145.

[VZKL06] Vilanova A., Zhang S., Kindlmann G., Laidlaw D. An introduction to visualization of diffusion tensor imaging and its applications. In *Visualization and Processing of Tensor Fields* (2006), Weickert J., Hagen H., (Eds.), Springer-Verlag, Berlin Heidelberg, pp. 121–153.

[WPG*97] Westin C.-F., Peled S., Gudbjartsson H., Kikinis R., Jolesz F. A. Geometrical diffusion measures for MRI from tensor basis analysis. In *ISMRM '97* (Vancouver Canada, April 1997), p. 1742.

[WZMK05] Wang L., Zhao Y., Mueller K., Kaufman A. The magic volume lens: An interactive focus+context technique for volume rendering. In *Proceedings of IEEE Visualization 2005* (Los Alamitos, CA, USA, 2005), Silva C. T., Gröller E., Rushmeier H., (Eds.), IEEE Computer Society, IEEE Computer Society Press, pp. 65–72.

[ZB02] Zhukov L., Barr A. H. Oriented tensor reconstruction: Tracing neural pathways from diffusion tensor MRI. In *Proceedings of IEEE Visualization '02* (Los Alamitos, CA, 2002), IEEE Computer Society, pp. 387–394.

Dense Glyph Sampling for Visualization

Louis Feng[1], Ingrid Hotz[2], Bernd Hamann[1], and Kenneth Joy[1]

[1] Institute for Data Analysis and Visualization (IDAV), Department of Computer Science, University of California, Davis, CA 95616-8562, USA
{zfeng,kijoy}@ucdavis.edu, hamann@cs.ucdavis.edu
[2] Zuse Institute Berlin, Division for Visualization and Data Analysis, Berlin Germany
hotz@zib.de

Summary. We present a simple and efficient approach to generate a dense set of anisotropic, spatially varying glyphs over a two-dimensional domain. Such glyph samples are useful for many visualization and graphics applications. The glyphs are embedded in a set of nonoverlapping ellipses whose size and density match a given anisotropic metric. An additional parameter controls the arrangement of the ellipses on lines, which can be favorable for some applications, for example, vector fields and distracting for others. To generate samples with the desired properties, we combine ideas from sampling theory and mesh generation. We start with constructing a first set of nonoverlapping ellipses whose distribution closely matches the underlying metric. This set of samples is used as input for a generalized anisotropic *Lloyd relaxation* to distribute samples more evenly.

1 Introduction

Anisotropic spot samples with certain characteristics, such as spatially varying density and size, have many applications in visualization and computer graphics ranging from glyph rendering and texture generation for visualization purposes [3, 8, 11–14], digital halftoning [15, 19, 25] to mesh generation [2, 16, 22]. While some of the desirable properties of the samples are similar across applications, the goals and appropriate sampling strategies are problem-dependent. When using the samples as input for texture generation as, for example, line integral convolution (LIC), it is important to avoid structural patterns. An ellipse alignment leads to distracting artifacts in the LIC texture. But an alignment of glyphs is desirable for other applications, for example, vector field visualization, where it supports the impression of flow.

To achieve the objectives listed earlier we have designed a method, which generates an anisotropic sample distribution in two main steps. First, we construct a set of nonoverlapping ellipses. This first sample set already exhibits most of the desired properties. Next, we use a generalized anisotropic Lloyd relaxation to distribute the spot samples more evenly. Our anisotropic Lloyd

relaxation is a straight-forward generalization of the isotropic version and is based on the work of Labelle and Shewchuk [16] and Du et al. [2]. A parameter in the Voronoi cell definition controls the alignment of the ellipses. We have applied our method to several test data sets and various vector and tensor fields.

2 Related Work

The generation of point or spot distributions with certain properties is the subject of research in different fields. Dependent on the specific needs many algorithms have been developed.

Generating uniformly distributed points with constant or varying density without large scale patterns has a long tradition in the area of noise generation, sampling, or halftoning. These fields are closely related, many sampling algorithms are directly used to generate noise textures. Some techniques use a form of stochastic sampling, where random points are added or rejected according to certain criterions. Such methods often suffer from low convergence rates. Other approaches use relaxation techniques, in particular Lloyd's relaxation [1, 18] and its variants resulting in high-quality blue-noise samples. To improve efficiency of the sampling algorithm several approaches have been suggested using tile sets, which then are repeatedly tiled across the plane. Using this strategy, for example, Ostromoukhov et al. introduced a very efficient isotropic blue-noise sampling method based on Penrose tiling [20]. Most of these methods assume that the samples are isotropic. For a survey on sampling techniques, we refer the reader to [5, 24]. Anisotropic settings can be found in the area of stippling or automatic mosaic generation, where objects of different size and shape are distributed on a plane [4, 6, 7]. Different from our definition, here the orientation of the distributed objects is not predefined by the metric but can change during relaxation. Most of the proposed methods use a Lloyd relaxation based on a generalized Voronoi cell definition, where the Euclidean distance of the objects is approximated.

The goal of generating an anisotropic distribution following a given metric also appears in the area of mesh generation. Shimada et al. [22] introduced a mesh generation approach using a close packing of ellipsoidal bubbles. The packing is performed using a particle system, where particles move according to repulsive and attractive forces. The equations of motion are solved numerically to yield a force-balancing configuration. A geometric approach for anisotropic mesh generation was chosen by Du et al. [2] and Labelle et al. [16]. Both methods define a generalized Voronoi tessellation based on a non-Euclidean metric using different distance approximations as basis for the final triangulation. Our work builds on the ideas introduced in these methods.

The use of glyphs for visualization of local field properties is common in visualization. The question of placing these glyphs has been subject of discussion in several contexts. The most common strategies are regular sampling, random sampling with or without Poisson property [13, 17], or procedural texture

generation, for example, using reaction diffusion. In vector field visualization, Turk and Banks proposed a method to place arrows along streamlines generated by streamline optimization [23]. Kindlmann introduced reaction-diffusion into the visualization community applying it to diffusion tensor MRI data [11]. Sanderson et al. used a reaction-diffusion model to generate spot noise based on the underlying vector field placing glyphs at the spot center [21]. Reaction diffusion provides automatic control of density, size, and placement of patterns but the specification of appropriate parameters is not trivial. Stable patterns only form for a very narrow band of values for the parameters. In addition it is computationally expensive. Recently, Kindelmann and Westin proposed a glyph packing algorithm in the context of diffusion tensor visualization [12]. Their work is built on a particle approach simulating attractive and repulsive forces. This work is an extension to our recent work for the generation of anisotropic noise samples [3] by adding a control over the alignment of the ellipses.

3 Assumptions and Goals

The starting point for the generation of the elliptic noise samples is a metric $g = (g_{ij})$ given over a domain $D \subset R^2$, which defines the sample properties. The metric can be user-defined or derived from scalar fields, vector fields, or tensor fields, see Sect. 6.4. The metric is given as a two-by-two symmetric, positive definite matrix depending on the location $P = (x, y) \in R^2$. We assume that the metric is nondegenerate everywhere. In general, it is spatially varying and anisotropic. Size and density of the ellipses are specified by the metric in their center $P_0 = (x_0, y_0)$. Their shape is defined as unit circle with respect to the metric g_0 in P_0, that is,

$$g_{0\,11}(x - x_0)^2 + 2g_{0\,12}(x - x_0)(y - y_0) + g_{0\,22}(y - y_0)^2 = 1. \tag{1}$$

Their half-axes are aligned to the eigenvectors and their squared principal radii $a^2(x_0, y_0)$ and $b^2(x_0, y_0)$ are scaled according to the reciprocal eigenvalues

$$a^2(x_0, y_0) = \frac{1}{\lambda_1(x_0, y_0)} \text{ and } b^2(x_0, y_0) = \frac{1}{\lambda_2(x_0, y_0)}, \tag{2}$$

where $\lambda_1(x_0, y_0)$ and $\lambda_2(x_0, y_0)$ are the eigenvalues of $g(x_0, y_0)$. The sample density is implicitly defined by the size of the ellipses. To make a glyph-based visualization reasonable, we further assume that the frequency of the generated spots is higher than the frequency of the change of the underlying metric. This means that density and eigen-directions do not vary much from one sample to its neighbors. In summary, we have designed our algorithm to generate noise samples with the following properties:

- Size and shape of the spots are determined by the local metric. By choosing the right scaling we can define the spots as unit circles, see (1).

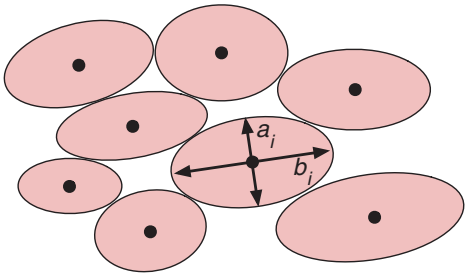


Fig. 1. Generalized Poisson disk property. The minimum distance of two sample points is defined by the local ellipses, which are not allowed to overlap

- The spots are closely packed without holes, resulting in an uniform density, defined as covered pixels per unit area.
- The spots are nonintersecting having a minimum distance, defined by a generalized Poisson disk property, see Fig. 1.
- The degree of alignment of the spots can be controlled by a parameter.

4 Algorithm

Texture generation can be divided into two independent steps:

1. Computation of a reasonable starting distribution of ellipses, where we generate a set of spot candidate based on a dense set of uniformly sampled jittered points, and then traverse the candidate set to select ellipses such that the resulting distribution fulfills a generalized Poisson disk property. This start distribution provides the basis for most of the properties of the resulting sample set.
2. Optimization of the starting distribution using an anisotropic Lloyd relaxation. Dependent on a parameter controlling the anisotropy, the relaxation more or less favors an alignment.

Both steps are important. The first step determines the number of samples, the density and the Poisson disk property. The second step leads to a more uniform sample distribution, approaching a stable configuration. In the following we shortly explain each step. For more details we refer to [3].

4.1 Generating the Initial Sample Set

The generation of the initial sample set is done in two steps. In the first step, a set of jittered grid points is generated as locations for the candidate spots. The initial set must have higher density than the target density. Four times the target density leads to good results. The candidate spots in each location are defined by the local metric as "unit-circle." For a general metric g,

these are ellipses defined by (1). At this stage, the generated candidate spots generally overlap. Once the initial set of points is generated, the algorithm traverses the set of points. A candidate spot is accepted when its ellipse does not overlap with the ellipses at any other already selected samples. The underlying regular grid structure of random points has the nice property that it supports efficient spatial search of neighboring points, therefore simplifying the checking process.

4.2 Anisotropic Voronoi Relaxation

By eliminating overlapping samples, holes can result in certain areas. To remove these artifacts we use a method similar to *Lloyd relaxation*.

Lloyd relaxation (also known as Voronoi iteration) is a method to generate evenly distributed samples. It is an iteration of constructing Voronoi tessellation and its centroids. In each iteration the sample points are moved to the cell centroid, which corresponds to the center of mass of the cell. The process converges against a centroidal Voronoi diagram, where each sample point lies in its cell centroid. This diagram minimizes the energy given as

$$E = \sum_{i \in I} \int_{Vor_i} \rho(x) ||r - r_i||^2 \mathrm{d}r, \tag{3}$$

where I is the index set for the samples, Vor_i the Voronoi cell of the ith sample, r_i its position, and ρ a local scalar density.

Because of the anisotropy of the metric, we use an anisotropic Voronoi diagram and an anisotropic centroid computation for the relaxation step. For the definition of the anisotropic Voronoi diagram and the centroid computation, we built on the works of Labelle and Shewchuk [16] and Du et al. [2]. Our method is a combination of these two methods, satisfying our demands. Depending on the special choice of the metric used to define the Voronoi cells the alignment of ellipses is more or less supported.

4.2.1 Definition of the Voronoi Regions

Let $\{P_i \in D, i \in I\}$ be the set of sample points resulting from our previous step, where I is an index set for the points. The most natural way of generalizing the Voronoi tessellation to other more general metrics would be to define a Voronoi cell $Vor(P_i)$ of a point P_i as the set of all points $P \in D$ that are at least as close to P_i as to any other point $P_j, j \neq i$, using the geodesic instead of the Euclidean distance. However, since the computation of this shortest path is difficult and computationally expensive, we use an approximate distance function for two points proposed by Labelle and Shewchuk, because it matches our conditions well, that is,

$$d^2(P, Q) = (P - Q)^T g(P)(P - Q). \tag{4}$$

The distance measure is not symmetric $d(P,Q) \neq d(Q,P)$. Also, the triangle inequality is not necessarily satisfied. Based on this approximate distance, a Voronoi cell of point P_i is defined as

$$Vor(P_i) = \{P \in D | d(P_i, P) \leq d(P_j, P) \quad \text{for all } j \in I \text{ with } i \neq j. \tag{5}$$

When using the metric defining the ellipses, this distance function guarantees that ellipses of our start configuration lie entirely inside the Voronoi cells. The resulting Voronoi cells are in general not convex and may not even be connected. Therefore, we define a localized version of the Voronoi cells considering only the part containing the sample point. For more details we refer to [3]. We define

$$Vor_r(P_i) = \{P \in D | i \in I_P \text{ and } d(P_i, P) \leq d(P_j, P) \text{ for all } j \in I_P \text{ with } i \neq j,$$
$$\text{with } I_P = \{i \in I | (P_i - P_j) \cdot (P - P_j) \geq 0, \forall j \neq i\}. \tag{6}$$

4.2.2 Centroid Definition

For the definition of the centroid we follow the idea of Du et al. [2], which is a straight-forward generalization of centroid definition as the center of mass to an anisotropic setting. The center of mass c_i of a Voronoi cell $Vor(P_i)$ is defined as

$$c_i = \frac{\int_{Vor(P_i)} d(r) r \ \mathrm{d}r}{\int_{Vor(P_i)} d(r) \ \mathrm{d}r}, \tag{7}$$

where d is an isotropic scalar density and $r = (x, y)$. By replacing the density d by the metric tensor g the centroid c_i is defined as

$$c_i = \left(\int_{Vor(P_i)} g(r) \ \mathrm{d}r \right)^{-1} \cdot \left(\int_{Vor(P_i)} g(r) \cdot r \ \mathrm{d}r \right). \tag{8}$$

As an integral over positive definite matrices, the left matrix is always invertible. When using an isotropic metric, this definition reduces to the standard weighted centroid definition. If the metric is uniform, that is, it does not depend on r, the anisotropic centroid definition coincides with isotropic uniform case.

4.3 Implementation

4.3.1 Intersection Test

The initial sampling requires intersection tests between neighboring samples. In the isotropic case, this intersection test is simply the circle to circle intersection test and can be done efficiently. In the more general case, the samples are represented by ellipses. The algebraic method of ellipse to ellipse intersection

test involve solving a quartic polynomial, which is computationally expensive and numerically unstable. We use polylines to approximate the ellipses during intersection test to reduce complexity. This approximation produces good results without the issues involved in the algebraic method.

4.3.2 Relaxation

For the computation of the Voronoi cells and the centroid, we use a discrete approach. Considering the domain as a set of uniform cells represented by their center R, the discretized (8) results in

$$\underbrace{c_i - P_i}_{\equiv T_i} = \underbrace{\left(\sum_{R \in Vor(P_i)} g(R) \right)}_{\equiv M_i}^{-1} \cdot \underbrace{\sum_{R \in Vor(P_i)} g(R) \cdot (R - P_i)}_{\equiv t_i}. \tag{9}$$

Instead of computing the Voronoi cell explicitly and using these cells for the centroid computation, we perform both computations in one step. We initialize all sample positions $P_i, i \in I$, with a zero vector t_i and zero matrix M_i. Next, we march through the discretized domain performing the following steps for each cell represented by the point R:

- Find the Voronoi Cell $Vor_r(P_i)$ containing point R to specify i, by comparing the distances to sample points lying inside a local bin.
- Update the matrix M_i and the vector t_i in the following way:

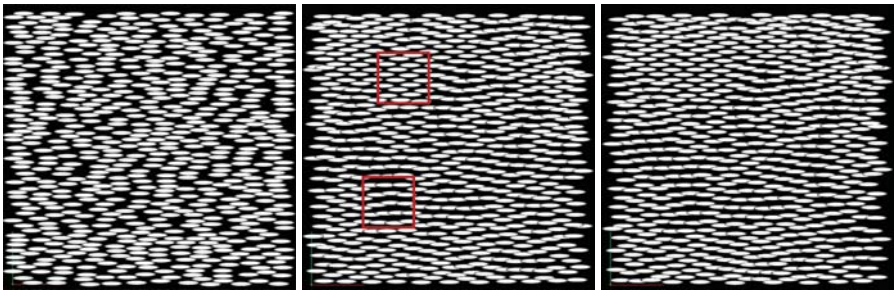
$$\begin{aligned} M_i &\rightarrow M_i + g(R) \\ t_i &\rightarrow t_i + g(R) \cdot (R - P_i). \end{aligned} \tag{10}$$

After traversing the entire domain, the new position of the sample points P_i, given by (9), is determined by the translation vector T_i, that is,

$$T_i = M^{-1} \cdot t_i \text{ and } P_i \rightarrow P_i' = P_i + T_i. \tag{11}$$

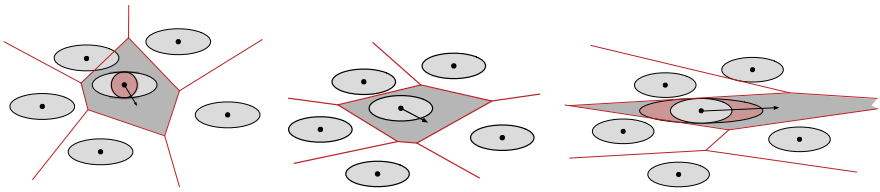
5 Structural Behavior

The distance approximation makes general statements about the convergence behavior of the point set difficult. For the quality of the results the effect of a couple of relaxation steps is more important than convergence. We can identify fix-points of the relaxation process for the uniform case, as, for example, hexagonal structures or any other point symmetric configurations. In our examples we observe that after several iterations, regions with hexagonal patterns are forming, see Fig. 2. Inside these regions the patterns become relatively stable quickly. Between these regions the structure still changes slightly even after many iterations.



(a) (b) (c)

Fig. 2. After many iterations, the sample positions converge and lead to stable patterns. This is especially visible for uniform data sets. After 50 iterations (**b**) we can already see the formation of basically two different structures (highlighted by the *red box*). These regions hardly change over the next 75 iterations (**c**). Looking at difference images between the single iterations we can see that there is still a fluctuation in the regions between the stable patterns. (**a**) The start configuration before relaxation



(a) (b) (c)

Fig. 3. Example of the Voronoi diagram of six points using three different uniform metrics. (**a**) Isotropic metric, (**b**) metric given by the ellipses, (**c**) metric with higher anisotropy

Dependent on the orientation of the ellipses in relation to the orientation of the hexagonal structure, these configurations may result in an alignment. A similar behavior can be observed for slowly varying fields. Whether an alignment is desirable depends on the specific application. While it generates artifacts when the samples are used as input for texture generation or in non-photorelalistic rendering applications, it enhances perception of flow field data sets.

This behavior can be controlled by the anisotropy of the metric used for the Voronoi cells in the relaxation. Figure 3 shows Voronoi cells for different types of anisotropies: isotropic, given by the ellipse and with an exaggerated anisotropy. The shape of the Voronoi cell determines the movement of the sample point in the next iteration. Especially if the start configuration is not very dense, the anisotropy given by the ellipses is not sufficient to prevent the samples from aligning.

In our implementation we have adjusted the anisotropy by multiplying the larger eigenvalue λ_2 with a positive parameter p. Since only the ratio of the eigenvalues is important it is enough to manipulate one eigenvalue. A value of λ_2/λ_1 results in an isotropic metric. A value larger than one leads to a higher degree of anisotropy. Since alignment often goes hand in hand with overlapping of the ellipses, we also implemented a relaxation with intersection test. In case of an intersection with neighboring ellipses the translation vector is shortened. While this prevents the ellipses from overlapping and reduces the alignment, it results in a less uniform distribution, see Fig. 4a. It is important

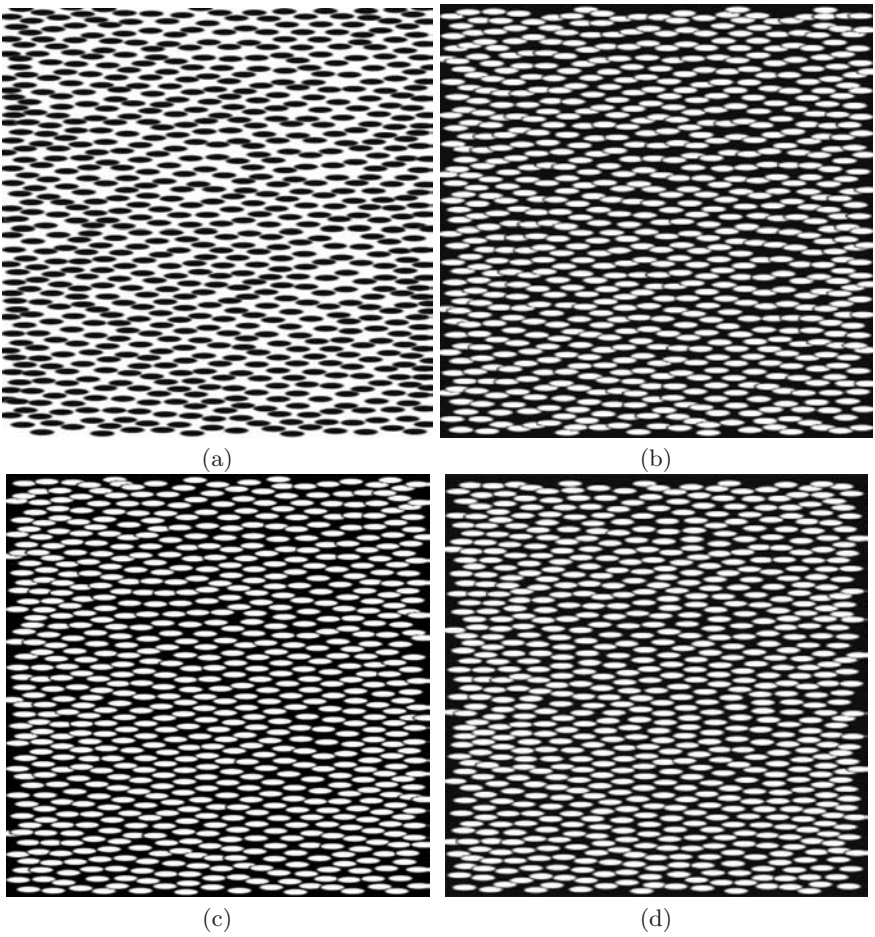


Fig. 4. Relaxation of uniform anisotropic samples. (**a**) The samples after ten relaxations with intersection test. The other images show results after ten iterations without intersection test using three different metrics for the relaxation: (**b**) original metric, (**c**) scaling of the of the larger eigenvalue with two, (**d**) scaling of the larger eigenvalues with three

to note that only the anisotropy of the relaxation process is influenced by the alignment parameter. Size and shape of the represented glyph are not changed and thus still represent the local field properties. It also does not change the general convergence properties but the characteristics of the resulting texture see Fig. 4b–d.

6 Results

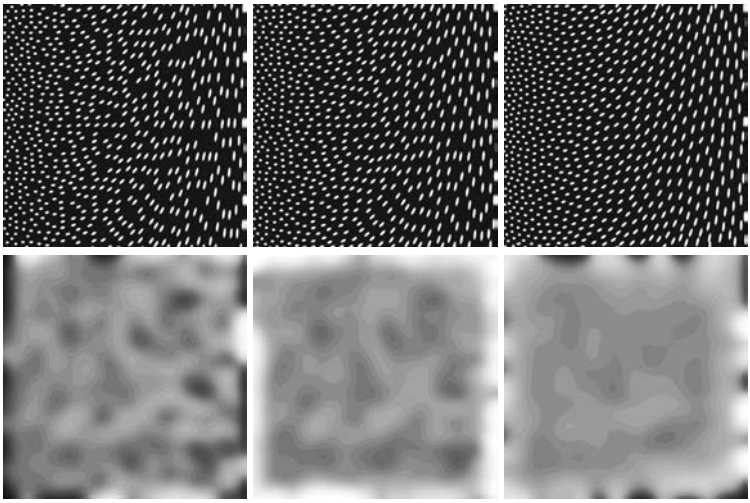
The evaluation of our algorithm is guided by the goals described in Sect. 3. We first discuss examples for a simple isotropic and anisotropic metric definition, which already exhibit most characteristic behaviors of our method. Then we show some results for applications in different contexts. Our main applications are related to visualization, but we also considered "artistic" image rendering applications. The use of glyph sampling with varying density and size is appropriate for any glyph-based visualization, using glyphs that can be embedded into an elliptical shape. For visualization purposes the main step is the definition of the metric, ensuring that it incorporates the most important features of the data. A further analysis of the results in frequency space can be found in [3].

6.1 Representation of the Metric by Sampling Shape and Density

Size and shape of the spots are determined by the local metric. Thus, each spot reflects the metric values at the sample point exactly. The scalar density d, which is defined as covered pixels per unit area, can be measured by using a Gaussian filter. The local density is then given as gray value. Because of the discrete structure of the samples we cannot expect a constant, but almost uniform density. An example for an anisotropic data set is shown in Fig. 5. The size of the Gaussian filter used for these examples is the same for both examples. It can be observed that the density is fast approaching a uniform distribution. After six relaxation steps there are no holes visible anymore. In particular, there is no dependence of the coverage on the size and shape of the samples. Close to the boundary a slightly higher density can be seen for both data sets. This is due to fixed boundary conditions.

6.2 Control of Alignment

To evaluate the influence of the shape of the Voronoi cells on the relaxation, we started with a uniform anisotropic data set. We used the same data set as for Fig. 2, where the initial sample set can be seen in (a). Figure 4 shows results using different relaxation methods after ten iterations. For the generation of Fig. 4a, b, the Voronoi cells are computed using the original metric as given by the ellipses. In Fig. 4a we performed an intersection test after each iteration.



(a) (b) (c)

Fig. 5. Relaxation of nonuniform anisotropic samples. The definition of the Voronoi cells uses the original metric. Top row shows the sample set. (**a**) first sample set, (**b**) after one iteration, (**c**) after ten iterations. The bottom row shows the images after applying a Gaussian blur

This enforces the Poisson disk property but also hinders the relaxation process. There are holes in the dataset even after ten iterations. Figure 4b shows the result without intersection test. We can observe hexagonal structures with and without alignment of the ellipses. There are almost no holes left, but in a few regions the ellipses start overlapping. For Fig. 4c, d we used an exaggerated anisotropy for the Voronoi cell computation. Especially Fig. 4d shows a very uniform structure with almost no overlapping and alignment along the major eigenvector.

6.3 Vector Field Visualization

One of the most direct vector field visualization methods is the use of arrows or other icons. We applied our glyph sampling method to provide a dense placement of glyphs without clustering based on a synthetic vector field. The major eigen-direction of the metric is determined by the direction of the vector field. The major eigenvalue is specified by vector magnitude $\lambda_1 = |v|$ and the minor eigenvalue λ_2 is defined as a constant. The metric is given as

$$g = \lambda_1 e_v \cdot e_v^T + \lambda_2 e_v^\perp \cdot e_v^{\perp T}. \tag{12}$$

Figure 6 shows the results for two different degrees of anisotropy in the relaxation after five iterations. The left image uses the original metric g, and in the right image the larger eigenvalue is scaled by a factor of three. Both images

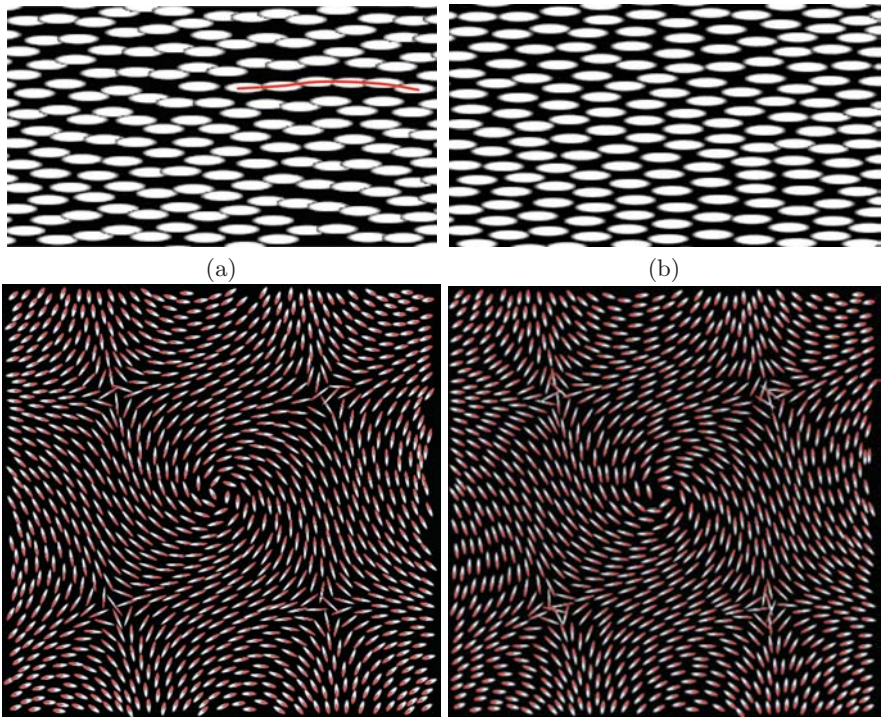


Fig. 6. Effect of manipulating the anisotropy value during the relaxation process: First row shows the tenth relaxation step of a uniform data set starting from the same sample configuration. Second row shows our method for vector field visualization. The left images (**a**,**c**) use the original metric for relaxation, for (**b**,**d**) the larger eigenvalue has been multiplied by three. It can be seen that the original metric favors an alignment along the field lines whereas the exaggerated anisotropy favors an alignment orthogonal to the field lines

show a uniform sampling of the ellipses, but the perception of flow is much better in (a). To achieve similar results, Turk and Banks proposed a method to place arrows along streamlines generated using streamline optimization [23]. Sanderson et al. [21] used a reaction-diffusion model to generate spot noise based on the underlying vector field and places glyphs at spot centers.

6.4 Tensor Field Visualization

To be able to use anisotropic noise for the visualization of tensor data, we must define a metric based on the given tensors. Some of the tensor fields we are interested in are already positive definite, for example, diffusion tensor fields. But other tensor fields, like stress or strain fields, also have negative

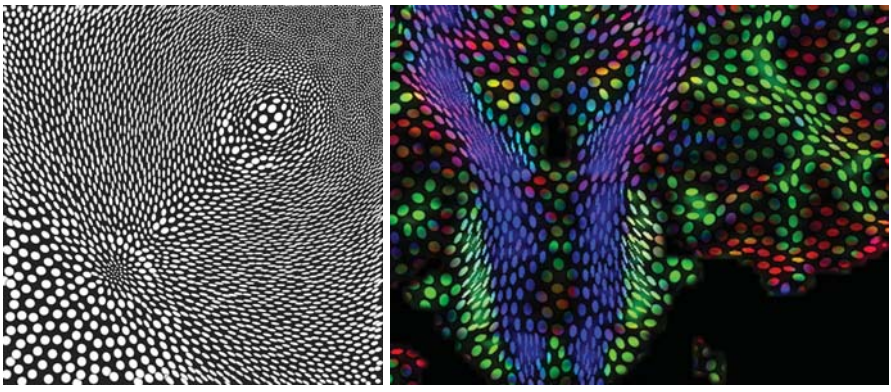
eigenvalues. To be able to treat such tensor fields we interpret them as distortion of a flat metric [9]. Assume that we have a positive definite tensor field T defined over a domain D. Let λ_1 and λ_2 be its eigenvalues and v_1 and v_2 the respective eigenvectors. We define the metric for the sample generation as

$$g = \frac{1}{\sqrt{\lambda_1}} v_1 \cdot v_1^T + \frac{1}{\sqrt{\lambda_2}} v_2 \cdot v_2^T \tag{13}$$

the resulting samples are ellipses aligned to v_1 and v_2 and scaled according to the eigenvalues. Depending on the application it may be necessary to normalize the eigenvalues.

Our first example is a stress field of a solid block with two applied loads with opposite sign, resulting from a three-dimensional numerical finite element simulation. Figure 7a shows a slice of this data set orthogonal to the applied forces. The displayed ellipses represent the shape of a unit sphere deformed according to the local stress field. Small ellipse half-axes indicate compression, large half-axis indicate expansion in the respective direction. Ellipses with high eccentricity mean strong shear forces.

We also applied our method to a slice of a diffusion tensor MRI dataset of a brain. The use of glyphs, ranging from simple ellipses to more advanced glyphs as superquadrics [10], is commonly done for visualizing such data sets. The glyphs are mostly placed in grid points or are randomly spread [17]. Figure 7b shows a result using our sample generation. We used a mask image representing the confidence values of the tensors as provided by Gordon



(a) (b)

Fig. 7. (**a**) Slice of a numerical simulation of a solid block with two forces acting on the block, one pushing and one pulling force. The image shows the tensor data as ellipses. The ellipses give an idea of directions of contraction or deletion inside the material. (**b**) A close-up view obtained after three relaxation steps of a diffusion MRI slice. The color code is the standard color map of encoding the major eigendirection. The projected tensors are represented by ellipses. Each ellipse is defined by the tensor value given at its center

Fig. 8. Mosaic-like images generated by our technique. The metric used for ellipse generation results from the gradient field of the blurred original image. The top image shows the result before relaxation and the bottom image after three relaxation steps

Kindlmann together with the data set. The color is used to represent the principal diffusion direction. The result is a uniform and dense representation of the data independent from the grid points. Similar results were obtained by Kindlmann et al. [12] using a particle simulating approach with repulsive and attractive forces.

6.5 Nonphotorealistic Rendering

Non-photorealistic rendering is often used to simulate painting or drawing styles an artist would use. There are many techniques to simulate these styles. Anisotropic noise samples can be used for generating "artistic images" where elements of the image have directional properties, such as paint brush direction or rectangular mosaic tiles. Our example images were generated by constructing a gradient vector field based on the intensity values of the images. To reduce noise in the vector field, the original images were blurred by applying a Gaussian filter. We defined a tensor metric over the image using the gradient vector field and its orthogonal vector field. The orthogonal vector field essentially points in the direction tangent to the boundary features in the images. One example can be seen in Fig. 8, the left image shows the ellipses before relaxation and the right image after three relaxations, using the original metric.

7 Conclusion

We have introduced a method to generate a dense set of uniformly spread glyphs. Besides the local control of size and density, the methods provides a parameter to control the alignment of the glyphs. This is a desirable property

not only in tensor field visualization. As for all glyph-based methods, the resolution of the representation is limited by the size of the glyphs that are used. The method as described is applicable for two-dimensional fields. A generalization to the three-dimensional case is possible, but computational efficiency and perceptual rendering issues will need to be addressed.

Our method is a purely geometric process. The centroid can be computed explicitly without involving numerics. In contrast to models using repulsive forces, the Voronoi cell based relaxation is very stable. Using a good start configuration of the samples only a few relaxation steps are needed to achieve a uniform distribution. Thus the method is reasonably fast.

Because of the lack of repulsive forces, the Voronoi relaxation does not necessarily preserve the Poisson disk property. We have shown that we can reduce the violation of the Poisson disk property by manipulating the shape of the Voronoi cell appropriately. The key entity thereby is the anisotropy of the metric used for the Voronoi cell definition. In contrast to expensive intersection tests, this approach does not hinder the relaxation process and does not introduce any additional computational costs.

Acknowledgments

The brain dataset is the courtesy of Gordon Kindlmann, Scientific Computing and Imaging Institute, University of Utah, and Andrew Alexander, W. M. Keck Laboratory for Functional Brain Imaging and Behavior, University of Wisconsin-Madison. This work was partially supported by the German Research Foundation DFG (Emmy-Noether Research group) and by the National Science Foundation under contracts ACI 9624034 (CAREER Award), through the Large Scientific and Software Data Set Visualization (LSSDSV) program under contract ACI 9982251, and a large Information Technology Research (ITR) grant; the National Institutes of Health under contract P20 MH60975-06A2, funded by the National Institute of Mental Health and the National Science Foundation. Further, we thank the members of the Visualization and Computer Graphics Research Groups at the Zuse Institute Berlin and the Institute for Data Analysis and Visualization (IDAV) at the University of California, Davis.

References

[1] Q. Du, V. Faber, and M. Gunzburger. Centroidal voronoi tessellations: Applications and algorithms. *SIAM Review*, 41(4):637–676, 1999.

[2] Q. Du and D. Wang. Anisotropic centroidal voronoi tessellations and their applications. *SIAM Journal on Scientific Computing*, 26(3):737–761, 2005.

[3] L. Feng, I. Hotz, B. Hamann, and K. Joy. Anisotropic noise samples. *IEEE Transactions on Visualization and Computer Graphics*, 14(2):342–354, 2008.

[4] L.-P. Fritzsche, H. Hellwig, S. Hiller, and O. Deussen. Interactive design of authentic looking mosaics using voronoi structures. In *Proceedings of 2nd International Symposium on Voronoi Diagrams in Science and Engineering*, Seoul, Korea, 2005.

[5] A. S. Glassner. *Principles of Digital Image Synthesis: Volume Two*. The Morgan Kaufmann Series in Computer Graphics and Geometric Modeling. Morgan Kaufmann, 1995. GLA a 95:2 1.Ex.

[6] A. Hausner. Simulating decorative mosaics. In E. Fiume, editor, *SIGGRAPH 2001, Computer Graphics Proceedings*, pp. 573–578, 2001.

[7] S. Hiller, H. Hellwig, and O. Deussen. Beyond stippling – methods for distributing objects on the plane. *Computer Graphics Forum*, 22(3):515–522, 2003.

[8] A. J. S. Hin and F. H. Post. Visualization of turbulent flow with particles. In *VIS '93: Proceedings of the 4th conference on Visualization '93*, pp. 46–53, 1993.

[9] I. Hotz, L. Feng, H. Hagen, B. Hamann, B. Jeremic, and K. I. Joy. Physically based methods for tensor field visualization. In *VIS '04: Proceedings of IEEE Visualization 2004*, pp. 123–130. IEEE Computer Society Press, 2004.

[10] G. Kindlmann. Superquadric tensor glyphs. In *Proceeding of The Joint Eurographics - IEEE TCVG Symposium on Visualization*, pp. 147–154, 2004.

[11] G. Kindlmann, D. Weinstein, and D. Hart. Strategies for direct volume rendering of diffusion tensor fields. *IEEE Transactions on Visualization and Computer Graphics*, 6(2):124–138, 2000.

[12] G. Kindlmann and C.-F. Westin. Diffusion tensor visualization with glyph packing. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization/Information Visualization 2006)*, 12(5):1329–1336, 2006.

[13] R. M. Kirby, H. Marmanis, and D. H. Laidlaw. Visualizing multivalued data from 2D incompressible flows using concepts from painting. In D. Ebert, M. Gross, and B. Hamann, editors, *IEEE Visualization '99*, pp. 333–340, San Francisco, 1999.

[14] M.-H. Kiu and D. C. Banks. Multi-frequency noise for lic. In *VIS '96: Proceedings of the 7th conference on Visualization '96*, pp. 121–126, Los Alamitos, CA, USA, 1996. IEEE Computer Society Press.

[15] D. E. Knuth. Digital halftones by dot diffusion. *ACM Transactions on Graphics*, 6(4):245–273, 1987.

[16] F. Labelle and J. R. Shewchuk. Anisotropic voronoi diagrams and guaranteed-quality anisotropic mesh generation. In *SCG '03: Proceedings of the nineteenth annual symposium on Computational geometry*, pp. 191–200, New York, NY, USA, 2003. ACM Press.

[17] D. H. Laidlaw, E. T. Ahrens, D. Kremers, M. J. Avalos, R. E. Jacobs, and C. Readhead. Visualizing diffusion tensor images of the mouse spinal cord. In D. Ebert, H. Hagen, and H. Rushmeier, editors, *IEEE Visualization '98 (VIS '98)*, pp. 127–134, Washington–Brussels–Tokyo, Oct. 1998. IEEE, IEEE Computer Society Press.

[18] S. P. Lloyd. Least square quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

[19] V. Ostromoukhov. A simple and efficient error-diffusion algorithm. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pp. 567–572. ACM Press, 2001.

[20] V. Ostromoukhov, C. Donohue, and P.-M. Jodoin. Fast hierarchical importance sampling with blue noise properties. *ACM Transactions on Graphics*, 23(3):488–495, 2004.

[21] A. R. Sanderson, C. R. Johnson, and R. M. Kirby. Display of vector fields using a reaction-diffusion model. In *VIS '04: Proceedings of the conference on Visualization '04*, pp. 115–122, Washington, DC, USA, 2004. IEEE Computer Society.

[22] K. Shimada, A. Yamada, and T. Itoh. Anisotropic triangulation of parametric surfaces via close packing of ellipsoids. *International Journal of Computational Geometry and Application*, 10(4):417–440, 2000.

[23] G. Turk and D. Banks. Image-guided streamline placement. *Computer Graphics*, 30(Annual Conference Series):453–460, 1996.

[24] R. Ulichney. *Digital halftoning*. MIT Press, Cambridge, MA, USA, 1987.

[25] L. Velho and J. de Miranda Gomes. Digital halftoning with space filling curves. In *Proceedings of the 18th annual conference on Computer graphics and interactive techniques*, pp. 81–90. ACM Press, 1991.

The Role of Tensor Fields for Satellite Gravity Gradiometry

Michael Schreiner

University of Buchs, Laboratory for Industrial Mathematics, Werdenbergstrasse 4, CH-9471 Buchs, Switzerland
schreiner@ntb.ch

Summary. Satellite gravity gradiometry is a recent method for a detailed determination of the gravitational potential of the earth. The measurements provided by a gradiometer are – for a so-called full gradiometer – all second-order derivatives of the potential, that is, the Hessian tensor. This leads to a nonstandard problem in potential theory.

After a short description of satellite gradiometry, we focus on uniqueness and existence problems. For this, a certain decomposition of the Hessian tensor is of importance. The decomposition allows in addition to transfer modern solution methods for scalar functions on the sphere to the tensorial case, by generalizing zonal functions to zonal tensor kernels.

1 Satellite Gravity Gradiometry

There is a growing public concern about the future of our planet, its climate, its environment, and expected shortage of natural resources. Any consistent and efficient strategy of protection against these threats depends on a profound understanding of the Earth system. In particular, the knowledge of the Earth mass distribution is of crucial importance for the exploration of processes driving deformation of the Earth surface and influencing ocean surface topography. Closely interrelated with mass transport and mass anomalies is the Earth's gravity field and its constituting ingredients.

If the Earth had a perfectly spherical shape and if the mass inside the Earth were distributed homogeneously or rotationally symmetric, then the line along which a test mass fell would be a straight line, directed radially and going exactly through the Earth's center of mass. The gravitational field obtained in this way would be spherically symmetric. In reality, however, the situation is more complex. The topographic features, mountains and valleys, are very irregular. The actual gravitational field is influenced by strong irregularities in density within the Earth. As a result, the gravitational force deviates from one place to the other from that of a homogeneous sphere.

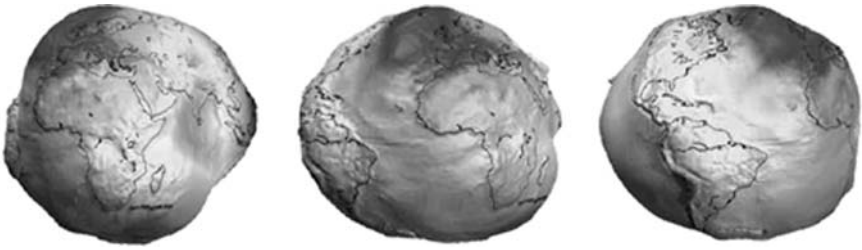


Fig. 1. Geoidal surface (GFZ-EIGEN-CG01C geoid [GFZ])

The knowledge of the gravitational field of the global Earth is of great importance for many applications from which we only mention a few significant examples, for example, geodesy, civil engineering, solid Earth physics, oceanography. A particular role is played for aspects of global "climate change" in the Earth system: Indeed, there is a growing awareness of global environmental problems (e.g., the CO_2-question, the rapid decrease of rain forests, global sea level changes, etc). What is the role of the future airborne methods and satellite missions in this context? They do not tell us the reasons for physical processes, but it is essential to bring the phenomena into one system (e.g., to make sea level records comparable in different parts of the world). In other words, equipotential surfaces such as the geoid (see Fig. 1) are viewed as an almost static reference for many rapidly changing processes and at the same time as a "frozen picture" of tectonic processes that evolved over geological time spans.

Indeed, the gravity field plays a peculiar dual role in Earth sciences. On the one hand, by comparing the actual field with that of an idealized Earth body (e.g., an idealized Earth in hydrostatic equilibrium) their deviations, called *gravity anomalies*, are derivable. The gravity anomalies indicate the state of mass imbalance in the Earth's interior. On the other hand, the *geoid*, that is, the equipotential surface at (mean) sea-level of a hypothetical ocean at rest, serves as the reference surface for all topographical features (for more details see, e.g., [ESA99]).

Internal density signatures of the Earth are reflected by gravitational field signatures, and gravitational field signatures smooth out exponentially with increasing distance from the Earth's body. As a consequence, positioning systems are ideally located as far as possible from the Earth, whereas gravity field sensors are ideally located as close as possible to the Earth. Following these basic principles, various positioning and gravity field determination techniques have been designed. Sensors may be sensitive to local or global features of the gravity field. Considering the spatial location of the data, we may differentiate between terrestrial (surface), airborne, and spaceborne methods. Regarding the data type we have various measurement principles of the gravity field leading to different types of data.

To be more precise, the *force of gravity* provides a directional structure to the space above the Earth's surface. It is tangential to the vertical plumb lines and perpendicular to all (level) equipotential surfaces. Any water surface at rest is part of a level surface. Level (equipotential) surfaces are ideal reference surfaces, for example, for heights. As already mentioned, the geoid is defined as that level surface of the gravity field which best fits the mean sea level. Gravity can be measured by absolute or relative gravimeters. The highest accuracy relative gravity measurements are conducted at the Earth's surface. Measurements on ships and in aircraft deliver reasonably good data only after the removal of inertial noise.

It should be pointed out that the terrestrial distribution of Earth's gravity data on a global scale is far from being homogeneous with large gaps, in particular over oceans and also over land. In addition, the quality of the data is very distinct. Thus, terrestrial gravity data coverage now and in the foreseeable future is far from being satisfactory. This is the reason why spaceborne measurements have to come into play.

The three satellite concepts under present operation are satellite-to-satellite tracking in the high–low mode (SST hi–lo), satellite-to-satellite tracking in the low–low mode (SST lo–lo), and satellite gravity gradiometry (SGG). Representatives of these three concepts are CHAMP (SST hi–lo), GRACE (SST lo–lo combined with SST hi–lo), GOCE (SGG combined with SST hi–lo). Common to all three concepts is that the determination of the Earth's gravity field is based on the measurement of the relative motion (in the Earth's gravity field) of test masses.

The concept of *satellite-to-satellite tracking* (SST) goes back to almost three decades. The original idea was to fly two satellites in an identical low orbit with a separation of a few hundred kilometers between the spacecraft (low–low SST). Between the satellites the distance and the Doppler frequency shift can be measured.

Satellite–gravity–gradiometry (SGG) is a technique of measuring the relative acceleration, not between free falling test masses like satellites, but of measuring test masses at different locations inside one satellite (see Fig. 2). Each test mass is enclosed in a housing and kept levitated (floating, without ever touching the walls) by a capacitive or inductive feedback mechanism. The difference in feedback signals between two test masses is proportional to their relative acceleration and exerted purely by the differential gravitational field. Non-gravitational acceleration of the spacecraft affects all accelerometers inside the satellite in the same manner and so ideally drops out during differencing. The rotational motion of the satellite affects the measured differences. However, the rotational signal (angular velocities and accelerations) can be separated from the gravitational signal if acceleration differences are taken in all possible (spatial) combinations (=full tensor gradiometer). Again low orbit means high sensitivity. The GOCE mission opens a completely new range of spatial scales to research.

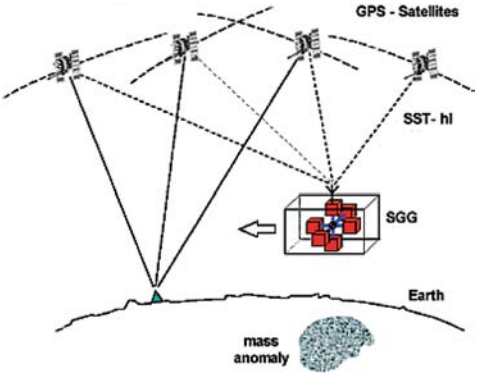


Fig. 2. The principle of satellite gravity gradiometry (from [ESA98])

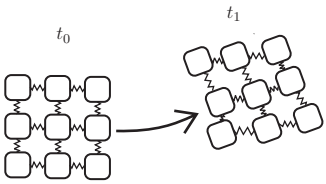


Fig. 3. The principle of a gradiometer

A simplified model of a gradiometer is sketched in Fig. 3. An array of test masses is connected with springs. Measured are differences between the displacements of opposite test masses. This yields information on the differences of the forces. They, however, are due to local differences of ∇V. Since the gradiometer itself is small, these differences can be identified with differentials, so that a so-called full gradiometer gives information on the whole tensor consisting of all second-order partial derivatives of V, that is, the Hessian matrix of V. V is the gravitational potential, which is – by Newton's law – given by

$$V(x) = \int_{\text{earth}} \rho(y) \frac{1}{|x-y|} dV(y), \quad x \in \mathbb{R}^3, \tag{1}$$

with the unknown density ρ inside the earth. V satisfies the Laplace equation outside of the earth, so that we are finally confronted with the following problem arising from satellite gradiometry:

$$\begin{gathered} \Delta V(x) = 0, \quad x \in \mathbb{R}^3 \backslash \text{earth} \\ V(x) = O(1/|x|), \quad \text{as } |x| \to \infty \\ \nabla \otimes \nabla V(x) \text{ or components are given for } |x| = H. \end{gathered} \tag{2}$$

This problem is nonstandard because of the following reasons:

- The problem is ill-posed since the data is not given at the boundary of the domain, but at a surface in the interior of the domain, that is, at a certain height. But the gravitational potential decreases with increasing height, and therefore the process of transforming the data down to the earth surface (often called downward continuation) is unstable.
- The measured data is not the usual data (as, e.g., Dirchlet or Neumann data), but components of the Hessian tensor $\nabla \otimes \nabla V$ of the gravitational potential. Thus, it is – at first sight – not clear which data ensures uniqueness of the problem.

2 Nonuniqueness Results

From potential theory one might expect that the solution of problem (2) is unique, when one scalar component of the Hessian tensor is prescribed for every point x at the sphere $\Omega_H = \{x \in \mathbb{R}^3 \mid |x| = H\}$. This is, however, not true in general. To see this, we construct a counterexample: If $v \in \mathbb{R}^3$ with $|v| = 1$ is given, the second-order directional derivative of V at the point x is

$$v^T \nabla \otimes \nabla V(x) v. \tag{3}$$

Given a potential V, we construct a vector field $v(x)$ on Ω_H, such that the second-order directional derivative (3) is zero. Assume that V is a solution of (2). For each $x \in \Omega_H$, we know that the Hessian tensor $\nabla \otimes \nabla V(x)$ is symmetric. Thus, there exists an orthogonal matrix $A(x)$ so that

$$A(x)^T (\nabla \otimes \nabla V(x)) A(x) = \operatorname{diag}(\lambda_1(x),\ \lambda_2(x),\ \lambda_3(x)), \tag{4}$$

where $\lambda_1(x), \lambda_2(x), \lambda_3(x)$ are the eigenvalues of $\nabla \otimes \nabla V(x)$. From the harmonicity of V it is clear that

$$0 = \Delta V(x) = \lambda_1(x) + \lambda_2(x) + \lambda_3(x). \tag{5}$$

Let $\mu_0 = 3^{-1/2}(1, 1, 1)^T$. We define the vector field $\mu : \Omega_H \to \mathbb{R}^3$ by

$$\mu(x) = A(x)\mu_0, \quad x \in \Omega_H. \tag{6}$$

Then it holds

$$\begin{aligned}
\mu^T(x)(\nabla \otimes \nabla V(x))\mu(x) &= \mu_0^T A(x)^T (\nabla \otimes \nabla V(x)) A(x) \mu_0 \\
&= \frac{1}{3}(1\ 1\ 1) \begin{pmatrix} \lambda_1(x) & 0 & 0 \\ 0 & \lambda_2(x) & 0 \\ 0 & 0 & \lambda_3(x) \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\
&= \frac{1}{3}(\lambda_1(x) + \lambda_2(x) + \lambda_3(x)) \\
&= 0.
\end{aligned}$$

Hence, we have constructed a vector field μ such that the second order directional derivative of V in the direction of $\mu(x)$ is zero for every point $x \in \Omega_H$. It can be easily seen that for a given V there exist many vector fields leading to the same uniqueness problems as the vector field μ. Observing these arguments we are led to the conclusion that the function V is undetectable from the directional derivatives corresponding to μ, see also [Schr94b].

As a matter of fact there do exist conditions under which only one quantity of the Hessian tensor yields a unique solution (at least up to low order spherical harmonics). To formulate these results, a certain decomposition of the Hessian tensor is necessary, which strongly depends on the separability of the Laplace operator with respect to polar coordinates.

3 Spherical Approach for SGG

In this section, we will reformulate the SGG problem in a spherical setting. For this we start with some basic facts concerning special functions on the sphere.

Any $x \in \mathbb{R}^3$ can be decomposed as $x = r\xi$, where the directional part is an element of the unit sphere: $\xi \in \Omega$, $\Omega = \{x \in \mathbb{R}^3 \mid |x| = 1\}$. Let $Y_{n,m} : \Omega \to \mathbb{R}^3$, $n = 0, 1, \ldots$, $m = 1, \ldots, 2n+1$ be an orthonormal set of spherical harmonics (see, e.g., [Mue66, Hob55, FGS98]). They are complete in $L^2(\Omega)$, that is, if $F \in L^2(\Omega)$, it holds with

$$F^\wedge(n,m) = (F, Y_{n,m}) = \int_\Omega F(\xi) Y_{n,m}(\xi)\, d\omega(\xi) \tag{7}$$

that

$$F(\xi) = \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} F^\wedge(n,m) Y_{n,m}(\xi), \quad \xi \in \Omega \tag{8}$$

($d\omega$ is the usual surface measure on Ω).

The outer harmonics $H_{n,m} : \mathbb{R}^3 \backslash \{0\} \to \mathbb{R}$ are defined by

$$H_{n,m}(x) = \frac{1}{|x|^{n+1}} Y_{n,m}(x/|x|). \tag{9}$$

They are harmonic functions and their restrictions coincide on Ω with the corresponding spherical harmonics. Any $F \in L^2(\Omega)$ can thus be identified with a harmonic potential via

$$\sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} F^\wedge(n,m) H_{n,m}. \tag{10}$$

This motivates the following reformulation of the SGG problem (see Fig. 4). Instead of looking for a harmonic function outside the earth, we search

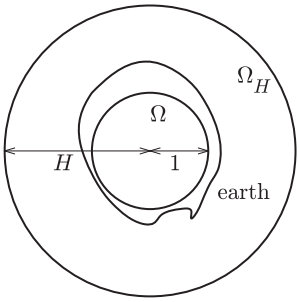


Fig. 4. The Runge–Walsh approach yields a problem in a spherical setting

for a harmonic function outside the unit sphere (assuming the units are chosen in such a way that the sphere with radius 1 is inside the earth and not too far away from the earth's boundary). The justification of this simplification is called Runge–Walsh approach (see [FM04]): To any harmonic function V_0 outside the earth and any given $\varepsilon > 0$, there exists a harmonic function V_1 outside the unit sphere such that

$$|V_1(x) - V_0(x)| < \varepsilon \tag{11}$$

for all x outside the earth. Thus, we finally arrive at the following problem:

$$\begin{gathered} \Delta V(x) = 0, \quad x \in \mathbb{R}^3,\ |x| > 1 \\ V(x) = O(1/|x|), \quad \text{as } |x| \to \infty \\ \nabla \otimes \nabla V(x) \text{ or components are given for } |x| = H. \end{gathered} \tag{12}$$

Using the outer harmonics, we can then represent the potential V by a series

$$\sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} F^{\wedge}(n,m) H_{n,m}. \tag{13}$$

4 Decomposition of Spherical Tensor Fields

We may represent the points $\xi \in \Omega$ in polar coordinates as follows

$$\begin{gathered} \xi = t\varepsilon^3 + \sqrt{1-t^2}(\cos\varphi\varepsilon^1 + \sin\varphi\varepsilon^2), \\ -1 \le t \le 1\ ,\ 0 \le \varphi < 2\pi\ ,\ t = \cos\vartheta\ , \end{gathered} \tag{14}$$

($\vartheta \in [0,\pi]$: (co-)latitude, φ: longitude, t: polar distance), that is,

$$\xi = (\sin\vartheta\ \cos\varphi,\ \sin\vartheta\ \sin\varphi,\ \cos\vartheta)^T\ . \tag{15}$$

Any element $\xi \in \Omega$ may be represented using its coordinates (φ, t) in accordance with (14). For the representation of vector and tensor fields on the unit

sphere, we use a local set of orthonormal unit vectors in the directions r, φ, and t. They are given by

$$\varepsilon^r(\varphi,t) = \begin{pmatrix} \sqrt{1-t^2}\cos\varphi \\ \sqrt{1-t^2}\sin\varphi \\ t \end{pmatrix}, \tag{16}$$

$$\varepsilon^\varphi(\varphi,t) = \begin{pmatrix} -\sin\varphi \\ \cos\varphi \\ 0 \end{pmatrix}, \tag{17}$$

$$\varepsilon^t(\varphi,t) = \begin{pmatrix} -t\cos\varphi \\ -t\sin\varphi \\ \sqrt{1-t^2} \end{pmatrix} \quad . \tag{18}$$

Obviously,

$$\varepsilon^t(\varphi,t) = \varepsilon^r(\varphi,t) \wedge \varepsilon^\varphi(\varphi,t) \quad . \tag{19}$$

The vectors ε^φ and ε^t mark the tangential directions. Since we associate ξ with its representations using the local coordinates φ and t, we identify $\varepsilon^r(\xi)$ with $\varepsilon^r(\varphi,t)$ etc. (cf. Fig. 5).

From (16)–(18) we immediately obtain a representation of the cartesian unit vectors in terms of the spherical ones:

$$\varepsilon^1 = \sqrt{1-t^2}\cos\varphi\varepsilon^r(\varphi,t) - \sin\varphi\varepsilon^\varphi(\varphi,t) - t\cos\varphi\varepsilon^t(\varphi,t), \tag{20}$$

$$\varepsilon^2 = \sqrt{1-t^2}\sin\varphi\varepsilon^r(\varphi,t) + \cos\varphi\varepsilon^\varphi(\varphi,t) - t\sin\varphi\varepsilon^t(\varphi,t), \tag{21}$$

$$\varepsilon^3 = t\varepsilon^r(\varphi,t) + \sqrt{1-t^2}\varepsilon^t(\varphi,t). \tag{22}$$

Gradient fields ∇F can be decomposed into a radial and a tangential component. More explicitly, the surface gradient ∇^* contains the tangential derivatives of the gradient ∇ as follows:

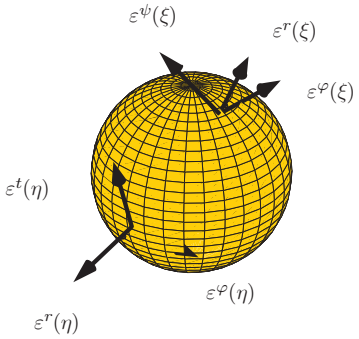


Fig. 5. The local triad ε^r, ε^φ, and ε^t with respect to two different points ξ and η on the unit sphere

$$\nabla = \varepsilon^r \frac{\partial}{\partial r} + \frac{1}{r}\nabla^*. \tag{23}$$

The surface curl gradient L^* is defined by

$$\mathrm{L}^*_\xi F(\xi) = \xi \wedge \nabla^*_\xi F(\xi), \qquad \xi \in \Omega, \tag{24}$$

$F \in \mathrm{C}^{(1)}(\Omega)$. According to its definition (24), $\mathrm{L}^* F$ is a tangential vector field perpendicular to $\nabla^* F$, that is,

$$\nabla^*_\xi F(\xi) \cdot \mathrm{L}^*_\xi F(\xi) = 0, \quad \xi \in \Omega. \tag{25}$$

$\nabla^* \cdot = \mathrm{div}^*$ and $\mathrm{L}^* \cdot = \mathrm{curl}^*$, respectively, denote the surface divergence and the surface curl given by

$$\nabla^*_\xi \cdot f(\xi) = \sum_{i=1}^{3} \nabla^*_\xi F_i(\xi) \cdot \varepsilon^i \tag{26}$$

and

$$\mathrm{L}^*_\xi \cdot f(\xi) = \sum_{i=1}^{3} L^*_\xi F_i(\xi) \cdot \varepsilon^i. \tag{27}$$

Note that the surface curl as defined by (27), that is,

$$\xi \mapsto \mathrm{L}^*_\xi \cdot f(\xi) = \mathrm{curl}^*_\xi f(\xi) = \mathrm{div}^*_\xi (f(\xi) \wedge \xi) = \nabla^*_\xi \cdot (f(\xi) \wedge \xi), \quad \xi \in \Omega, \tag{28}$$

represents a scalar-valued function on the unit sphere Ω in $\mathbb{R}^3$.

The aforementioned relations can be understood from the well-known role of the Beltrami operator Δ^* in the representation of the Laplace operator Δ:

$$\Delta_x = \left(\frac{\partial}{\partial r}\right)^2 + \frac{2}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\Delta^*_\xi \ . \tag{29}$$

For a more detailed introduction to this operators, see [FGS98].

Next we come to second order tensor fields on the unit sphere, that is, $\mathbf{f} : \Omega \to \mathbb{R} \otimes \mathbb{R}$. Tensor fields can be separated into their tangential and normal parts. We set

$$\mathbf{p}_{*,\mathrm{nor}}\mathbf{f} = (\mathbf{f}\xi) \otimes \xi, \tag{30}$$

$$\mathbf{p}_{\mathrm{nor},*}\mathbf{f} = \xi \otimes (\xi^T \mathbf{f}), \tag{31}$$

$$\mathbf{p}_{*,\mathrm{tan}}\mathbf{f} = \mathbf{f} - \mathbf{p}_{*,\mathrm{nor}}\mathbf{f} = f - (\mathbf{f}\xi) \otimes \xi, \tag{32}$$

$$\mathbf{p}_{\mathrm{tan},*}\mathbf{f} = \mathbf{f} - \mathbf{p}_{\mathrm{nor},*}\mathbf{f} = \mathbf{f} - \xi \otimes (\xi^T \mathbf{f}), \tag{33}$$

$$\begin{aligned} \mathbf{p}_{\mathrm{nor,tan}}\mathbf{f} &= \mathbf{p}_{\mathrm{nor},*}(\mathbf{p}_{*,\mathrm{tan}}\mathbf{f}) = \mathbf{p}_{*,\mathrm{tan}}(\mathbf{p}_{\mathrm{nor},*}\mathbf{f}) \\ &= \xi \otimes (\xi^T \mathbf{f}) - (\xi^T \mathbf{f}\xi)\xi \otimes \xi. \end{aligned} \tag{34}$$

The operators $\mathbf{p}_{\mathrm{nor,nor}}$, $\mathbf{p}_{\mathrm{tan,nor}}$, and $\mathbf{p}_{\mathrm{tan,tan}}$ are defined analogously. A vector field $\mathbf{f} : \Omega \to \mathbb{R} \otimes \mathbb{R}$ is called normal if $\mathbf{f} = \mathbf{p}_{\mathrm{nor,nor}}\mathbf{f}$ and tangential if $\mathbf{f} =$

$\mathbf{p}_{\tan,\tan}\mathbf{f}$. It is called left normal if $\mathbf{f} = \mathbf{p}_{\mathrm{nor},*}\mathbf{f}$, left normal/right tangential if $\mathbf{f} = \mathbf{p}_{\mathrm{nor},\tan}\mathbf{f}$, and so on.

The constant tensor fields $\mathbf{i}_{\tan}$ and $\mathbf{j}_{\tan}$ can be defined using the local triads by

$$\mathbf{i}_{\tan} = \varepsilon^{\varphi} \otimes \varepsilon^{\varphi} + \varepsilon^{t} \otimes \varepsilon^{t}, \quad \mathbf{j}_{\tan} = \xi \wedge \mathbf{i}_{\tan} = \varepsilon^{t} \otimes \varepsilon^{\varphi} - \varepsilon^{\varphi} \otimes \varepsilon^{t}. \tag{35}$$

Spherical tensor fields can be discussed in an elegant manner by the use of certain differential processes. Let $u \in \mathrm{c}^{(1)}(\Omega)$ be a vector field given in its coordinate form by

$$u(\xi) = \sum_{i=1}^{3} U_i(\xi)\varepsilon^i, \quad \xi \in \Omega, \quad U_i \in \mathrm{C}(\Omega)^{(1)}. \tag{36}$$

Then we define the operators $\nabla^* \otimes$ and $\mathrm{L}^* \otimes$ by

$$\nabla_\xi^* \otimes u(\xi) = \sum_{i=1}^{3} (\nabla_\xi^* U_i(\xi)) \otimes \varepsilon^i, \quad \xi \in \Omega, \tag{37}$$

$$\mathrm{L}_\xi^* \otimes u(\xi) = \sum_{i=1}^{3} (\mathrm{L}_\xi^* U_i(\xi)) \otimes \varepsilon^i, \quad \xi \in \Omega. \tag{38}$$

Clearly, $\nabla^* \otimes u$ and $\mathrm{L}^* \otimes u$ are left tangential. But it is an important fact that even if u is tangential, the tensor fields $\nabla^* \otimes u$ and $\mathrm{L}^* \otimes u$ are generally not tangential. It is obvious that the product rule is valid. To be specific, let $F \in \mathrm{C}^{(1)}(\Omega)$ and $u \in \mathrm{c}^{(1)}(\Omega)$, then

$$\nabla_\xi^* \otimes (F(\xi)u(\xi)) = \nabla_\xi^* F(\xi) \otimes u(\xi) + F(\xi)\nabla_\xi^* \otimes u(\xi), \quad \xi \in \Omega. \tag{39}$$

In view of the above equations and definitions, we finally introduce operators $\mathbf{o}^{(i,j)} : \mathrm{C}^{(2)}(\Omega) \to \mathbf{c}^{(0)}(\Omega)$ by

$$\mathbf{o}_\xi^{(1,1)} F(\xi) = \xi \otimes \xi F(\xi), \tag{40}$$

$$\mathbf{o}_\xi^{(1,2)} F(\xi) = \xi \otimes \nabla_\xi^* F(\xi), \tag{41}$$

$$\mathbf{o}_\xi^{(1,3)} F(\xi) = \xi \otimes \mathrm{L}_\xi^* F(\xi), \tag{42}$$

$$\mathbf{o}_\xi^{(2,1)} F(\xi) = \nabla_\xi^* F(\xi) \otimes \xi, \tag{43}$$

$$\mathbf{o}_\xi^{(3,1)} F(\xi) = \mathrm{L}_\xi^* F(\xi) \otimes \xi, \tag{44}$$

$$\mathbf{o}_\xi^{(2,2)} F(\xi) = \mathbf{i}_{\tan}(\xi) F(\xi), \tag{45}$$

$$\mathbf{o}_\xi^{(2,3)} F(\xi) = \left(\nabla_\xi^* \otimes \nabla_\xi^* - \mathrm{L}_\xi^* \otimes L_\xi^*\right) F(\xi) + 2\nabla_\xi^* F(\xi) \otimes \xi, \tag{46}$$

$$\mathbf{o}_\xi^{(3,2)} F(\xi) = \left(\nabla_\xi^* \otimes \mathrm{L}_\xi^* + \mathrm{L}_\xi^* \otimes \nabla_\xi^*\right) F(\xi) + 2\mathrm{L}_\xi^* F(\xi) \otimes \xi, \tag{47}$$

$$\mathbf{o}_\xi^{(3,3)} F(\xi) = \mathbf{j}_{\tan}(\xi) F(\xi), \tag{48}$$

$\xi \in \Omega$.

After our above considerations it is not difficult to prove the following lemma.

Lemma 1. *Let $F : \Omega \to \mathbb{R}$ be sufficiently smooth. Then the following statements are valid:*

1. $\mathbf{o}^{(1,1)}F$ *is a normal tensor field.*
2. $\mathbf{o}^{(1,2)}F$ *and* $\mathbf{o}^{(1,3)}F$ *are left normal/right tangential.*
3. $\mathbf{o}^{(1,1)}F$ *and* $\mathbf{o}^{(3,1)}F$ *are left tangential/right normal.*
4. $\mathbf{o}^{(2,2)}F$, $\mathbf{o}^{(2,3)}F$, $\mathbf{o}^{(3,2)}F$ *and* $\mathbf{o}^{(3,3)}F$ *are tangential.*
5. $\mathbf{o}^{(1,1)}F$, $\mathbf{o}^{(2,2)}F$, $\mathbf{o}^{(2,3)}F$ *and* $\mathbf{o}^{(3,2)}F$ *are symmetric.*
6. $\mathbf{o}^{(3,3)}F$ *is skew-symmetric.*
7. $\left(\mathbf{o}^{(1,2)}F\right)^T = \mathbf{o}^{(2,1)}F$ *and* $\left(\mathbf{o}^{(1,3)}F\right)^T = \mathbf{o}^{(3,1)}F$.
8. *For* $\xi \in \Omega$,

$$\operatorname{trace} \mathbf{o}_\xi^{(i,k)}F(\xi) = \begin{cases} F(\xi) & \text{for } (i,k) = (1,1) \\ 2F(\xi) & \text{for } (i,k) = (2,2) \\ 0 & \text{for } (i,k) \neq (1,1),\ (2,2) \end{cases}.$$

The tangent representation theorem (cf. [Bac66, Bac67]) asserts that if $\mathbf{p}_{\tan,\tan}\mathbf{f}$ is the tangential part of a tensor field $\mathbf{f} \in \mathbf{c}^{(2)}(\Omega)$, as defined already, then there exist unique scalar fields $F_{2,2},\ F_{3,3},\ F_{2,3},\ F_{3,2}$ such that

$$\int_\Omega F_{2,2}(\xi)\, d\omega(\xi) = \int_\Omega F_{3,3}(\xi)\, d\omega(\xi) = 0, \tag{49}$$

$$\int_\Omega F_{3,2}(\xi)(\varepsilon^i \cdot \xi)\, d\omega(\xi) = \int_\Omega F_{2,3}(\xi)(\varepsilon^i \cdot \xi)\, d\omega(\xi) = 0,\ i = 1,2,3, \tag{50}$$

and

$$\mathbf{p}_{\tan,\tan}\mathbf{f} = \mathbf{o}^{(2,2)}F_{2,2} + \mathbf{o}^{(2,3)}F_{2,3} + \mathbf{o}^{(3,2)}F_{3,2} + \mathbf{o}^{(3,3)}F_{3,3}. \tag{51}$$

Furthermore, the following orthogonality relations may be formulated: Let $F, G : \Omega \to \mathbb{R}$ be sufficiently smooth. Then for all $\xi \in \Omega$, it holds $\mathbf{o}_\xi^{(i,k)}F(\xi) \cdot \mathbf{o}_\xi^{(i',k')}F(\xi) = 0$ whenever $(i,k) \neq (i',k')$.

The adjoint operators $O^{(i,k)}$ satisfying

$$\int_\Omega \mathbf{o}^{(i,k)}F(\xi) \cdot \mathbf{f}(\xi)\ d\omega(\xi) = \int_\Omega F(\xi)\ O^{(i,k)}\mathbf{f}(\xi)\ d\omega(\xi), \tag{52}$$

that is, in shorthand notation

$$(\mathbf{o}^{(i,j)}F, \mathbf{f}) = (F, O^{(i,k)}\mathbf{f}) \tag{53}$$

for all sufficiently smooth functions $F : \Omega \to \mathbb{R}$ and tensor fields $\mathbf{f} : \Omega \to \mathbb{R}\otimes\mathbb{R}$ can be deduced by elementary calculations. We omit the details here (see, e.g., [FGS98]) and immediately formulate a generalization of the Helmholtz

decomposition theorem. As it is well-known, the classical version guarantees that every tangential vector field can be written as a sum of an irrotational and a solenoidal field. For tensor fields we have

Theorem 1. *(Helmholtz decomposition theorem) Let* $\mathbf{f}$ *be of class* $\mathbf{c}^{(2)}(\Omega)$. *Then there exist uniquely defined functions* $F_{i,k} \in \mathrm{C}^{(2)}(\Omega)$, $(i,k) \in \{(1,1), (1,2),\ldots,(3,3)\}$ *with* $(F_{i,k}, Y_0)_{L^2\Omega} = 0$ *for all spherical harmonic* Y_0 *of degree* 0, *if* $(i,k) \in \{(1,2), (1,3),(2,1),(2,3),(3,1),(3,2)\}$ *and* $(F_{i,k}, Y_1)_{L^2\Omega} = 0$ *for all spherical harmonics* Y_1 *of degree 1 if* $(i,k) \in \{(2,3),(3,2)\}$, *in such a way that*

$$\mathbf{f} = \sum_{i,k=1}^{3} \mathbf{o}^{(i,j)} F_{i,k}, \tag{54}$$

where the functions $\xi \mapsto F_{i,k}(\xi)$, $\xi \in \Omega$, *are explicitly given using the operators* $O^{(i,k)}$ *and the Green's function to the spherical Beltrami operators, see [Fre79] or [FGS98].*

This decomposition will be of crucial importance to get uniqueness results for the satellite gravity gradiometry problem.

5 Existence and Uniqueness of the SGG Problem

Suppose that the function $H : \mathbb{R}^3 \setminus \{0\} \to \mathbb{R}$ is twice continuously differentiable. We want to show how the Hessian matrix restricted to the unit sphere Ω, that is,

$$\mathbf{h}(\xi) = \nabla_x \otimes \nabla_x H(x)|_{|x|=1}, \quad \xi \in \Omega, \tag{55}$$

can be decomposed according to the rules of Theorem 1. To evaluate

$$\nabla_x \otimes \nabla_x H(x) = \left(\xi \frac{\partial}{\partial r} + \frac{1}{r}\nabla_\xi^*\right) \otimes \left(\xi \frac{\partial}{\partial r} + \frac{1}{r}\nabla_\xi^*\right) H(r\xi), \tag{56}$$

we first see that

$$\xi \frac{\partial}{\partial r} \otimes \xi \frac{\partial}{\partial r} H(r\xi) = \xi \otimes \xi \left(\frac{\partial}{\partial r}\right)^2 H(r\xi), \tag{57}$$

$$\xi \frac{\partial}{\partial r} \otimes \frac{1}{r}\nabla_\xi^* H(r\xi) = -\frac{1}{r^2}\xi \otimes \nabla_\xi^* H(r\xi) + \frac{1}{r}\xi \otimes \nabla_\xi^* \frac{\partial}{\partial r} H(r\xi), \tag{58}$$

$$\frac{1}{r}\nabla_\xi^* \otimes \xi \frac{\partial}{\partial r} H(r\xi) = \frac{1}{r}\mathbf{i}_{\tan}(\xi)\frac{\partial}{\partial r} H(r\xi) + \frac{1}{r}\nabla_\xi^* \left(\frac{\partial}{\partial r} H(r\xi)\right) \otimes \xi, \tag{59}$$

$$\frac{1}{r}\nabla_\xi^* \otimes \frac{1}{r}\nabla_\xi^* H(r\xi) = \frac{1}{r^2}\nabla_\xi^* \otimes \nabla_\xi^* H(r\xi). \tag{60}$$

Summing up these terms we obtain

$$\begin{aligned}\nabla_x \otimes \nabla_x H(x)|_{|x|=1} = {} & \xi \otimes \xi \left(\frac{\partial}{\partial r}\right)^2 H(r\xi)|_{r=1} \\ & + \xi \otimes \nabla_\xi^* \left(\frac{\partial}{\partial r} H(r\xi)|_{r=1} - H(\xi)\right) \\ & + \left(\nabla^* \frac{\partial}{\partial r} H(r\xi)|_{r=1}\right) \otimes \xi \\ & + \nabla_\xi^* \otimes \nabla_\xi^* H(\xi) \\ & + \mathbf{i}_{\tan}(\xi) \frac{\partial}{\partial r} H(r\xi)|_{r=1}.\end{aligned} \tag{61}$$

Using (57)–(60) and the definition of the $\mathbf{o}^{(i,k)}$-operators we finally arrive at

$$\begin{aligned}\nabla_x \otimes \nabla_x H(x)|_{|x|=1} = {} & \mathbf{o}_\xi^{(1,1)} \left(\left(\frac{\partial}{\partial r}\right)^2 H(r\xi)|_{r=1}\right) \\ & + \mathbf{o}_\xi^{(1,2)} \left(\frac{\partial}{\partial r} H(r\xi)|_{r=1} - H(\xi)\right) \\ & + \mathbf{o}_\xi^{(2,1)} \left(\frac{\partial}{\partial r} H(r\xi)|_{r=1} - H(\xi)\right) \\ & + \mathbf{o}_\xi^{(2,2)} \left(\frac{1}{2} \Delta_\xi^* H(\xi) + \frac{\partial}{\partial r} H(r\xi)|_{r=1}\right) \\ & + \mathbf{o}_\xi^{(2,3)} \frac{1}{2} H(\xi).\end{aligned} \tag{62}$$

In particular, if we consider an outer (solid spherical) harmonic $H_{n,m} : x \mapsto H_{n,m}(x)$, $H_{n,m}(r\xi) = r^{-(n+1)} Y_{n,m}(\xi)$, $r > 0$, $\xi \in \Omega$, we obtain the following decomposition of the Hessian matrix at height H:

$$\begin{aligned}\nabla \otimes \nabla H_{n,m}(h\xi) = {} & (n+1)(n+2) \frac{1}{H^{n+3}} \mathbf{o}_\xi^{(1,1)} Y_{n,m}(\xi) \\ & - (n+2) \frac{1}{H^{n+3}} \left(\mathbf{o}_\xi^{(1,2)} Y_{n,m}(\xi) + \mathbf{o}_\xi^{(2,1)} Y_{n,m}(\xi)\right) \\ & - \frac{(n+1)(n+2)}{2} \frac{1}{H^{n+3}} \mathbf{o}_\xi^{(2,2)} Y_{n,m}(\xi) + \frac{1}{2} \frac{1}{H^{n+3}} \mathbf{o}_\xi^{(2,3)} Y_{n,m}(\xi).\end{aligned} \tag{63}$$

Keeping in mind that any solution of (12) can be expressed as a series of outer harmonics and using the completeness of the spherical harmonics in the space of square-integrable functions on the unit sphere, it follows that the SGG problem is uniquely solvable (up to some low order spherical harmonics) by the $O^{(1,1)}$, $O^{(1,2)}$, $O^{(2,1)}$, $O^{(2,2)}$, and $O^{(2,3)}$ components. To be more specific, we formulate the following theorem:

Theorem 2. *Let V be a solution of (12). Then the following statements are valid:*

1. $O^{(i,k)}\nabla \otimes \nabla V(H\xi) = 0$ *if* $(i,k) \in \{(1,3),(3,1),(3,2),(3,3)\}$.
2. $O^{(i,k)}\nabla \otimes \nabla V(H\xi) = 0$ *for* $(i,k) \in \{(1,1),(2,2)\}$ *if and only if* $V = 0$.
3. $O^{(i,k)}\nabla \otimes \nabla V(H\xi) = 0$ *for* $(i,k) \in \{(1,2),(2,1)\}$ *if and only if* $V|_\Omega$ *is constant.*
4. $O^{(2,3)}\nabla\otimes\nabla V(H\xi) = 0$ *if and only if* $V|_\Omega$ *is linear combination of spherical harmonics of degree 0 and 1.*

This theorem gives a detailed information of which tensor components of the Hessian tensor ensure the uniqueness of the problem, see also [Schr94, FGS98, FMN02].

6 Tensorial Zonal Kernels

In the previous sections, we have seen which tensor components of the Hessian matrix ensure unique solutions of the SGG problem. An essential tool for these results is Fourier series with respect to spherical harmonics and outer harmonics, respectively. The extension to the tensor case was done with certain differential operators. The definition of these operators is made without an explicit use of a coordinate system on the sphere. Therefore, we have no problems with singularities, which are an intrinsic feature of every global coordinate system on the sphere.

In this chapter, we shortly discuss – or better motivate – how the described separation of tensor fields in their $O^{(i,k)}$-components can be used for solution schemes. Using spherical harmonics for the solution of such inverse problems is frequency oriented, that is, the functions are globally supported and show no space localization. Therefore, they are not the best choice for the solution of problems where the data density is varying or when there are regions without data, as it is the case for many satellite problems (polar gaps). Alternative methods (we describe them first in the scalar case) are based on zonal kernels on the sphere.

A zonal kernel K is defined by a function $K : [-1,1] \to \mathbb{R}$, which is used as a spherical functions in the following way:

$$\xi \mapsto K(\xi \cdot \eta), \quad \xi \in \Omega, \tag{64}$$

where η is a fixed point on the unit sphere. In other words, we get the kernel by a rotation of a scalar function around the axis η. Figure 6 gives an impression of the so-called Abel–Poisson kernels (see, e.g., [FGS98]). Using the addition theorem of spherical harmonics and the Funk–Hecke formula (see, e.g., [Mue66, FGS98]), we obtain the Legendre transform of these kernels:

$$\begin{aligned} K(\xi \cdot \eta) &= \sum_{n=0}^{\infty} \sum_{m=1}^{2n+1} \int_\Omega K(\zeta \cdot \eta) Y_{n,m}(\zeta) d\omega(\zeta) Y_{n,m}(\xi) \\ &= \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} 2\pi \int_{-1}^{1} K(t) P_n(t) dt \; P_n(\xi \cdot \eta) \end{aligned}$$

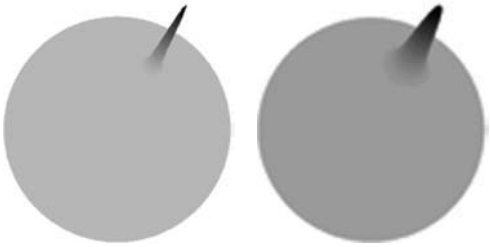


Fig. 6. Two examples of zonal functions on the unit sphere

(P_n is the Legendre polynomial). From this we can see that the harmonic extension of K, that is,

$$x \mapsto \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \frac{1}{|x|^{n+1}} 2\pi \int_{-1}^{1} K(t) P_n(t) dt \; P_n(\frac{x}{|x|} \cdot \eta) \tag{65}$$

is also a zonal function. Thus, the isotropic kernels are well-suited for the representation of harmonic functions.

Furthermore, this approach can be easily generalized to the tensorial case: Using (17), we see that the components of the Hessian tensor of (65) at the height H can be easily obtained. As an example, we give the formula for the $\mathbf{o}^{(1,2)}$-component of (65):

$$-\sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \frac{-(n+2)}{|x|^{n+3}} 2\pi \int_{-1}^{1} K(t) P_n(t) dt \; \mathbf{o}^{(1,2)}_{x/|x|} P_n(\frac{x}{|x|} \cdot \eta). \tag{66}$$

Thus, we can easily transfer modern methods for the solution and regularization with zonal kernels (such as splines, wavelets, etc.) to the tensorial situation. For a more detailed description of these methods, the reader is referred to [Bay00, FGS98, FMN02, Glo02, Hess03, FGlS98, Schr04, FS04].

References

[Bac66] Backus, G.E. Potentials for tangent tensor fields on spheroids. Arch. Ration. Mech. Anal., 22, 210–252 (1966)

[Bac67] Backus, G.E. Converting vector and tensor equations to scalar equations in spherical coordinates. Geophys. J.R. Astron. Soc., 13, 61–101 (1967)

[Bay00] Bayer, M. Geomagnetic field modelling from satellite data by first and second generation wavelets. Doctoral Thesis, Geomathematics Group, University of Kaiserslautern, Shaker, Aachen (2000)

[ESA98] ESA. European views on dedicated gravity field missions: GRACE and GOCE. ESD-MAG-REP-CON-001 (1998)

[ESA99] ESA. Gravity field and steady-state ocean circulation missions. ESTEC, Noordwijk, SP-1233(1) (1999)

[Fre79] Freeden, W. Über eine Klasse von Integralformeln der Mathematischen Geodäsie. Veröff. Geod. Inst. RWTH Aachen, Heft 27 (1979)

[FGS98] Freeden, W., Gervens, T., Schreiner, M. Constructive Approximation on the Sphere (with Applications to Geomathematics). Oxford Science Publications, Oxford (1998)

[FGlS98] Freeden, W., Glockner, O., Schreiner, M. Spherical panel clustering and its numerical aspects. J. Geod. 72, 586–599 (1998)

[FM04] Freeden, W., Michael, V. Multiscale Potential Theory (With Applications to Geoscience), Birkhüser, Boston, Basel, Berlin (2004)

[FMN02] Freeden, W., Michel, V., Nutz, H. Satellite-to-satellite tracking and satellite gravity gradiometry (Advanced techniques for high-resolution geopotential field determination). J. Eng. Math., 43, 19–56 (2002)

[FS04] Freeden,W., Schreiner, M. Multiresolution analysis by spherical up functions. Constr. Approx., 23(3), 241–259 (2004)

[GFZ] GFZ-EIGEN-CG01C geoid from the GeoForschungs-Zentrum Potdsam (http://www.gfz-potsdam.de) (2005)

[Glo02] Glockner, O. On numerical aspects of gravitational field modelling from SST and SGG by harmonic splines and wavelets (with application to CHAMP data). Doctoral Thesis, Geomathematics Group, University of Kaiserslautern, Shaker, Aachen (2002)

[Hess03] Hesse, K. Domain decomposition methods in multiscale geopotential determination from SST and SGG. Doctoral Thesis, Geomathematics Group, University of Kaiserslautern, Shaker, Aachen (2003)

[Hob55] Hobson, E.W. The Theory of Spherical and Ellipsoidal Harmonics. Chelsea Publishing Company, New York (1955)

[Mue66] Müller, C. Spherical harmonics. Lecture Notes in Mathematics 17, Springer, Berlin (1966)

[Schr94] Schreiner, M. Tensor spherical harmonics and their application to satellite gradiometry. Doctoral Thesis, Geomathematics Group, University of Kaiserslautern, Shaker, Aachen (1994)

[Schr94b] Schreiner, M. Uniqueness problems in satellite gradiometry. In: Proceedings of the 8th Conference of the European Consortium for Mathematics in Industry, Kaiserslautern, September 6–10 (1994)

[Schr04] Schreiner, M. Wavelet approximation by spherical up functions. Habilitation Thesis, Geomathematics Group, University of Kaiserslautern, Shaker, Aachen (2004)

Tensor Visualization and Defect Detection for Nematic Liquid Crystals using Shape Characteristics

T.J. Jankun-Kelly[1], Song Zhang[1], A.C. Callan-Jones[2], Robert A. Pelcovits[3], V.A. Slavin[4], and David H. Laidlaw[4]

[1] Department of Computer Science and Engineering, Bagley College of Engineering, Mississippi State University, Mississippi State, MS, USA
{tjk, szhang}@cse.msstate.edu
[2] Institut Curie, UMR CNRS 168, 26rue dUlm, 75248 Paris Cedex 05, France
Andrew.Callan-Jones@curie.fr
[3] Department of Physics, Brown University, Providence, RI, USA
pelcovits@physics.brown.edu
[4] Department of Computer Science, Brown University, Providence, RI, USA
vslavin@gmail.com, dhl@cs.brown.edu

Summary. Two alternate sets of tensor shape characteristics are introduced for the study of nematic liquid crystals, a little studied problem in tensor visualization. One set of characteristics are based on the physics of the liquid crystal system (a real, symmetric, traceless tensor); the other set is an application of the well known Westin DT-MRI shape characteristics. These shape metrics are used both for direct tensor visualization and for detection of defects within the liquid crystal matrix.

1 Introduction

Liquid crystals are substances that exhibit properties between those of a conventional liquid and those of a solid crystal. Liquid crystals are important for technological applications such as display devices and are also relevant to biological research. Physicists use numerical simulations to understand how molecular structures influence observed macroscopic behavior.

Since liquid crystals have very rich topological structures, point- and vector-visualization do not fully convey the structural information. As physicists use tensors to describe the molecular structures, visualization methods are needed to reconstruct the liquid crystal structures from the tensor data.

This chapter presents an overview of tensor visualization methods for nematic liquid crystals (NLCs). Section 2 introduces the behavior of liquid

crystals and the mathematical formulations that describe them. This formalism is extended in Sect. 3 to introduce shape characteristics that describe the tensors in a manner more amenable for visualization. The unique traceless form of this tensor poses new challenges in tensor depiction methods. The two different shape characteristics introduced are mapped on superellipsoids in Sect. 4 to visually distinguish between different types of NLC tensors, while the other set of characteristics are applied in defect detection in Sect. 5. Taken together, these our contributions provide powerful tools for the interrogation and understanding of liquid crystal simulations.

2 Background

Liquid crystals are an intermediate state of matter between the liquid and solid phases [1]. In this middle-state, the material exhibits both liquid-like and solid-like properties. In NLCs, the liquid crystalline phase that we focus on here, molecules do not occupy regular set of positions within a crystal matrix but do tend to favor certain orientations of the molecular axis. Over a small volume of space, the molecules exhibit a random distribution in space but possess an overall average orientation. This orientation – called the *director* **n** – and the strength of the molecular distribution's alignment with it are the fundamental properties that describe the liquid crystal. Since the director specifies an axis but not a direction, it is headless – the sign of the vector does not matter.

NLCs are usually composed of elongated molecules (*calamitic* or rod-like), though disc-like molecules (*discotic* or disc-like) and molecules of intermediary shapes are also possible [2]. There are four possible types of configurations for a macromolecular collection of NLC molecules (Fig. 1):

Isotropic The molecules have no overall alignment. There is no preferred director and no appreciable strength of alignment in any direction. Optically, the crystal has the same index of refraction regardless of orientation or polarization. This state is equivalent to an ordinary liquid.

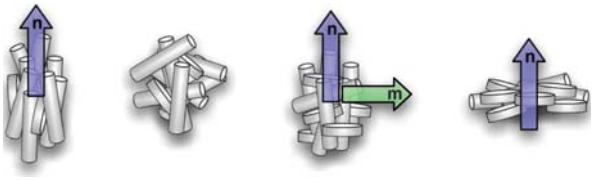


Fig. 1. The four major configuration of a nematic liquid crystal system. From *left to right*: positive uniaxial, isotropic, biaxial, negative uniaxial. Both calamitic (rod-like) and discotic (disc-like) molecules are illustrated

Positive Uniaxial The molecules are on average parallel to the director. There is one optical axis (the director) and light traveling along this axis refracts isotropically. Calamitic molecules dominate in a positive uniaxial state and have their main axis of orientation aligned with the director. Any discotic molecules orient their long axis along the director. The overall configuration is optically positive [2]. Most studied NLC configurations tend toward positive uniaxial configurations under temperatures of interest [1].

Negative Uniaxial Molecules are on average perpendicular to the director. The configuration still has one optical axis, though it is optically negative [2]. Negative uniaxial configurations are usually composed of discotic molecules as the discs are orthogonal to the director; however, calamitic molecules can form negative uniaxial configurations if confined to a plane perpendicular to the director [2].

Biaxial The crystal posses two axes of alignment. The strongest alignment is along the director; however, there is significant, anisotropic alignment along a secondary axis. Because of this anisotropy, biaxial nematics have two optical axes [3]. Biaxiality may be due to mixed concentrations of calamitic and discotic molecules, naturally biaxial molecules, or intermediate states within a positive to negative uniaxial transition [2]. As in the uniaxial cases, biaxial arrangements may be positive or negative depending on the type of alignment with the director. In addition, *neutral biaxial* nematics occur when there is equal positive and negative uniaxial alignment along the primary axes – neither dominates.

Because of the alignment of the NLC molecules, different forms of symmetries are exhibited by the four configurations. The isotropic state has spherical symmetry as there is no preferred orientation. Both unaxial configurations are rotationally symmetric about the director – the configuration is indistinguishable from any rotation of that configuration about the director. Finally, biaxial systems have reflection symmetry – reflections about either the director or the secondary axis preserve the configuration. These symmetries will be exploited when discussing the NLC glyphs in Sect. 4.

Each individual NLC molecule has a primary axis – along the rod for calamitic molecules or along the major axis for ellipsoidal discotics. Though the director is determined by the statistical average of these axes over a volume, the strength of the alignment with the director can vary – two configurations with the same director may be weakly or strongly aligned. The strength of alignment can be thought of as the overall fidelity of the director as a

representation for the collection of molecules; alternatively, it represents the overall order within the system. This *scalar order parameter* S is derived from the statistical average of the angle each molecule makes with the director [4]:

$$S = \frac{1}{2}\left\langle 3\cos^2\theta - 1\right\rangle, \tag{1}$$

where $\langle x\rangle$ represents the statistical average and θ is the angle between the primary axis of the NLC molecule and the director. S is one when all the molecules align with the director (i.e., are positive uniaxial) and $-\frac{1}{2}$ if the molecules are aligned primarily in the plane orthogonal to the director (i.e., are negative uniaxial). In the isotropic state, the molecules are randomly oriented, $\langle\cos^2\theta\rangle = \frac{1}{3}$ in the sampled region, and S is zero [4]. Biaxiality is not directly measured by the scalar order parameter; increases in the amount of biaxiality decrease the order parameter towards zero. Like the director, the change of the scalar order parameter is an important feature in NLC systems.

The alignment of an NLC system is a macromolecular, volumetric value. As such, it is represented by a second-order tensor. The NLC *alignment tensor* – alternatively known as the *order parameter tensor* or the *Q tensor* – encapsulates both the director and the average alignment [4, 5].

Without biaxiality, the alignment tensor for a collection of molecules is

$$\mathbb{Q} = S\left(\mathbf{n}\otimes\mathbf{n}\right) - \frac{1}{3}S\mathbb{I}, \tag{2}$$

where $\mathbf{n}\otimes\mathbf{n}$ is the outer product of the director and $\mathbb{I}$ is the identity tensor, noting that $\mathbb{Q}$, S, and $\mathbf{n}$ vary in time and space. For a biaxial system, two scalar order parameters may be used, one each for the major and medium axes [4]; alternatively, a single biaxial parameter b can be used – this is the approach used here. Regardless of the presence or absence of biaxaility, the alignment tensor is always symmetric and traceless.

The evolution of an NLC system depends on the $\mathbb{Q}$ tensor directly; thus, modeling the tensor is sufficient to describe the system [5]. Though an induced director and scalar order parameter are sufficient to describe the initial configuration of an NLC system, the dynamics of the system evolve these parameters. Thus, new $\mathbf{n}$, S, and b must be extracted from the updated $\mathbb{Q}$ tensor. These parameters are found via eigenanalysis: Assuming the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ of $\mathbb{Q}$ are sorted such that $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$, then, in the eigenvector frame, $\mathbb{Q}$ is represented as [6]

$$[\mathbb{Q}]_\epsilon = \begin{bmatrix} \frac{2}{3}S & & \\ & -\frac{1}{3}S - b_s & \\ & & -\frac{1}{3}S + b_s \end{bmatrix} = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}. \tag{3}$$

In addition, the eigenvector $\mathbf{v}_1$ corresponding to λ_1 (the major eigenvector) is the overall director $\mathbf{n}$ represented in the lab frame (the frame of the simulation), while $\mathbf{v}_2$ is the axis of biaxiality. Here, biaxiality represents the divergence of the NLC from a pure uniaxial configuration. Since uniaxial systems

are rotationally symmetric, a difference in the second and third eigenvalues indicates a loss of this symmetry – an effective decrease in the order along the director. The biaxiality parameter b represents this loss of rotational symmetry: It corresponds to the magnitude of the biaxial terms from the spherical decomposition of the NLC probability density function [6, 7].

Given the eigenanalysis of a $\mathbb{Q}$ tensor, or equivalently its diagonalization, many physical characteristics may be inferred. As mentioned, the eigenvalue with the largest absolute magnitude determines the effective scalar order parameter and the director. If this value is positive, then the configuration is either positive uniaxial or positive biaxial; if negative, the inverse is true. The medium eigenvalue (the second largest in absolute magnitude) determines if biaxiality is present: If it is not equal to the minor (smallest in absolute magnitude) eigenvalue, then the system is biaxial. Otherwise, it is uniaxial or isotropic. All three eigenvalues will be equal only when the system is isotropic and thus $\mathbb{Q} = \mathbb{0}$; the alignment tensor was designed to be zero only in cases of isotropy [5]. Finally, when $\lambda_1 = -\lambda_2 \neq 0$, the system is neutral biaxial. In this case, the sign of the scalar order parameter is ambiguous – either $\mathbf{v}_1$ or $\mathbf{v}_2$ could be the dominant director. This *neutral biaxial line* between the isotropic and maximally biaxial configurations will play an important role in the formulation of our tensor shape metrics in the next section.

3 Tensor Shape Characteristics for NLC Alignment

Second order tensors are represented by 3×3 matrices in a given frame of reference. Algebraically, these matrices belong to different groups depending on their degrees of freedom. By exploiting these degrees of freedom, we can derive *shape characteristics* – metrics over the possible configurations of the tensor that describe the "shape" or form of the tensor. This shape determines the visual geometry of the tensor (Sect. 4) and where rapid shape changes cause topological defects (Sect. 5).

Shape characteristics have been widely used in diffusion tensor imaging (DT-MRI). Westin et al. [8] specified a set of three characteristics describing the geometric behavior of water diffusion based upon the tensor signal: Diffusion is either predominantly linear, planar, or spherical. Because the DT-MRI tensor is a second order real symmetric positive definite tensor, these three shape metrics fully specify the geometry; excluding orientation, second-order real symmetric tensors have only three degrees of freedom [9].

Though the NLC alignment tensor is real and symmetric, it is not positive definite – it is traceless. Thus, one of the two approaches can be taken to specify shape characteristics for NLC systems: Either a set of characteristics based upon real symmetric traceless tensors must be used, or the alignment tensor must be modified to become positive definite. Both approaches are taken in this section.

3.1 Shape Characteristics for Real Symmetric Traceless Tensors

The $\mathbb{Q}$ tensor is a special case of a second-order real symmetric traceless tensor; its eigenvalues are limited to the range $\lambda \in \left[-\frac{1}{3}, \frac{2}{3}\right]$. In general, the eigenvalues of a second-order real symmetric are only restricted to sum to zero and to be real:

$$[\mathbb{T}]_\epsilon = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix} = \begin{bmatrix} \lambda_1 & & \\ & -\frac{1}{2}\lambda_1 - \delta_s & \\ & & -\frac{1}{2}\lambda_1 + \delta_s \end{bmatrix}, \tag{4}$$

where δ represents the divergence from "uniaxiality": $\delta_s = \operatorname{sgn}(\lambda_1)\,\delta$, $\delta \in \left[0, \frac{1}{2}\lambda_1\right]$. The $\mathbb{T}$ tensor exhibits the major properties of an NLC system: It varies between (1) a positive uniaxial representation, (2) a maximally biaxial state, finally arriving at (3) a negative uniaxial arrangement [10] (Fig. 2). Correspondingly, these states have (1) $\lambda_1 > 0$ and $\lambda_2 = \lambda_3 < 0$, (2) $|\lambda_1| = |\lambda_2|$, and (3) $\lambda_1 < 0$ and $\lambda_2 = \lambda_3 > 0$, respectively. This variation is characterized by three shape metrics:

$$\mu_+^{\mathrm{sf}} = \begin{cases} -\frac{2\lambda_3}{\lambda_1} & \lambda_1 \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mu_\delta^{\mathrm{sf}} = \frac{2\delta}{|\lambda_1|}$$

$$\mu_-^{\mathrm{sf}} = \begin{cases} 0 & \lambda_1 \geq 0 \\ \frac{2\lambda_3}{\lambda_1} & \text{otherwise} \end{cases}$$

Note that the isotropic case when the tensor is $\nvDash$ corresponds to a maximally biaxial tensor with 0 eigenvalues.

The above shape characteristics are *scale-free* – they are independent of the magnitude of the eigensystem. NLC alignment tensors, however, are *scale-dependent* as the range of allowed eigenvalues is constrained. Given a maximum eigenvalue $\lambda_{\max} > 0$, this constraint facilitates the definition of four shape characteristics:

$$\mu_+^{\mathrm{sd}} = \begin{cases} -\frac{2\lambda_3}{\lambda_{\max}} & \lambda_1 \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

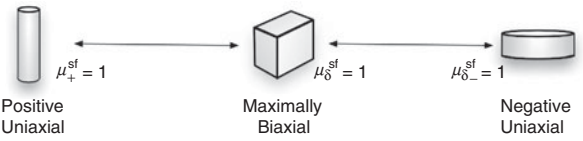


Fig. 2. Characteristic scale-free tensor geometry for second-order real symmetric traceless tensors

$$\mu_{\delta}^{\mathrm{sd}} = \frac{2\delta}{\lambda_{\max}} = \frac{|\lambda_1 + 2\lambda_3|}{\lambda_{\max}} \tag{6}$$

$$\mu_{i}^{\mathrm{sd}} = 1 - \mathrm{sgn}\,(\lambda_1)\,\frac{\lambda_1}{\lambda_{\max}} \tag{7}$$

$$\mu_{-}^{\mathrm{sd}} = \begin{cases} 0 & \lambda_1 \geq 0 \\ \frac{2\lambda_3}{\lambda_{\max}} & \text{otherwise} \end{cases} \tag{8}$$

With the additional degree of freedom (the scale), the shape metrics distinguish between the isotroptic and neutral/maximally biaxial cases – the scale essentially separates the two shapes along the neutral biaxial line (Fig. 3).

The scale-dependent shape metrics form a barycentric space over all possible second-order real symmetric traceless tensors: $\mu_{+}^{\mathrm{sd}} + \mu_{\delta}^{\mathrm{sd}} + \mu_{i}^{\mathrm{sd}} + \mu_{-}^{\mathrm{sd}} = 1$ and all are nonnegative for any given tensor. Thus, they determine the "type" of such tensors: The metric with the largest value identifies the tensor geometry to which it is most similar. For example, an NLC collection is "predominantly positive uniaxial" if $\mu_{+}^{\mathrm{sd}} \gg \left[mu_{i}^{\mathrm{sd}}, \mu_{\delta}^{\mathrm{sd}}, \mu_{-}^{\mathrm{sd}}\right]$.

Equations (5)–(8) are mostly suitable for NLC systems; however, they induce a symmetry about the neutral biaxial line. In contrast, the $\mathbb{Q}$ tensor space is not symmetric about this line – $\lambda_{\max} = \frac{2}{3}$ when $S > 0$ but $\lambda_{\max} = -\frac{1}{3}$ otherwise. Thus, there are combinations of the scale-dependent metrics that are not physically valid (e.g., when $\mu_{-}^{\mathrm{sd}} > 0.5$). To avoid invalid combinations, two different $\lambda_{\max}$'s corresponding to the positive and negative spaces can be used as long as they are made continuous at the neutral biaxial line [10]:

$$\mu_{u^+} = \begin{cases} -3\lambda_3 & \lambda_1 \geq 0 \\ 0 & \lambda_1 < 0 \end{cases} \tag{9}$$

$$\mu_{b} = |3\lambda_1 + 6\lambda_3| \tag{10}$$

$$\mu_{i} = \begin{cases} 1 - \frac{3}{2}\lambda_1 & \lambda_1 \geq 0 \\ 1 + 3\lambda_1 & \lambda_1 < 0 \end{cases} \tag{11}$$

$$\mu_{u^-} = \begin{cases} 0 & \lambda_1 \geq 0 \\ 6\lambda_3 & \lambda_1 < 0. \end{cases} \tag{12}$$

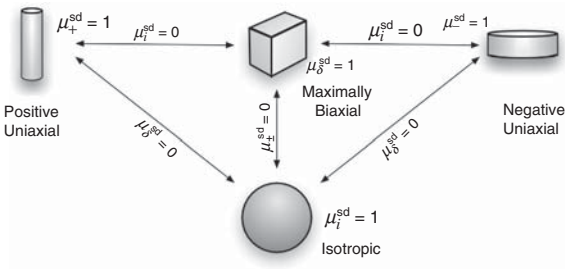


Fig. 3. Characteristic scale-dependent geometry for second-order real symmetric traceless tensors

These metrics split possible $\mathbb{Q}$ tensors into a positive barycentric space described by $\left(\mu_{u^+}, \frac{1}{2}\mu_b, \mu_i\right)$ and a negative barycentric space consisting of $(\mu_{u^-}, \mu_b, \mu_i)$; the difference in μ_b terms is due to the antisymmetry in maximum eigenvalue used. Like before, the largest appropriate metric describes the system – large values of μ_b indicate a biaxial shape that is positive, negative, or neutral depending on the sign of S.

These NLC shape metrics can also be expressed in terms of the primary quantities of the NLC system, S and b:

$$\begin{aligned}
\mu_{u^+} &= \begin{cases} S - 3b & S \geq 0 \\ 0 & S < 0 \end{cases} \\
\mu_b &= 6b \\
\mu_i &= \begin{cases} 1 - S & S \geq 0 \\ 1 + 2S & S < 0 \end{cases} \\
\mu_{u^-} &= \begin{cases} 0 & S \geq 0 \\ -2\left(S + 3b\right) & S < 0. \end{cases}
\end{aligned}$$

The characteristics are the basis of the tensor glyphs described in Sect. 4.

3.2 Using DT-MRI Characteristics for NLC Alignment

Shape metrics for DT-MRI have been well studied; thus, it behooves us to explore their suitability for NLC studies. However, as previously mentioned, the $\mathbb{Q}$ tensor is traceless, not positive definite. If a positive definite form of the tensor could be found, then the Westin shape metrics [8] could be used directly. We explore such a formulation in this section.

The definition of the NLC alignment tensor is based upon the outer product of the director (2). In order for the tensor to be zero everywhere when isotropic, it was made traceless. If we can derive a positive definite form of the alignment tensor, then existing methods for visualizing such tensors can be utilized. To accomplish this, the alignment tensor can be offset [11]

$$\mathbb{D}_\epsilon = \mathbb{Q}_\epsilon + \frac{1}{3}\mathbb{I}. \tag{13}$$

This modifies the system as follows:

$$[\mathbb{D}]_\epsilon = \begin{bmatrix} \lambda_1 + \frac{1}{3} & & \\ & \lambda_3 + \frac{1}{3} & \\ & & \lambda_2 + \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3}S' & & \\ & -\frac{1}{3}S' + b' & \\ & & -\frac{1}{3}S' - b' \end{bmatrix} = \begin{bmatrix} \lambda_1' & & \\ & \lambda_2' & \\ & & \lambda_3' \end{bmatrix}, \tag{14}$$

assuming $S > 0$ and sorting such that $\lambda_1' \geq \lambda_2' \geq \lambda_3' \geq 0$. Note that the new effective scalar order and biaxiality parameters S' and b' may be different from those based upon the $\mathbb{Q}$ approach; in this approach, S' is always nonnegative (see Sect. 3.3). By performing this offset, the medium eigenvalue of $\mathbb{Q}$ has become the minor eigenvalue of $\mathbb{D}$; when $S < 0$, λ_1 and λ_2 switch

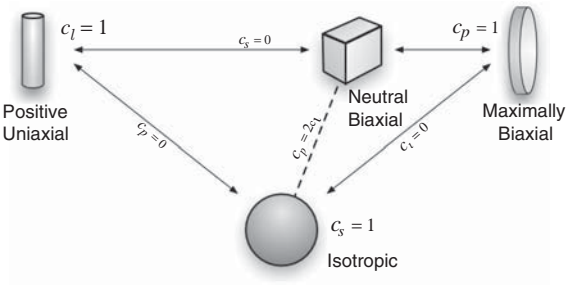


Fig. 4. Characteristic geometry for NLC systems using the Westin metrics. Neutral biaxiality (where the alignment with two major eigenvector is equal) occurs where $2c_l = c_p$ (*dashed line*)

roles and positions in the matrix. The offsetting could cause ambiguity in glyph representations, as discussed in Sect. 4; defect detection is not affected as demonstrated in Sect. 5. Since the smallest possible eigenvalue of $\mathbb{Q}$ is $-\frac{1}{3}$, all eigenvalues of $\mathbb{D}$ are nonnegative. Thus, $\mathbb{D}$ is a positive definite real symmetric tensor.

For DT-MRI, the Westin shape metrics indicate the amount of linear, spherical, or planar anisotropy within the tensor; these correspond to the metrics c_l, c_p, and c_s respectively. Linear anisotropy constrains water diffusion to a line, planar anisotropy possesses movement only in a plane, and so on. Adapting these metrics to $\mathbb{D}$, they describe the amount of positive uniaxiality, isotropy, and biaxiality, respectively [12] (Fig. 4):

$$c_l = \lambda'_1 - \lambda'_2 \tag{15}$$
$$c_s = 3\lambda'_3 \tag{16}$$
$$c_p = 2\,(\lambda'_2 - \lambda'_3)\,. \tag{17}$$

As before, these characteristics form a barycentric space over all possible $\mathbb{D}$. Expressed in terms of the NLC system parameters, these characteristics are

$$c_l = S' - b'$$
$$c_s = 1 - S' - 3b'$$
$$c_p = 4b'.$$

Streamtubes and isosurfaces of these metrics will be used in Sect. 5 to depict defects in crystal cores.

3.3 Reconciling the Two Approaches

While they use different characterizations, the $\mathbb{Q}$ shape characteristics described by (9)–(12) and the Westin-based characteristics for $\mathbb{D}$ in (15)–(17) are reconcilable. The two systems are piecewise linear functions of the other;

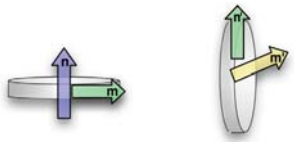


Fig. 5. Consequences of choosing a different reference frame when finding the director. The effective director is always depicted vertically, causing the rotation. Color indicates the same base eigenvector of $\mathbb{Q}$ (thus $\mathbf{m} = \mathbf{n}'$). In the figure, $\mathbf{m}' = \mathbf{v}_3$ is parallel to the plane of the disc

however, the choice of system effects the interpretation of the physics from each. The key observation is that while the $\mathbb{Q}$ formulation uses the dominant eigenvalue to determine its effective director, $\mathbb{D}$ uses the most positive eigenvalue to determine its effective director. This change in effective director (the vector chosen to measure S/S' along) affects the calculations of NLC parameters and the orientation of the eigensystem: When S is negative, S' will be positive due to the change of effective director. The Westin-metric approach always chooses a reference frame that is positive (Fig. 5).

In regions where $\lambda_1 > 0$, there is no difference between the values of S/S' and b/b' using the two approaches as the system is positive uniaxial. When $\lambda_1 < 0$, however, the Westin-metric approach uses the scalar order parameter associated with the largest positive eigenvalue, not the one associated with the largest in absolute magnitude, creating a difference. In this case, the magnitude of S' will be smaller than $|S|$, and b' will exceed $\frac{1}{6}$, the maximum value for the $\mathbb{Q}$ approach [10]. Physically, the effective biaxiality b' increases due to the lack of rotational symmetry about $\mathbf{v}_2$ – the chosen director in the Westin-based approach. As $\mathbb{Q}$ approaches a pure negative uniaxial arrangement, $\mathbf{v}_2$ aligns in the plane of the disc-like configuration and λ_1' tends towards λ_2'. While there is rotational symmetry about $\mathbf{v}_1$, the director used by $\mathbb{Q}$, there is only reflective symmetry about $\mathbf{v}_2$ – the hallmark of biaxial nematics (Fig. 5). In this configuration, $\mathbb{D}$ is maximally biaxial with $c_p = 1$.

Given a set of shape characteristics for $\mathbb{Q}$ or $\mathbb{D}$, the equivalent characteristics in the alternate formulation can be derived. When $\lambda_1 \geq 0$ or equivalently $\lambda_1' \geq \frac{1}{3}$

$$\begin{aligned} \mu_{u^+} &= c_l - \tfrac{1}{2}c_p & c_l &= \mu_{u^+} + \tfrac{1}{3}\mu_b \\ \mu_b &= \tfrac{3}{2}c_p & c_p &= \tfrac{2}{3}\mu_b \\ \mu_i &= c_s + \tfrac{3}{4}c_p & c_s &= \mu_i - \tfrac{1}{2}\mu_b. \end{aligned}$$

Similarly, when $\lambda_1 < 0$ and $\lambda_1' < \frac{1}{3}$:

$$\begin{aligned} \mu_{u^-} &= c_p - 2c_l & c_p &= \mu_{u^-} + \tfrac{2}{3}\mu_b \\ \mu_b &= 3c_l & c_l &= \tfrac{1}{3}\mu_b \\ \mu_i &= c_s & c_s &= \mu_i. \end{aligned}$$

4 Shape-Based Tensor Glyphs for NLC Visualization

The alignment tensor encapsulates four major parameters of the local volumetric alignment: The director $\mathbf{n}$, the direction of maximum biaxiality $\mathbf{m}$, the uniaxial scalar order parameter S, and the biaxial divergence b. A visual depiction of these four quantities needs to incorporate the two orientations and the different types of symmetry – axial and reflective – possessed by the different NLC alignments. This section discusses a glyph-based approach to visually encapsulate these characteristics.

A tensor glyph uses the eigen-decomposition of the tensor's representation in order to depict the tensor's geometry; this approach has been used in DT-MRI to show the local water's Brownian motion in conjunction with other visualizations (see Zhang et al. [13] for a survey). Three major approaches have been used to depict NLC tensor geometry: Cylindrical [14], box [15], and ellipsoidal [16] glyphs. Each of these approaches captures portions of the NLC behavior, but fail to depict the full range:

- Cylindrical glyphs are ideal for positive uniaxial arrangements; their azimuthal axial symmetry is the same as that for calamitic nematic molecules. However, since the non-azimuthal axes have equal magnitude, they are inappropriate for biaxial or discotic arrangements.
- Box glyphs are suited for biaxial alignments. They possess reflective symmetry without rotational symmetry, and each axis may take on a separate length. However, the lack of rotational symmetry makes them inappropriate for uniaxial regions.
- Ellipsoidal glyphs address both these issues; they exhibit axial rotational symmetry at one extreme and reflective symmetry at another. But, as previously shown [17], they are ambiguous with respect to viewing – different view projections of the same ellipsoid produce glyphs that are indistinguishable from glyphs of a different tensor.

To address these limitations, we use a system of superquadric tensor glyphs [10]. Superquadrics are a higher order parameterization of quadric surfaces; their shape deformation allow a wide range of symmetry and geometry [18]. For this work, we use superellipsoids:

$$\mathbf{e}\left(\theta,\phi,\alpha,\beta\right) = \begin{bmatrix} \cos_s^{\alpha}\theta\cos_s^{\beta}\phi \\ \sin_s^{\alpha}\theta\cos_s^{\beta}\phi \\ \sin_s^{\beta}\phi \end{bmatrix} \quad \begin{matrix} -\pi \le \theta \le \pi \\ -\frac{\pi}{2} \le \phi \le \frac{\pi}{2}, \end{matrix} \tag{18}$$

where α, β control the shape parametrization and $x_s^y = \mathrm{sgn}(x)|x|^y$. One advantage of superellipsoids is their comprehensibility under rotation and deformation; unlike ellipsoids, superellipsoids are perceptually distinct under rotations and shape parameter change [19, 20]. Superellipsoids have been used in DT-MRI visualization utilizing the Westin shape metrics successfully [17]; our work is inspired by these glyphs but uses a parameterization more appropriate to NLC physics.

4.1 NLC Tensor Glyph Generation

Given a $\mathbb{Q}$ tensor sampled at a position $\mathbf{x}$, an NLC tensor glyph is created in six steps. First, we perform an eigenanalysis of $\mathbb{Q}$ to extract the major, medium, and minor eigenvalues $\lambda_1, \lambda_2, \lambda_3$ (sorted by absolute magnitude) and their corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. From the eigenanalysis, the important parameters of the NLC configuration are derived: $\mathbf{n} = \mathbf{v}_1$, $\mathbf{m} = \mathbf{v}_2$, $S = \frac{3}{2}\lambda_1$, and $b = \left|\frac{1}{2}\lambda_1 + \lambda_3\right|$[1] (from (3)). The glyph generation then proceeds over the following five steps: We first determine the tensor's shape characteristics, then calculate the superellipsoid parameters, scale by the glyph scaling factors, and finally align the glyph based upon the eigenframe. This section explains these steps.

The shape of the tensor glyph is ultimately based upon the geometry of the NLC alignment tensor itself. The NLC shape metrics of (9)–(12) are used to determine the type of shape present within the tensor. After these shape characteristics are known, these must be mapped unto the space of superellipsoids to emulate the types of symmetry the configuration possess. In addition, the transition between the different symmetry types must be continuous – a NLC system halfway between pure positive uniaxial and maximally biaxial should be partially axially symmetric and partially reflective symmetric. Continuity is granted by the nature of superellipsoids – they are designed to perform smooth transitions based upon the shape parameter. Matching the symmetry types requires finding four appropriate shapes and then interpolating between their corresponding shape parameters smoothly. Fortunately, superellipsoids possess the shapes we need – cylinders for positive uniaxial tensors, spheres for isotropic configurations, a top-like shape whose disc is reminiscent of negative uniaxial alignments with an axis for the director, and a box-like shape for biaxial configurations (corners of Fig. 6). By interpolating between the shape parameters α and β for the canonical shapes, we generate a spectrum of tensor shapes:

$$\begin{aligned} \mu_{u^+} \geq 0, \mu_{u^-} = 0 &\Rightarrow \begin{cases} \alpha = (1 - \mu_b)^{\gamma_b} \\ \beta = (1 - \mu_{u^+})^{\gamma_u} \end{cases} \\ \mu_{u^+} = 0, \mu_{u^-} > 0 &\Rightarrow \begin{cases} \alpha = (1 - \mu_b)^{\gamma_b} \\ \beta = 1 + 3\mu_{u^-}^{\gamma_u}, \end{cases} \end{aligned} \tag{19}$$

where the γ_i are sharpness control parameters after the work of Kindlmann [17]. Using this parameterization, α is related to the uniaxial strength while β corresponds to biaxiality.

After the glyph template is determined, the next step is to determine its scale. We cannot use the eigenvalues directly since the tensor is traceless – some of the eigenvalues may be negative or zero. Instead, we encode the

[1] Finding b via the sum of the major and minor eigenvalues is more numerically stable than the traditional difference of the medium and minor eigenvalues due to the potential small difference in the latter values (which causes loss of significant digits [21]).

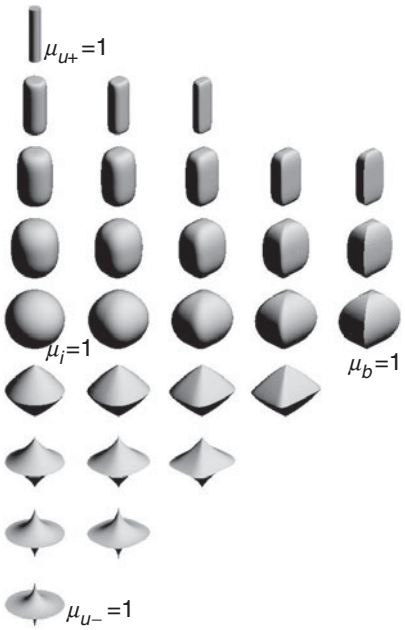


Fig. 6. Nematic liquid crystal tensor glyphs. The uniaxial order parameter increases vertically; biaxiality increases left-to-right. The center horizontal line includes the isotropic and neutral biaxial cases; positive and negative alignments are above and below this line, respectively. The director is vertical in all cases

biaxiality via the difference in the off-director axes scales and the strength of uniaxial alignment as the scale along the director:

$$\mu_{u^+} \geq \mu_{u^-} \Rightarrow \begin{cases} s_x = s_{\min} + (s_{\max} - s_{\min})\,\mu_i \\ s_y = s_{\min} + (s_{\max} - s_{\min})\,(1 - \mu_{u^+}) \\ s_z = s_{\max} \end{cases} \tag{20}$$

$$\mu_{u^+} < \mu_{u^-} \Rightarrow \begin{cases} s_x = s_{\min} + (s_{\max} - s_{\min})\left(1 - \frac{1}{2}\mu_b\right) \\ s_y = s_{\max} \\ s_z = s_{\max}. \end{cases} \tag{21}$$

Typical values of the minimum and maximum scaling factors $s_{\min},\ s_{\max}$ are 0.1 and 0.5, respectively; these produce shapes that fit inside a unit cube. Changing the scale factors allows the glyphs to fit in different volume regions. These scales are continuous along the neutral biaxial line; there, $\mu_i = 1 - \frac{1}{2}\mu_b$.

The last step in generating the glyph is to orient it correctly in space. The director $\mathbf{n}$ and the direction of biaxiality $\mathbf{m}$ provide the frame for the glyph, the third axis of the frame is chosen to ensure that glyph possesses a right-handed coordinate system:

$$\begin{aligned} \mathbb{V}(\mathbb{Q}) = &\ [\mathrm{sgn}\,(\mathbf{v}_2 \cdot \mathbf{v}_1)\,\mathbf{v}_3 \quad \mathbf{v}_2 \quad \mathbf{v}_1] \\ = &\ [\mathbf{l} \qquad\qquad\quad \mathbf{m} \quad \mathbf{n}]\,. \end{aligned} \tag{22}$$

This construction aligns the major axis of the superellipsoid $\hat{\mathbf{z}}$ with $\mathbf{n}$ and the medium axis $\hat{\mathbf{y}}$ with $\mathbf{m}$. The minor axis of the superellipsoid encodes the smallest scaling factor from above (21).

Taken together, (9)–(12) and (18)–(22) specify the following glyph parameterization:

$$\begin{aligned}\mathbf{e}'(\mathbb{Q}, \mathbf{x}, \theta, \phi) &= \mathbb{V}(\mathbb{Q})\,\mathbb{S}(\mathbb{Q})\,\mathbf{e}(\theta, \phi, \alpha(\mathbb{Q}), \beta(\mathbb{Q})) + \mathbf{x} \\ &= \begin{bmatrix}\mathbf{l} & \mathbf{m} & \mathbf{n}\end{bmatrix} \begin{bmatrix} s_x & & \\ & s_y & \\ & & s_z \end{bmatrix} \begin{bmatrix} \cos_s^\alpha \theta \cos_s^\beta \phi \\ \sin_s^\alpha \theta \cos_s^\beta \phi \\ \sin_s^\beta \phi \end{bmatrix} + \mathbf{x}. \end{aligned} \tag{23}$$

Figure 6 depicts the space of glyphs. There is a smooth transition between purely aligned arrangements (left of the figure) towards biaxial arrangements (right of the figure); as biaxiality increases, the minor and medium scaling factors diverge as well. The direction of the director is clearly encoded. Finally, positive and negative alignment is distinguished by the "pinching" in the plane orthogonal to the director – negative alignments have a sharp crease orthogonal to $\mathbf{n}$. The effectiveness of these glyphs will be demonstrated next.

4.2 Application and Discussion

To motivate the effectiveness of the glyphs, we utilize an unstructured grid simulation of a biomolecule in a nematic medium [22] as an example. The embedded biomolecule induces discontinuities in the director alignment – what starts out in a uniaxial alignment will diverge into both positive and negative biaxial cases. The example (Fig. 7) depicts context and close-ups using two approaches: An ellipsoid-based approach [23] and our approach.

Both methods present an effective high-level overview of the tensor field. It is in the detail window that the differences between the techniques become apparent. In the ellipsoid view, differences in the ellipsoids indicate a possible change in director orientation or scalar order strength; however, it is unclear which possibility dominates. In addition, the director in the central region is ambiguous as it cannot be read directly from the ellipsoid. The ambiguity is caused both by the glyph shape and the offset of the eigenvectors used to create the glyph – the eigenvectors are all offset by $\frac{2}{3}$. In our approach, the behavior is clear – the director changes along with the sign of the order parameter. The strong difference in shape preattentively communicates the presence of the large region of negative biaxial alignment in the center of the zoomed region. These negative alignments cannot be gleaned from the ellipsoidal approach.

Figure 8 further illustrates the issues with offset-based approaches for traceless tensor visualization. After applying an offset δ, it may no longer be the case that $|\lambda_1+\delta| > |\lambda_2+\delta|$; if the major eigenvalue was originally negative, it becomes the smallest of the new values when positive. If one naively extracts the director from this new system, an incorrect choice is made – the

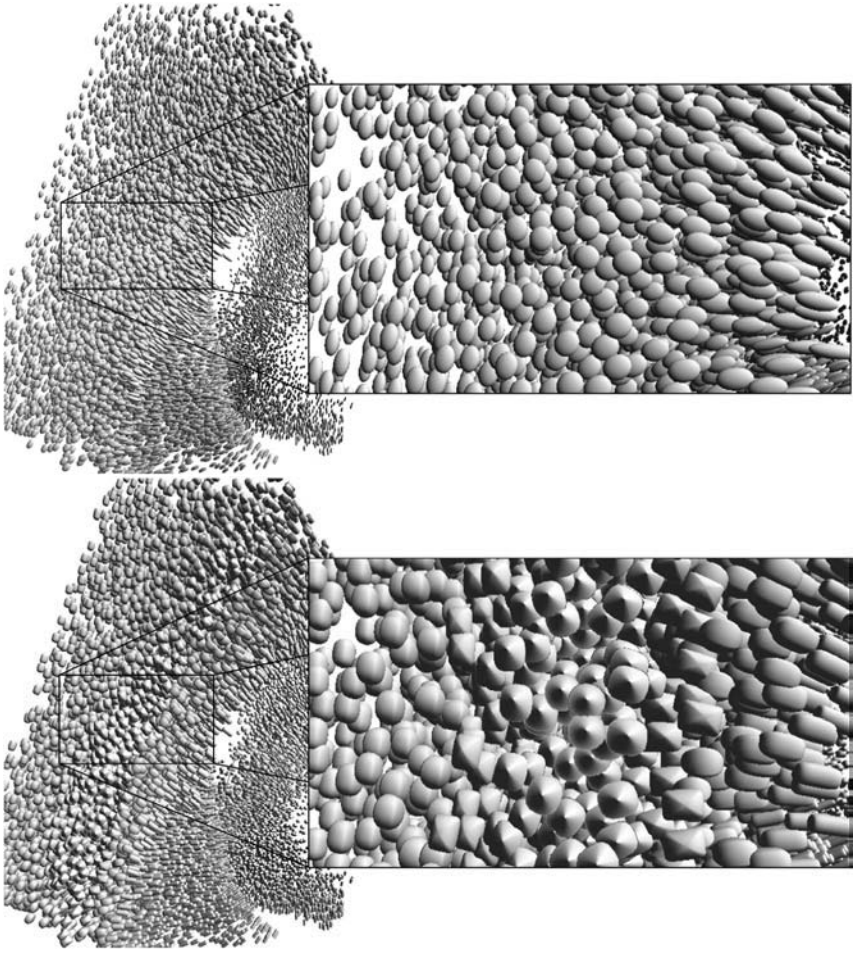
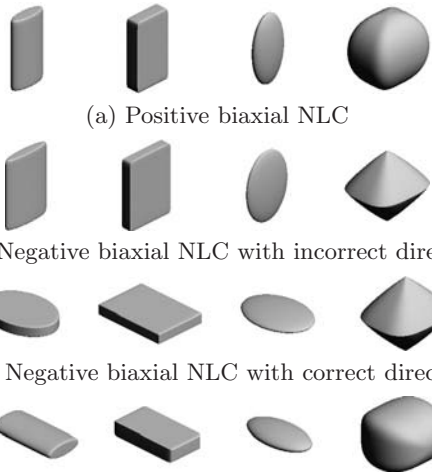


Fig. 7. Comparison of an ellipsoid, eigenvalue-offset approach (*top*) to our approach (*bottom*). The highlighted region contains a defect, indicated by the change in orientation and glyph shape in the bottom image. This change is more difficult to perceive using previous approaches

largest modified eigenvalue is not the largest in absolute magnitude within the original system (Fig. 8b). Even if the correct orientation is extracted (Fig. 8c), the inappropriateness (for cylindrical) or rotational ambiguity (for box and ellipsoid) of the glyph is still present. This ambiguity has been noticed in liquid crystal physics [15] and in general tensor visualization [17]. Our approach avoids these issues by using a distinguishable glyph based upon the system's physics.



(a) Positive biaxial NLC

(b) Negative biaxial NLC with incorrect director

(c) Negative biaxial NLC with correct director

(d) Rotated positive biaxial NLC

Fig. 8. NLC Glyphs using a cylinder, box, ellipsoid, and our method. In the second row (**b**), the wrong director was chosen after the offset was applied due to the major eigenvalue sign flip. Even with the correct director (**c**), there is visual ambiguity for the first three glyphs, especially under rotation (**d**)

5 Detecting Topological Defects with Shape Characteristics

Topological defects play an important role in the structures of NLC. The NLC phase is characterized by long-range orientational order. Under some conditions, such as a quench from a disordered to an ordered phase, confinement of the material, or the synthesis of inverted nematic emulsions, the molecules exhibit discontinuities in this order – these discontinuities are known as topological defects. In NLCs, the main types of defects are monopoles and disclination lines. Figure 9 shows a cutting plane through these defects, illustrating two types of director field patterns, wedge and twist, which can in principle be present. The wedge pattern corresponds to a rotation of the director about an axis parallel to the local direction of the disclination line, while the twist case corresponds to a rotation axis perpendicular to the disclination line direction. All three defects shown are topologically equivalent and can be transformed into each other via suitable continuous rotations of the director field. Finding the three-dimensional structure of these defects is of significant interest to NLC scientists.

Several approaches have been taken to track defects in fluid liquid crystal simulations. Billeter et al. [24] explored a thermal quench of a Gay–Berne nematic [25] consisting of 65,536 molecules in a box with periodic boundary conditions. To track the disclination loops that coarsen and annihilate (cancel each other) after the quench, a cubic lattice of molecules bins was created

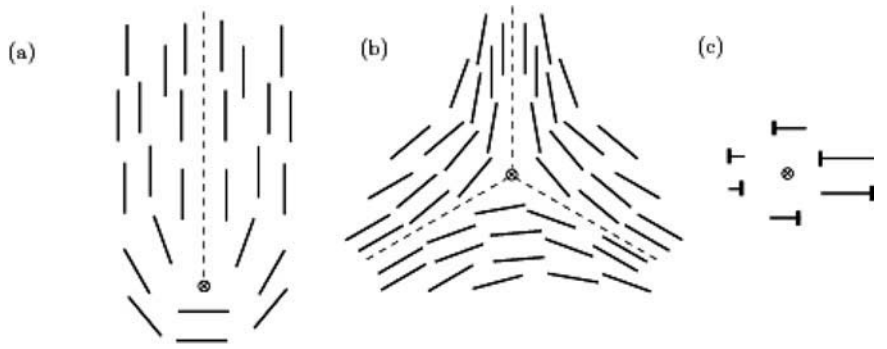


Fig. 9. 2D schematic drawing of $\pm 1/2$ disclination line segments. The disclination cores are indicated by a "$\otimes$", and the disclination line is locally perpendicular to the plane of the page. (**a**) $+1/2$ wedge defect; (**b**) $-1/2$ wedge defect; and (**c**) $+1/2$ twist defect. The rotation of the director in (**c**) is about an axis lying in the plane of the page perpendicular to the disclination line and the nail head representation indicates a projection of the director from below the plane of the page

with the bin size chosen to be approximately equal to the defect core size. Once an average director for each bin was computed, point defects in two dimensions and line defects in three dimensions can be located by traversing a loop around faces of the lattice and checking whether the last director traversed is closer to the first director (no half-integer defect) or its antipode (in which case there will be a defect) [26]. This method was found to be reliable in locating disclination lines: The lines always formed closed loops and a physically reasonable coarsening sequence emerged after the quench. Müller matrices have also been used to simulate Schlieren textures in order to detect defects [27]. The aforementioned approaches work on structured data; for 3D unstructured data, a defect detection has been reported that is based upon a measure of total angular change in molecular orientation over a node neighborhood via the use of a nearest neighbor path [28]. This method replicates the 2D structured grid results where appropriate.

All of the methods discussed above do not show any information about the nature of the director pattern around the defect, namely, the topological charge of the defect and its 3D wedge-like or twist-like nature. To address this problem, in this section we introduce a visualization framework for defect detection [11, 12]. Visualization of these defects help physicists identify the defect cores as well as the molecular structures around them. The interactivity in the framework helped the physicists better understand defect behavior.

5.1 Defect Visualization Method

To visualize defects, we apply second order positive definite tensor visualization methods to NLC data using the offset approach described by (13)–(14).

This visualization technique treats $\mathbb{D}$ as a tensor that describes water self-diffusion as measured via diffusion tensor magnetic resonance imaging [29].

For the visualization, we sampled the NLC data onto a regular grid using an approximating cubic B-spline as the sampling kernel (24):

$$h(\mathbf{x}) = \begin{cases} \frac{1}{6}\left(3r^3 - 6r^2 + 4\right) & 0 \le r \le 1 \\ \frac{1}{6}(2-r)^3 & 1 < r \le 2 \\ 0 & r > 2, \end{cases} \tag{24}$$

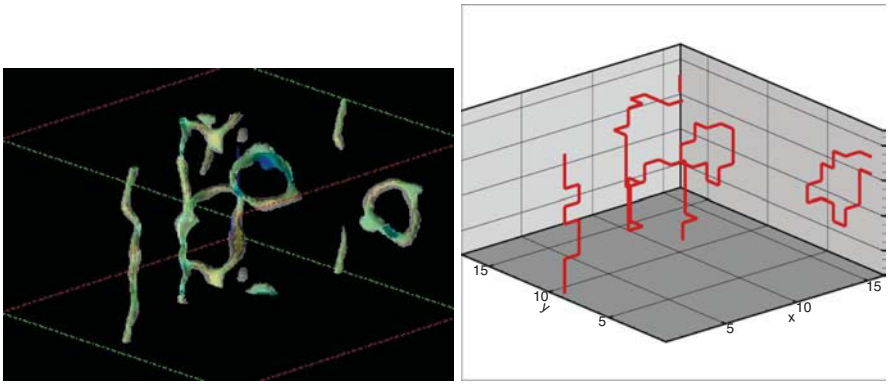
where $r = \frac{2\|\mathbf{x}-\mathbf{x}'\|}{\rho}$ is the sample radius, calculated from the sample point $\mathbf{x}$, the kernel's center $\mathbf{x}'$, and the kernel width ρ. The width of the kernel function was designed so that it would include about 30 molecules (about $\rho = 7.3$); this was empirically determined to be the size of important features in the simulation results. In doing so we converted the discrete molecular simulation data into a regularly sampled, continuous second-order diffusion tensor field. Samples were taken sufficiently closely that we could evaluate a continuous version at any spatial location within the regular grid without introducing aliasing artifacts via interpolation [11]. The sampling was repeated for each time step as needed.

Additional processing of the data was necessary. We introduced padding to preserve the features on the boundary in the resampled field. The original NLC sample is topologically equivalent to a hypertorus because of periodic boundary conditions. We extended the original spatial dimensions of the field by reproducing a portion of the data in every direction. This way the features that wrapped from one side of the data to the next appeared to be continuously repeating thus allowing for better visual coherence.

We calculated integral paths through the principal eigenvector field to create streamtubes in regions with sufficient linear anisotropy [29]. These streamtubes represent the average molecular orientation in regions outside topological defects. Redder color corresponding to larger c_l. Because the defects occur where the relative ordering of molecules is very small, we expected small linear anisotropy values there and no streamtubes.

Regions of low linear anisotropy can represent defects and can be visualized by plotting isosurfaces of linear anisotropy values. Defects form closed linear structures and so these isosurfaces are tori (Fig. 10). The anisotropy value chosen was varied to produce the desired visualization effect. As the linear anisotropy value changed, the diameter of the tube comprising the surface changed as well.

NLC physicists are particularly interested in the structure of the core of the defect, that is, the spatial region within the blue defect tubes. Therefore, we introduced more tools to allow for more detailed exploration. We augmented the visualization of the tensor field features with color-coded cutting planes that display the values of the Westin metrics at various locations in the system. We also mapped the values of the Westin metrics onto the isosurfaces forming



(a) (b)

Fig. 10. Visualization of the disclinations in the simulation data using (**a**) isosurfaces for $c_l = 0.12$, colormap on c_p from blue ($c_p \approx 0$) to red ($c_p \approx 0.7$), and isosurfaces for maximal biaxiality ($\lambda_2' = 1/3$); (**b**) the method of Billeter et al. which was based on standard lattice techniques

the tubes. We enhanced this feature by introducing a tool to probe the values of the field, which allowed for quantitative analysis of the Westin metrics at areas of interest.

Such exploratory and interactive framework for tensor data analysis allows the researchers to map a variety of different data parameters onto the visual representation of their data. This effectively increases the dimensionality of the visualization. Many more data dimensions can be visualized in context with each other than previously was possible. The researchers can pick and choose which features of the defect data they want to explore.

The above tools were built to analyze only one time step. By stitching together the visualization environments as frames of a movie we were able to build an animation of the evolution of the system.

5.2 Case Study: Quenched Gay–Berne Nematic

We applied the visualization methods in the previous section to the simulated data of a quenched Gay–Berne nematic. We analyze the results of the visualization in three aspects: the locations, the surrounding director structures, and the core structures of the disclination lines.

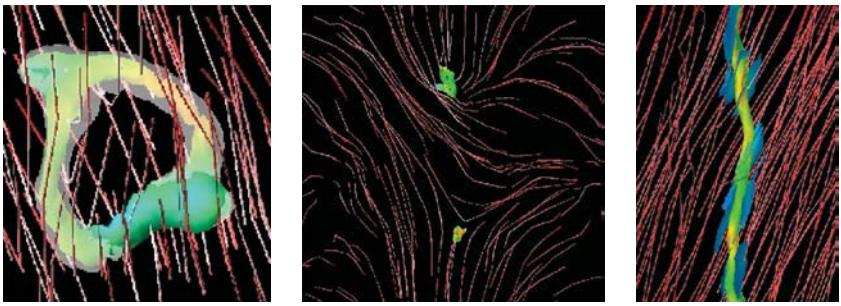
5.2.1 Location of Disclination Lines

Disclination lines are the focus of the NLC defect visualization. Figure 10 shows a visualization of the disclination lines in the Gay–Berne nematic [24] 33,000 timesteps after a thermal quench from the isotropic to the nematic phase. The disclination lines are identified based on the criterion that the

core of a line must be characterized by a very small value of the uniaxial nematic order parameter S or equivalently the Westin metric c_l even if the core is biaxial. The left figure shows isosurfaces of the Westin metric c_l and the intermediate eigenvalue λ_2' of $\mathbb{D}$ (13) using the Advanced Visualization System (AVS) software package [30]. The green–yellow–blue surfaces in the main figure are $c_l = 0.12$ isosurfaces; the color variation on these isosurfaces from blue to green to red indicates increasing value of c_p, a measure of the planar order in the defect core. Within these isosurfaces $c_l < 0.12$ and thus the isosurfaces identify the disclination lines. The gray isosurfaces surrounding the predominantly green ones correspond to $\lambda_2' = 1/3$, demonstrating that the disclination cores include a tube of maximal biaxiality, in agreement with the Landau–Ginzburg theory results [31]. The right figure shows for comparison of the same data visualized using standard methods [24] where a cubic lattice of bins was created and disclinations were found using standard lattice techniques [26]. The general locations of the disclination lines in the two figures are the same aside from the data padding added in the left figure.

5.2.2 Director Structure Around Disclinations

Figure 11 demonstrates how streamtubes can help determine the local topological charge and the rotation axis of the director. The timestep shown is 148,000 after the thermal quench. The streamtubes wrapping through the loop and around the top of the loop indicate that the top is wedge-like with charge $+1/2$ (recall Fig. 9a), whereas the streamtubes near the bottom of the loop indicate that this portion is wedge-like of charge $-1/2$ (recall Fig. 9b).



(a) (b) x-y plane (c) y-z plane

Fig. 11. Director structure around disclinations. (**a**) A wedge-twist loop. The red streamtubes (the thin lines) are along the director field. Redder color corresponding to larger c_l (streamtubes were selected so that $c_l \geq 0.2$). (**b**) The two lines that are roughly parallel to the z axis are viewed from above and appear to be purely wedge-like, with charge $+1/2$ on top and $-1/2$ on the bottom. However, the red streamtubes and blue streamsurfaces in (**c**) (where the lines are viewed along the x axis) indicate that the lines have a twist-like character as well

It is also clear that the disclination loop is twist-like on the left side, as the streamtubes have a component parallel to the disclination loop.

Figure 11 also indicates how using streamtubes as well as streamsurfaces provides various clues in determining the topological charge and director rotation axis $\hat{\boldsymbol{\Omega}}$ of a disclination that would otherwise be very difficult to do. From Fig. 11b which shows a top view of two relatively straight disclination line segments, it might be tempting to conclude that the two disclinations are simply $+1/2$ and $-1/2$ wedge lines. However, examination of the red streamtubes and blue streamsurfaces in a side view of these defects (Fig. 11c) shows a more complicated picture. While the disclinations are on average parallel to the z-axis, the director field outside the core regions and the planar ordering of the director indicated by the streamsurfaces is considerably tilted out of the x-y plane. Further 3D observation confirms that this line has twist as well as wedge character. This does not come as a surprise as the twist Frank elastic constant is less than the splay and bend constants in the Gay–Berne nematic [32].

5.2.3 Core Structure

Using the new visual representation, physicists were able to study the behavior of the tensor field in the core of the defect line and answer a longstanding question in liquid crystal physics: are defect cores characterized by maximum biaxiality or total degeneracy of the eigenvalues (i.e., all three eigenvalues equal to 1/3)? The term maximum biaxiality refers to a case where the three eigenvalues are as different from one another as possible. Degeneracy corresponds to total disorder of the molecular alignment, whereas biaxiality implies some degree of order. It is a subtle question of energetics as to which possibility will be found. The visualization methods presented here offer an immediate answer to this important question, namely that the defect cores are in fact biaxial. This answer is immediately apparent in Fig. 10a, in which the gray isosurfaces correspond to the condition of maximal biaxiality, and does not require a detailed numerical analysis. Thus, the question can be readily answered at any time step of the simulation or for any new simulation done under different initial conditions (including, e.g., varying the molecular parameters). The detailed structure of the core can be explored quantitatively by measuring the Westin metrics on a plane orthogonal to the defect line (Fig. 12).

6 Combining the Two Methods

It is possible to combine the two methods discussed in order to depict the tensor behavior near regions of defect. Though the examples of the defect detection methods were structured data, the metrics can be adapted to unstructured data. Figure 13 shows the detected defect structure using the metrics from Sect. 5 (left) and a method specifically designed for unstructured

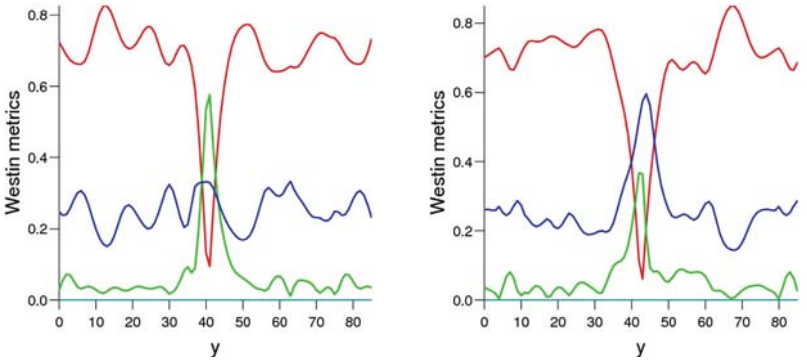


Fig. 12. The Westin metrics c_l (*red*), c_p (*green*), and c_s (*blue*) as functions of the coordinate y for the disclination shown in Fig. 11b; the two disclination lines are roughly parallel to the z axis. The value of the x coordinate is chosen to coincide with the center of each disclination; z is roughly a third of the distance along the depth of the simulation box

Fig. 13. Selected tensors for the defect region for the human Ig antibody dataset using a method adapted from Sect. 5 (*left*). The highlighted region has strong correspondence to the defect nodes detected via changes in director [28] (*right*)

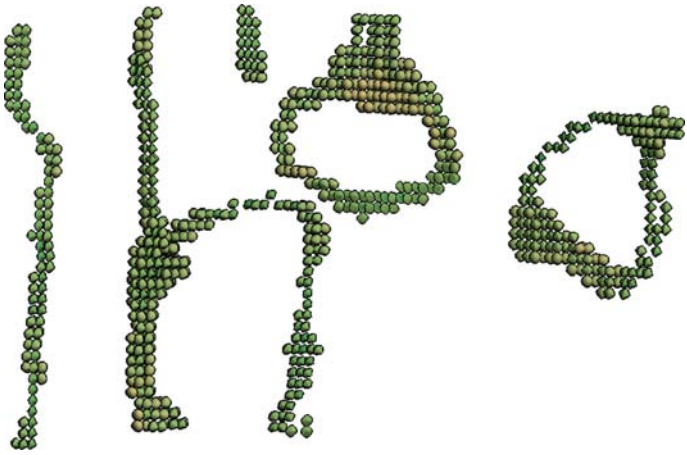


Fig. 14. Tensor visualization using the NLC glyphs from Sect. 4 depicting the structure near the defects discovered in Sect. 5. See Fig. 10a for comparison

NLC simulations [28] (right). The defect structures are very similar, with the glyphs indicating low-biaxiality tensors via their lack of reflection symmetry. Here, low c_l corresponds to regions of low biaxiality (low μ_b) below the neutral biaxial line (thus $\mu_+ = 0$).

Similar visualizations can be performed for the regular grid data set from Sect. 5 (Fig. 14). Note how the planar isotropy indicated by the Westin metrics translates to pinched, negative uniaxial glyphs (the more green glyphs). This corresponds to the change of reference between the Westin-based and traceless tensor-based approaches discussed in Sect. 3.3. The isotropic glyphs near the defect (high c_s) remain spheroid (high μ_i, more brownish).

7 Summary and Conclusions

Shape characteristics are powerful tools that facilitate the visualization and understanding of tensor fields. In this chapter, we have demonstrated how two characterizations of NLC alignment tensors are used to visualize the spatial distribution of alignments and to depict defect structures within the liquid crystal. The two characteristics, one based upon a general, traceless tensor and the other on positive definite tensors, capture the transition between the different canonical NLC alignments, and we have provided a conversion process between the two methods. We also explored how the different mathematical formulations result in different interpretations of the physics via changes in frames of reference.

The shape characteristics-based visualization methods described here provide several benefits to NLC scientists. The glyph-based visualization unambiguously depicts the different classes of NLC alignments without the

orientation uncertainty of ellipsoids while preserving the symmetry properties of the given alignment. The director field stream visualizations can not only be used to determine the location of the defects (as earlier visualization methods could also do), but also to assess the core structure of the defect and director field structure outside the core. Both methods have been used by NLC scientists to positive effect – the former method to understand the behavior of novel NLC media, the latter to answer outstanding questions in NLC physics. Together, our shape characteristic-based visualization methods are effective tools for NLC exploration.

Acknowledgments

We thank Ketan Mehta for his work on unstructured grid-based disclination detection and Gordon Kindlmann for our in-depth discussions. Rajendran Mohanraj and Huangli Wu, formerly of the Mississippi State SimCenter, provided the human Ig anitibody simulation data. This work was partially funded through the National Science Foundation EPSCoR program via award No. 0132618.

References

[1] P. J. Collings. *Liquid Crystals: Nature's Delicate Phase of Matter*, 2nd. edition. Priceton University Press, Princeton, NJ, 2002.
[2] P. K. Mukherjee. Critical behavior of uniaxial–biaxial nematic phase transition. *The Journal of Chemical Physics*, 109(7):2941–2946, 1998.
[3] I. Dierking. *Textures of Liquid Crystals*. Wiley, New York, 2003.
[4] N. Morrtram and C. Newton. Introduction to Q-tensor theory. Technical Report, Department of Mathematics, University of Strathclyde, 2004.
[5] P. G. de Gennes. *The Physics of Liquid Crystals*. Clarendon, Oxford, 1974.
[6] R. J. Low. Measuring order and biaxiality. *European Journal of Physics*, 23:111–117, 2002.
[7] C. Zannoni. Distribution functions and order parameters. In G.R. Luckhurst and G.W. Gray, editors, *The Molecular Physics of Liquid Crystals*, chapter 3, Academic Press, London, pp. 51–83, 1979.
[8] C.F. Westin, S. Peled, H. Gudbjartsson, R. Kikinis, and F. A. Jolesz. Geometrical diffusion measures for MRI from tensor basis analysis. In *Proceedings of the Fifth Annual Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM '97)*, Vancouver Canada, p. 1742, 1997.
[9] G. Kindlmann. *Visualization and Analysis of Diffusion Tensor Fields*. PhD thesis, University of Utah, 2004.

[10] T.J. Jankun-Kelly and K. Mehta. Superellipsoid-based, real symmetric traceless tensor glyphs motivated by nematic liquid crystal alignment visualization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization/Information Visualization 2006)*, 12(5):1197–1204, 2006.

[11] V. A. Slavin, R. A. Pelcovits, G. Loriot, A. Callan-Jones, and D. H. Laidlaw. Techniques for the visualization of topological defect behavior in nematic liquid crystals. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization/Information Visualization 2006)*, 12(5):1323–1328, 2006.

[12] A. C. Callan-Jones, R. A. Pelcovits, V. Slavin, S. Zhang, D. H. Laidlaw, and G. B. Loriot. Simulation and visualization of topological defects in nematic liquid crystals. *Physical Review E*, 74(6), 2006.

[13] S. Zhang, D. H. Laidlaw, and G. L. Kindlmann. Diffusion tensor MRI visualization. In C. D. Hansen and C. R. Johnson, editors, *The Visualization Handbook*, chapter 16, Elsevier Academic Press, Amsterdam, pp. 327–340, 2004.

[14] J. E. Anderson, P. Watson, and P. J. Bos. Comparisons of the vector method and tensor method for simulating liquid crystal devices. *Liquid Crystals*, 28:109–115, 2001.

[15] A. Sonnet, A. Kilian, and S. Hess. Alignment tensor versus director: Description of defects in nematic liquid crystals. *Physical Review E*, 52: 718–722, 1995.

[16] M. Gregory Forest, Q. Wang, and H. Zhou. Methods for the exact construction of mesoscale spatial structures in liquid crystal polymers. *Physica D*, 152–153:288–309, 2001.

[17] G. L. Kindlmann. Superquadric tensor glyphs. In O. Deussen, C. Hansen, D. A. Keim, and D. Saupe, editors, *Proceedings of the Eurographics/IEEE VGTC Symposium on Visualization (VisSym '04)*, pp. 147–154, 2004.

[18] A. H. Barr. Superquadrics and angle-preserving transformations. *IEEE Computer Graphics and Applications*, 1(1):11–22, 1981.

[19] C. D. Shaw, D. S. Ebert, J. M. Kukla, A. Zwa, I. Soboroff, and D. Aaron Roberts. Data visualization using automatic, perceptually-motivated shapes. In *Proceedings of Visual Data Exploration and Analysis*. SPIE, 11 April 1998.

[20] C. D. Shaw, J. Hall, C. Blahut, D. S. Ebert, and D. Aaron Roberts. Using shape to visualize multivariate data. In *Proceedings of the Workshop on New Paradigms in Information Visualization and Manipulation (NPIVM-99)*, ACM Press, New York, pp. 17–20, 1999.

[21] R. L. Burden and J. Douglas Faires. *Numerical Analysis*, chapter 1, PWS Publishing Co., pp. 17–18, 1994.

[22] H. Wu and R. Mohanraj. Computational simulations for study of a liquid-crystal-based sensor system. In J. Graef, H. Lim, R. Shivaji, B. Soni, and J. Zhu, editors, *Proceedings of the Sixth Mississippi State-UAB Conference on Differential Equations and Compuational Simulations*, May 2005.

[23] T. Tsuji and A. D. Rey. Orientation mode selection mechanisms for sheared nematic liquid crystalline materials. *Physical Review E*, 57(5):5609–5625, 1998.

[24] J. L. Billeter, A. M. Smondyrev, G. B. Loriot, and R. A. Pelcovits. Phase-ordering dynamics of the Gay-Berne nematic liquid crystal. *Physical Review E*, 60, 1999.

[25] J. G. Gay and B. J. Berne. Modification of the overlap potential to mimic a linear site–site potential. *Journal of Chemical Physics*, 74(6):3316–3319, 1981.

[26] M. Zapotocky, P. M. Goldbart, and N. Goldenfeld. Kinetics of phase ordering in uniaxial and biaxial nematic films. *Physical Review E*, 51(2):1216–1235, 1995.

[27] C. Chiccoli, O. D. Lavrentovich, P. Pasini1, and C. Zannoni. Monte carlo simulations of stable point defects in hybrid nematic films. *Physical Review Letters*, 79(22):4401–4404, 1997.

[28] K. Mehta and T. J. Jankun-Kelly. Detection and visualization of defects in 3d unstructed models of nematic liquid crystals. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization/Information Visualization 2006)*, 12(5):1045–1052, 2006.

[29] S. Zhang, C. Demiralp, and D. H. Laidlaw. Visualizing diffusion tensor MR images using streamtubes and streamsurfaces. *IEEE Transactions on Visualization and Computer Graphics*, 9(4):454–462, 2003.

[30] C. Upson, T. Faulhaber Jr., D. Kamins, D. H. Laidlaw, D. Schlegel, J. Vroom, R. Gurwitz, and A. van Dam. The application visualization system: A computational e nvironment for scientific visualization. *Computer Graphics and Applications*, 9(4):30–42, 1989.

[31] N. Schopohl and T. J. Sluckin. Erratum: Defect core structure in nematic liquid crystals [Physical Review Letters 59, 2582 (1987)]. *Physical Review Letters*, 60:755, 1988.

[32] M. P. Allen, M. A. Warren, M. R. Wilson, A. Sauron, and W. Smith. Molecular dynamics calculation of elastic constants in Gay-Berne nematic liquid crystals. *Journal of Chemical Physics*, 105:2850–2858, 1996.

A Tensor Approach to Elastography Analysis and Visualization

D. Sosa-Cabrera[1], M.A. Rodriguez-Florido[2,1], E. Suarez-Santana[1,2], and Juan Ruiz-Alzola[1,2]

[1] Center for Technology in Medicine, University of Las Palmas de Gran Canaria, Spain
dario@ctm.ulpgc.es, {marf,esuarez,jruiz}@itccanarias.org
[2] Canary Islands Institute of Technology, SPAIN
{marf,esuarez,jruiz}@itccanarias.org

Summary. Elastography measures the elastic properties of soft tissues using principally ultrasound (US) or magnetic resonance (MR) signals. The elastic behavior of tissues can be analyzed with tensor signal processing. In this work, we propose an analysis of elastography through the deformation tensor and its decomposition into both strain and vorticity tensors. The vorticity gives information about the rotation of the inclusions (simulated tumors) that might be helpful in the discrimination between malign and benign tumors without using biopsy. The tensor strain field visualizes in one image the standard scalar parameters that are usually represented separately in elastography. By using this technique physicians would have complementary information. In addition, it offers them the possibility of extracting new discriminant and useful parameters related to the elastic behavior of tissues. Although clinical validation is needed, synthetic experiments from finite element and ultrasound simulations present reliable results.

1 Introduction

Changes in tissue stiffness are correlated with pathological phenomena, a fact that can help the diagnosis of several diseases such as breast and prostate cancer [7, 8] or cardiovascular dysfunctions [14, 23]. There are mechanical ways to estimate the biomechanical properties of the tissue such as indentation, which is mostly used for thin layers of ex vivo tissue [11, 28]. Many different approaches estimate and image the elastic properties of tissues adapting conventional ultrasound, MR, TC, or nuclear imaging protocols and studying the tissue's response to some stimulus. There are also some investigations applying optical methods [6], These techniques may be referred to as elasticity imaging. A review is found in [17].

Elastography [15], that is, the imaging of the elastic properties of soft tissues, is well established in the literature. In quasi-static USE (ultrasound

elastography), the standard mechanical set-up to produce a controlled compression on the tissue consists of a mechanical arm that holds the ultrasonic probe together with a plate. The use of the plate generates a more homogeneous stress field. However, in the last years, many researchers have focused on simplifying the elastographic set-up. There are comparatives studies [4] that show the capability of freehand elastography (technique that makes the compression with the probe and without the plate) to detect lessons such as breast carcinomas [16], although it has a lower SNR. Different ways of generating the stimulus give different names to the elastographic modalities, as it is the case of sonoelasticity [12], which uses external applied vibrations instead of quasi-static compressions.

The displacement field from which researchers normally obtain the strain is estimated with different techniques. We refer to papers such as [27], which uses a time-domain cross-correlation technique, or [18], which uses an iterative phase-zero estimation, among others. Some researchers visualize the estimated displacement and strain fields following the path in [15]; they focus on the *forward problem*. Some others solve the so-called *inverse problem* [2]; based on the displacement and strain fields, they calculate mechanical properties of the tissue such as Young's modulus, by using the constitutive elasticity equations. In the former, either axial strain or lateral strain [15], Poisson's ratio [20], or shear strain [9] elastograms are visualized. The *inverse problem* approach deals with Young's modulus visualization, the shear modulus [5], or other related parameters. A comparative study between this two approaches can be found in [5].

Although elastography has been shown to be capable of detecting breast tumors in vivo [7], biopsies are still needed to assess their malignancy. Malignant tumors are known to form ramified boundaries that become firmly bound to the surrounding tissue, as opposed to the benign ones, which have smooth borders and are loosely attached to its surrounding. Garra et al. [7] and Konofagou et al. [9] have made efforts trying to reduce the use of biopsies, with USE.

In this chapter, we present the theory and procedure to assess the rotation of the tumor by visualizing the vorticity image aided with the strain tensor field as an effort to better determine the tumor infiltration, supported with a phantom study. Preliminary promising results show the potential of the technique in the diagnosis and prognosis of tumors, by detecting their infiltrating nature in a noninvasive way. This tensorial approach as well as its visualization can be extended to other elasticity imaging techniques.

This work is organized as follows. In Sect. 2, we present the mathematical basis from where we obtain the parameters that we use for visualization. In Sect. 3, we expose the motivation and preliminary investigations that made us deepen in this research. Section 4 presents the set up of the synthetic experiments and their results, and discusses different visualization methods. Finally, Sect. 5 contains the conclusions and an outline of the future work.

2 Background: The Theory of Elasticity

The theory of elasticity is the branch of the physics, specifically of the continuum mechanics, that studies the stresses in solids and the deformation or flow of materials. The branch concerned with the study of liquids and gases is the so-called fluid dynamics. The continuum makes reference to the hypothetical continuous material since the molecular structure of matter and empty spaces or gaps are disregarded. It is also supposed that the functions entering the theory are continuous functions except at surfaces that separate different regions of continuity. The theory of elasticity provides a consistent set of equations, which may be solved to obtain a unique point-wise description of the distribution of the forces experienced at each point (stresses), and the displacements caused (strains), for a particular loading and geometry. We explain in the following sections the main participants in this theory, deepening on the strain that we use for the visualizations. With the strain, we have a relative measure of the elasticity of different tissues that can be used for the diagnosis of several pathologies. Thus, it is not necessary to make any assumption about isotropy nor problem configurations to shorten the independent constitutive equations (1). Neither is it necessary to add computational cost to the procedure. The visualization scheme applied here to the strain tensor can equally be applied to the stress tensor.

2.1 Tensors

Tensors and tensor fields are basic tools in differential geometry and physics to describe geometric and physical quantities that remain invariant under coordinate transformations, as the ones we face in the theory of elasticity.

In this work, a tensor of order (or rank)n living in a d-dimensional space is a mathematical object with n indices and d^n components, which obeys certain transformation rules. In a three-dimensional Euclidean space, such as ordinary physical space, the number of components of a tensor is 3^n. Scalars and vectors are particular cases of tensors of order zero and one, respectively. The strain and stress tensors are both second order tensors, therefore defined by $3^2 = 9$ components.

Some applications include diffusion-tensor magnetic-resonance imaging (DT-MRI) (see first seven chapters), fluid dynamics, electromagnetism, differential geometry, satellite gravity gradiometry (see previous chapter by Schreiner), and nematic liquid crystals visualizations (chapter by Kelly et al.) among others.

2.2 Constitutive Equations

The classical elastic constitutive equations, also called the generalized Hooke's law, are nine equations expressing the stress components as linear homogeneous functions of the nine strain components:

$$T_{ij} = c_{ijrs} E_{rs}. \tag{1}$$

These nine equations contain 81 constants, but since both T_{ij} and E_{rs} are symmetric there are only six independent equations, each containing six terms, giving 36 elastic constants or moduli. For the case of an isotropic material, all the elastic constants can be expressed in terms of two independent ones.

2.3 Stress

In general, two types of external forces can be applied to a solid body: surface forces applied to its boundary (arising from contact with another solid or fluid pressure) and body forces that are distributed throughout the volumetric elements inside the body (gravitational, magnetic, and inertial forces).

Given a point within a body and a plane that passes through that point, the stress vector or surface traction t on that plane at that point is the force per unit area required to keep that body in place, if the material on one side of the plane were instantly removed (Fig. 1), and is defined as

$$\lim_{\Delta S \to 0} \frac{\Delta f}{\Delta S}, \tag{2}$$

where ΔS is an element of area in the interior of the volume V, with normal $\hat{\mathrm{n}}$ and Δf is the surface force acting on ΔS.

The traction vectors on planes perpendicular to the coordinate axes (Fig. 1) are especially useful because, when the vectors acting at a point on three such mutually perpendicular planes are given, the stress vector at that point on any other plane inclined arbitrarily can be expressed on terms of these three given special vectors. The Cauchy tetrahedron analysis results in the stress boundary equations:

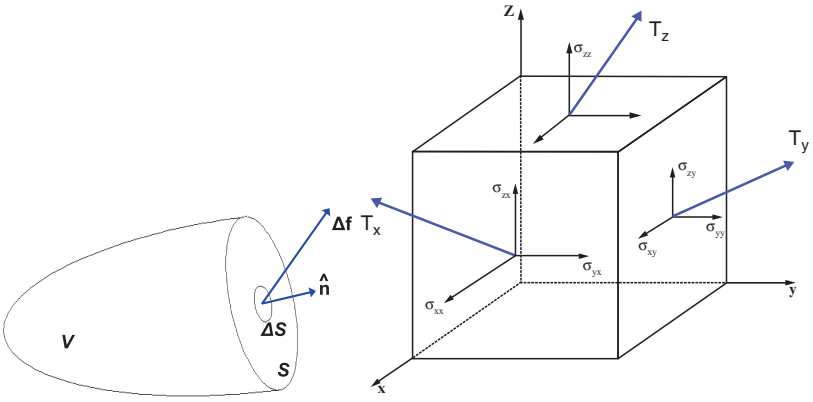


Fig. 1. *Left:* Stress definition; *right:* stress components

$$T_i^{(n)} = T_{ji} n_j. \tag{3}$$

The components of these traction vectors form the Cauchy stress tensor in a given basis and are usually displayed in one matrix T (4) in which each row contains the components of each traction vector corresponding to each plane. The first subscript identifies the plane while the second the component. This tensor is symmetric for most of the materials, being an exception polar medium that have magnetic properties [19].

$$T = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \tag{4}$$

2.4 Strain and Deformation

The simplest case for the understanding of strain is the uniaxial tension test of metal or concrete test tubes. Strain is defined as the change in length per unit of initial length. Thus if a 100 cm test tube is stretched to 100.1 cm, the strain is 0.001% or 0.1%. With this we can characterize the tensional state. However, if the test tube is stretched first to 150 cm and then compressed back to 100.1 cm, the same statement about the strain would not describe what had happened. When plastic deformation occurs, history of straining needs to be monitored. With noninfinitesimal displacements for some materials, fracture may also happen, making invalid the formulation of small-strain theory. Infinitesimal strain theory gives good results for the practical purposes under study for strains around 1% and below.

Three simple 2D cases are presented in Fig. 2. The dotted lines show the initial positions of the element. The unit extensions $\boldsymbol{\epsilon}_x$ and $\boldsymbol{\epsilon}_y$ are defined as a change in length per unit initial length, and the shear strain is defined as half the decrease in the right angle initially formed by the lines parallel to the X and Y axes ($\boldsymbol{\gamma}_{xy} = \pi/2 - \psi = \theta_2 + \theta_1$). In the limit, when ΔX and ΔY tend to zero

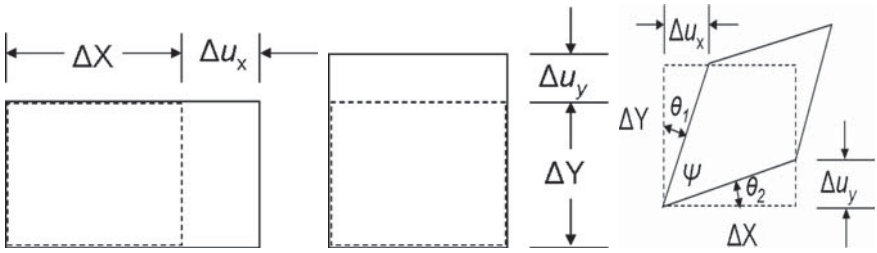


Fig. 2. Elementary cases of strain: (*Left*) uniaxial extension in x direction; (*center*) uniaxial extension in y direction; (*right*) pure shear without rotation. $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$

$$\epsilon_x = \frac{\partial u_x}{\partial X}, \qquad \epsilon_y = \frac{\partial u_y}{\partial Y}, \qquad \epsilon_{xy} = \frac{1}{2}\gamma_{xy} = \frac{1}{2}\left(\frac{\partial u_x}{\partial Y} + \frac{\partial u_y}{\partial X}\right), \tag{5}$$

where (u_x, u_y) are the particle displacement components in the X and Y directions. The angles θ_1 and θ_2 have been approximated by their tangents, as we are restricting this theory for small displacements and small angles compared to one radian. The $\frac{1}{2}$ factor appears when extracting the definitions from the displacement gradient matrix, which will be seen below.

The complexity of this problem is given by the fact that all three kinds of simple strain occur simultaneously and in three dimensions while there are some relative displacement that do not produce strain and stresses, for example, a rigid body transformation with rotation. The general problem is to express the strain in terms of the displacements by separating off that part of the displacement distribution which does not contribute to the strain. The results for the relative displacements analysis of a general three-dimensional case are presented here while a detailed study can be found in [13].

Let $\mathbf{u} = (u_x, u_y, u_z)$ be a 3D displacement vector field. In mechanical engineering it is well-known that the *displacement gradient matrix*, also known as *Jacobian matrix* or the *unit relative displacement matrix* (6), can be decomposed into the strain tensor (the symmetric part) with the elongational strains on the diagonal and the shearing strains on the off-diagonal, and the vorticity tensor (the antisymmetric part) that contains only the vorticity components [13],

$$\begin{bmatrix} \frac{\partial u_x}{\partial x} & \frac{\partial u_x}{\partial y} & \frac{\partial u_x}{\partial z} \\ \frac{\partial u_y}{\partial x} & \frac{\partial u_y}{\partial y} & \frac{\partial u_y}{\partial z} \\ \frac{\partial u_z}{\partial x} & \frac{\partial u_z}{\partial y} & \frac{\partial u_z}{\partial z} \end{bmatrix} = \begin{bmatrix} \frac{\partial u_x}{\partial x} & \frac{1}{2}\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x}\right) & \frac{1}{2}\left(\frac{\partial u_x}{\partial z} + \frac{\partial u_z}{\partial x}\right) \\ \frac{1}{2}\left(\frac{\partial u_y}{\partial x} + \frac{\partial u_x}{\partial y}\right) & \frac{\partial u_y}{\partial y} & \frac{1}{2}\left(\frac{\partial u_y}{\partial z} + \frac{\partial u_z}{\partial y}\right) \\ \frac{1}{2}\left(\frac{\partial u_z}{\partial x} + \frac{\partial u_x}{\partial z}\right) & \frac{1}{2}\left(\frac{\partial u_z}{\partial y} + \frac{\partial u_y}{\partial z}\right) & \frac{\partial u_z}{\partial z} \end{bmatrix} + \begin{bmatrix} 0 & \frac{1}{2}\left(\frac{\partial u_x}{\partial y} - \frac{\partial u_y}{\partial x}\right) & \frac{1}{2}\left(\frac{\partial u_x}{\partial z} - \frac{\partial u_z}{\partial x}\right) \\ \frac{1}{2}\left(\frac{\partial u_y}{\partial x} - \frac{\partial u_x}{\partial y}\right) & 0 & \frac{1}{2}\left(\frac{\partial u_y}{\partial z} - \frac{\partial u_z}{\partial y}\right) \\ \frac{1}{2}\left(\frac{\partial u_z}{\partial x} - \frac{\partial u_x}{\partial z}\right) & \frac{1}{2}\left(\frac{\partial u_z}{\partial y} - \frac{\partial u_y}{\partial z}\right) & 0 \end{bmatrix}. \tag{6}$$

For abbreviation we will represent this decomposition as $\boldsymbol{\nabla}\mathbf{u} = \boldsymbol{E} + \boldsymbol{\Omega}$. The tensor $\boldsymbol{E}$ measures the changes of shape locally (stretching or shortening), while $\boldsymbol{\Omega}$ informs about modifications related to rotations, and are defined by

$$\boldsymbol{E} = \begin{bmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{yx} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{zx} & \epsilon_{zy} & \epsilon_{zz} \end{bmatrix}, \text{ and } \boldsymbol{\Omega} = \begin{bmatrix} 0 & \omega_{xy} & \omega_{xz} \\ \omega_{yx} & 0 & \omega_{yz} \\ \omega_{zx} & \omega_{zy} & 0 \end{bmatrix}. \tag{7}$$

The columns in the strain matrix individually represent the unit relative displacement vectors, the first in the x-direction, the second in the y-direction, and the third in the z-direction of three infinitesimal elements, each of one initially parallel to x-direction, to y-direction, or to z-direction. From this, we can illustrate a physical interpretation of the shear components $\epsilon_{xy} = \epsilon_{yx}$, $\epsilon_{xz} = \epsilon_{zx}$ and $\epsilon_{yz} = \epsilon_{zx}$. When there is no rotation, ϵ_{xy} is equal to the

displacement in the y-direction of the two extremes of the infinitesimal element initially parallel to the x-direction divided by the length of that element. When there are both rotation and shear, the shear represents only part of the unit lateral relative displacement. Poisson's ratio is the ratio of lateral to axial strains: $\nu = \frac{-\epsilon_{yy}}{\epsilon_{xx}}$, or in a general way $\nu_{ij} = \frac{-\epsilon_{ii}}{\epsilon_{jj}}$, as it can be different for each direction. Its maximum value is 0.5, which denotes no volume change during the deformation.

When the strain matrix components are zero, the local motion of the material in the neighborhood of a point is an infinitesimal rigid-body motion, through a small angle (compared to one radian). Vorticity can be extracted from this rotation tensor and is related to the amount of circulation or rotation. More strictly, the local angular rate of rotation.

3 Methods

As it was mentioned in Sect. 1, most representations for elasticity imaging are scalar parameters, components of the strain tensor. Besides, tumor characterization have not been solved. With the same goals but different means than other investigations such as [7], which utilizes the discrepancy in sonograms and elastograms, and [9] through the shear strain visualization, we propose the study and analysis of the deformation tensor, and therefore, the strain tensor and the vorticity to asses tumor infiltration.

Tensor formulations are not widely extended in signal processing and related fields, but there are some cases in the biomedical field such as diffusion tensor imaging (DTI) and cardiac strain-rate imaging (SRI) that uses tensor visualizations. DTI visualization techniques are quite well developed as it can be inferred from first four chapters, where different models for DT-MRI are presented, and from next three chapters, where higher-level analysis of diffusion images are discussed. The application of these techniques to the visualization of the strain tensor fields is not obvious. The strain tensor is symmetric but does not satisfy the positive semidefinite condition. Hence, it is not possible to draw an ellipse with a negative value for one of its axis. However, in our case that condition is not a barrier, because clinically the sign of the strain tensor eigenvalues does not corrupt the information given with this representation regarding the different elastic properties of tissues (positive and negative represents, respectively, material stretching or shortening, in the direction of the corresponding eigenvector). If we represent the axes of the ellipse with the absolute values of the eigenvalues, we have the information of the rate of the deformation; it can be appreciated how much deformation from the total has been absorbed by the different tissues. Further explanation solving the representation of the eigenvalues' sign is found in the Sect. 4.3. Different tensor visualization methods can be found in this book in chapters by Hlawitschka et al., Feng et al., and Schreiner.

Two different cases of boundary condition between the inclusion (the simulated tumor) and its surrounding (the background) were studied: one of them with the inclusion loosely bounded to the background and the other fully connected, simulating, respectively, the benign and malign tumors. Different cases describing several possibilities at the boundaries of the region of interest (the homogeneous background and the inclusion, here and after the ROI) have been studied too, and we will mainly refer to two of them: symmetric and nonsymmetric boundary conditions. In both of them the movement at the lower horizontal boundary of the ROI is restricted in both directions x and y (2D case) and at the upper horizontal boundary perfect slip conditions are considered. In the symmetric case, the two vertical boundaries have no restrictions for x and y displacement. In the nonsymmetric case, one of these lateral boundaries has a restriction for the displacement on the x direction.

Experiments done through simulations made us observe what intuitively could be expected from an unbounded inclusion inside an homogeneous background axially compressed with nonsymmetric boundary conditions. As it can be seen in Fig. 3 the inclusion rotates, and this does not happen with perfectly symmetric boundary conditions. This simulation solves the problem with finite element analysis (FEA), but the algorithms currently used in elastography cannot appreciate that rotation in a displacement image is due to their limitation to detect small displacements. However, parameters such as vorticity can give information about the differential rotation in the areas or volumes of study.

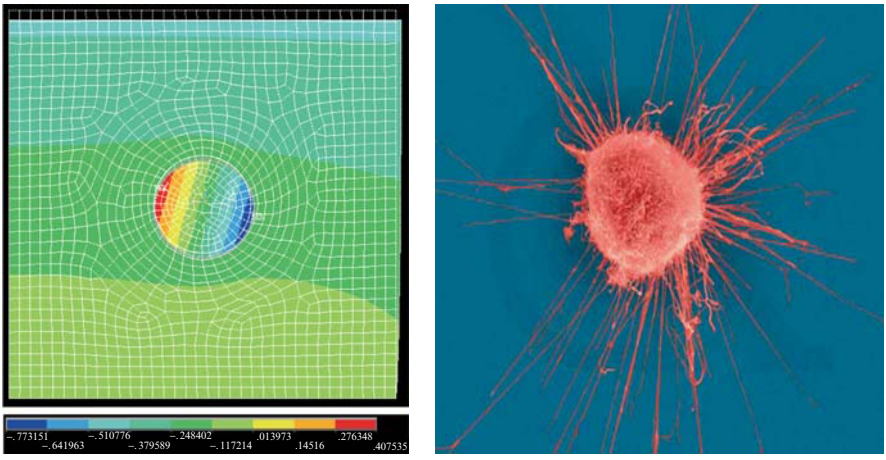


Fig. 3. *Left:* axial displacement image corresponding to an unbounded inclusion three times stiffer than the background with nonsymmetric boundary conditions. *Right:* a malign tumor (source [1]). It can be seen how much it is ramified while a benign one would be smooth allowing its rotation

A main contribution of this work is the hypothesis of infiltrating tumors being less prone to rotate than noninfiltrating ones when the surrounding tissue is pressed, since the former are more tightly attached to the normal tissue than the latter. Hence, we expect the vorticity to be zero at the background and nonzero values for the infiltrating tumors, not as with noninfiltrating ones, which should have also values around zero. We also expect the strain tensor to behave differently for this two cases, especially at the boundaries of the tumor. This difference can be assessed in the visualizations as we show next.

4 Experimental Set-up and Results

We have developed an experimental setting that has proved to be very useful to obtain preliminary results, prior to a clinical experimentation. The setting consists of a synthesis stage, where a digital phantom of the tumor anatomy with surrounding tissue, the effect of the compression applied with the ultrasonic probe, and the US images are simulated. This setting allows us to easily experiment under different conditions and mechanical characteristics that we expect to correlate with real clinical data. The analysis stage estimates the deformation and elastic properties from the US data and provides useful visualizations.

4.1 Synthesis

Tumors can be simulated digitally [25] by an homogeneous background (the surrounding normal tissue) and a centered circular inclusion (the tumor) using FEA software. Mechanical properties are endowed to both the background and the inclusion. The level of infiltration is simulated by defining contact points and assigning to them a friction coefficient. For the specific setting reported in this work the mechanical properties are as follows: for both inclusion and background, the Poisson's ratio used was $\nu = 0.495$, [11] thus assuming conditions of virtual incompressibility. The background has a modulus of 21 kPa [11], and the inclusion was three times stiffer.

The FE models were generated with applications to elastography in mind with an area of $40 \times 40\,\mathrm{mm}^2$ for the background, and 5 mm radius for the circular inclusion. The phantom was meshed using eight noded quadrilateral elements. Two different cases of boundary condition between the inclusion and the surrounding were studied (simulating the benign and malign tumors), and two cases studying the external interactions of the ROI as described in Sect. 3. We have been carrying out extensive experiments, with different mechanical properties, geometries, and boundary conditions in order to simulate many real situations, which we cannot report in this work due to space limitations.

A compression was applied to the entire top side of the digital phantom, being 1% of its height. The top surface allow perfect slip conditions, but the

bottom surface has movement restriction in both directions (axial and lateral). The sides of the target are free to move.

The coordinates of the nodes corresponding to the meshes of the test object before and after the compression are the outputs from the FEA program, and are considered as ground-truth data. These two sets of (x, y) coordinates were then computed with ultrasound simulation codes obtaining 15 frames of each, containing the A-lines (amplitude echoes of the acoustic waves) for the ROI. Comparing each pair of frames (pre and post) we estimated 15 displacement fields for our phantom. Averaging them all we produce elastograms as in the usual scalar way of visualization (top of Fig. 5) where the white areas represent less strain and hence, a stiffer region.

4.2 Analysis

Several techniques have been proposed to estimate the tensor strain components (Sect. 1). All of them estimate a displacement vector field due to an applied deformation load. In this work, we are not discussing which one is better, but how we can obtain more useful information from the displacement vector field. We computed the cross-correlation of the pre and post-compression A-lines in order to obtain that displacement field. Further information on this technique can be found in [10, 27]. From the estimated displacement field (the vector field), the gradient is calculated (6), after which we decompose it in the strain and the vorticity tensor fields. In Fig. 4, the vector field at each point is overlaid on the strain tensor with different boundary conditions to show the potentiality of this visualization and the information behind it. Below we have plotted the strain tensor field for a 3D case. This case is not the same as the 2D one and different solutions might be adopted. As there are three eigenvectors and eigenvalues for each voxel instead of two, it becomes more complex and the sign of the eigenvalues must be considered in a different way. More information on this issue can be found in [24].

The strain tensor field is noisy and susceptible to be filtered. [21] proposed a noniterative anisotropic method to regularize vector and higher order tensor fields having in mind DT-MRI applications. We extended its use to the strain tensor field, obtaining a more aligned and anisotropically averaged tensor field, while the edges are preserved. Results are shown in Fig. 8.

4.3 Visualization

Although elastography is based on the estimation of tensor data (the strain tensor), its visualization has been traditionally based on scalar representations. Some researchers represent the axial strain or other related parameters in gray scale, the darker areas representing stiffer regions (less strain) and some others in color images. The gray scale has been the standard for ultrasound imaging in general and also for elastography. However, the color images

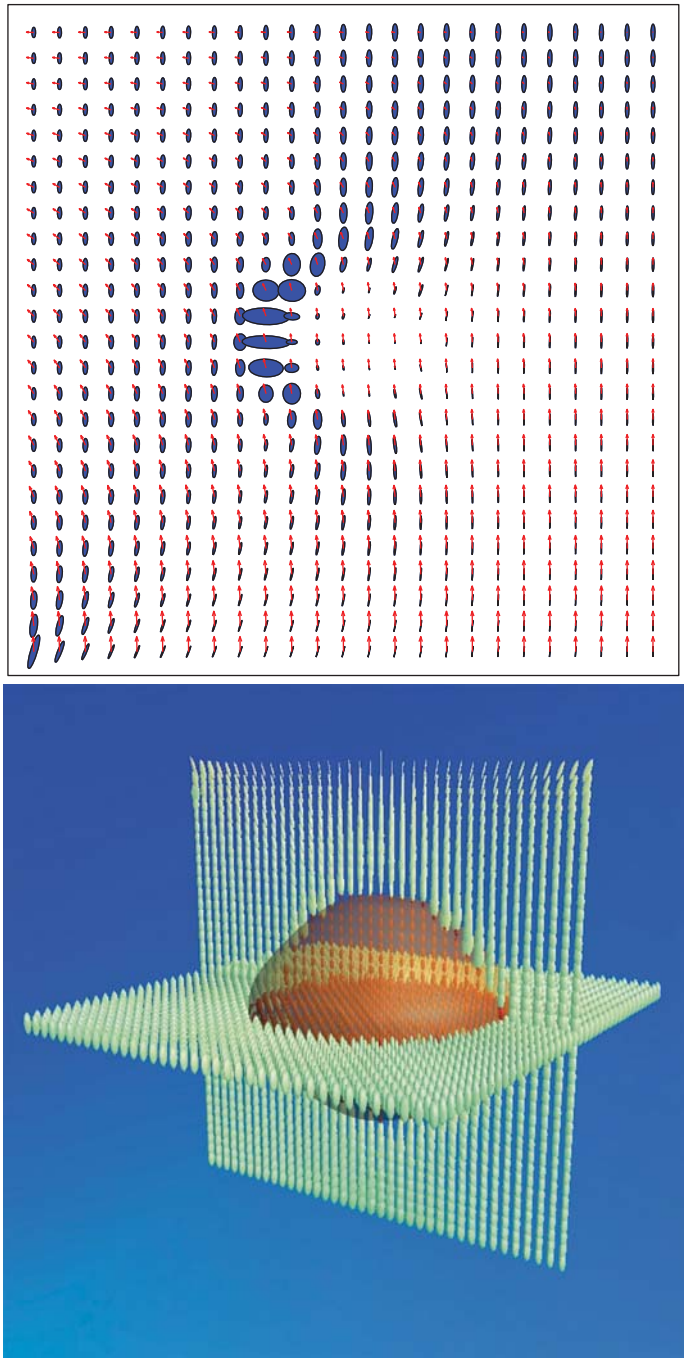


Fig. 4. Displacement vector field overlaid on the strain tensor field for a virtual phantom, ideal case. *Above*: unbounded case (benign) with nonsymmetric boundary conditions. *Below*: a 3D case with symmetric boundary conditions

for the elastograms are becoming widely used. This is due to the fact that elastograms are usually represented beside the B-mode image or overimposed to it. Therefore, it is easy to appreciate the colored image in contrast to the gray scale one. In this chapter both of them are used for the reader to have the reference and the possibility to choose. In any case the clinician should decide which image he or she prefers, being our mission to give them different tools and possibilities.

It is important to know that the visualization of tensor fields improve the understanding and interpretation of tensor data. In the last years, tensor fields visualization has achieved great interest thanks to improvement of graphics hardware, and the advances in nervous fiber visualization given by DT-MRI dataset.

It is proposed here a technique similar to that in [22], used for the visualization of myocardial strain-rate tensors, that is, visualize tensors as ellipsoids colored according to the sign of the largest eigenvalue representing stretch or compression in the principle direction. We have colored blue for shortening and red for stretching, when displaying the tensorial image alone or overlaid on a gray scale image. For the cases where the tensorial image is displayed over a color image, the ellipses are drawn in black for shortening, in order to distinguish from the background colors. Obviously, this is just a convention that may be adapted to the different circumstances to achieve a better visualization.

In our case, as it can be seen in Figs. 5–8, all the ellipses appear in blue. As we are applying a compression in the axial direction to a incompressible target, the Poisson's effect distribute that deformation expanding the target in both lateral directions. Therefore, the axes of the ellipses aligned with the axial direction, where the compression is applied, have higher values. This representation visualizes the complete tensor in one plot giving information about magnitude, direction, and the ratio between the strain principle values. This representation can detect possible abnormalities and regions with different local behavior than the surrounding. Different in vivo experiments are planned to validate the results.

When visualizing the vorticity, we have represented the scalar value of

$$\omega_z \equiv 2 \cdot \omega_{xy} = -2 \cdot \omega_{yx}, \tag{8}$$

which represent the infinitesimal angle of the pure rotation experimented locally (considering the movement a compound of elongational terms and rotational ones, as seen in Sect. 2, (7)) [26].

Another experiment was carried out imaging a commercial breast elastography phantom (Model 059, CIRS Inc., Virginia, USA) to a depth of 4 cm with a linear array of 128 elements with a 12-MHz centroid frequency, 80% fractional bandwidth transducer, of a PC-based Ultrasonix 500RP ultrasound machine (Ultrasonix Medical Corporation, Burnaby, BC, Canada). The experiment was realized applying a 2% freehand compression to the breast phantom

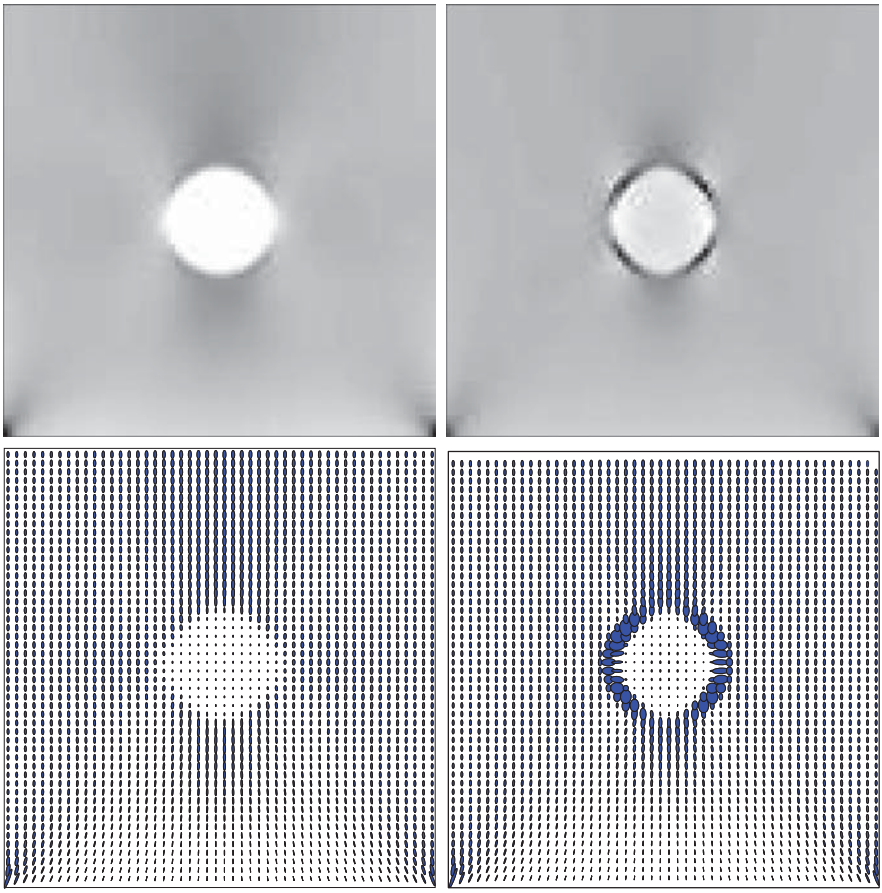


Fig. 5. Scalar and tensorial images comparison for a virtual phantom, ideal case. *Top*: axial elastograms; *bottom*: tensor elastograms. *Left*: bounded case (malign); *right*: unbounded case (benign)

with the linear probe. The phantom contained lesions three times stiffer than the background. By so they can be detected on elastograms. The result is shown in Fig. 7

5 Discussion

The decomposition of the *displacement gradient matrix* in the strain tensor and the vorticity yields the tensor visualization of the first and the scalar representation of the second. The color-coded ellipses representation of the strain tensor has a big potential in studying the elastic behavior of tissues. The experiments presented in this work show a homogeneous axially compressed

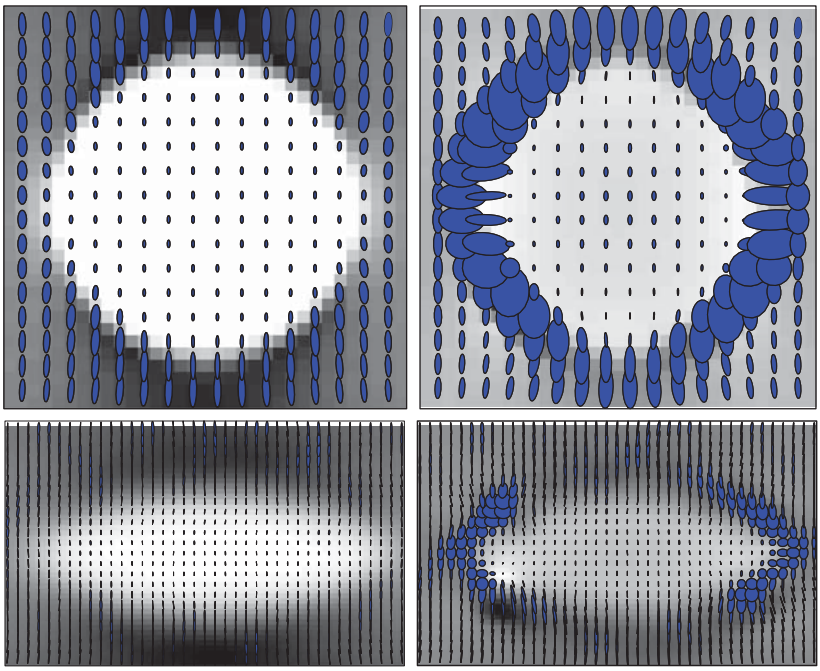


Fig. 6. Tensorial images overlaid on the axial strain. Comparison of the virtual phantom between the ideal case and the US simulation. *Top*: Ideal case; *bottom*: US simulation; *Left*: bounded case (malign); *right*: unbounded case (benign). Note that the tensor field dimensions are different between the ideal case and the US simulated phantoms because several ultrasonic beams cover the areas corresponding to the nodes of the FE, resulting in the elliptical shape of the inclusions

target with symmetrical boundary conditions, where the color coding of the ellipses seems to be trivial, but other settings such as antisymmetric boundary conditions may show its potential utility. Beside the color-coded information about the stretching or shortening of the tissue locally, the ellipses present the principal directions of deformation, their magnitude, and the ratio between these magnitudes. While the scalar representations show the deformation in axial or lateral directions, the ellipses show the magnitude and also the direction in which the tissue is deformed the most at that point. Many applications in different elastographic modalities such as intravascular elastography may be favored by the tensorial representation, for example, detecting calcification in the arteries [14]. Further investigation has to be done to find other ways to visualize the strain tensor in elastography suitable to handle both the information and the utility for the physicians, extracting information clinically useful for the diagnosis and prognosis of diseases such as breast and prostate cancer.

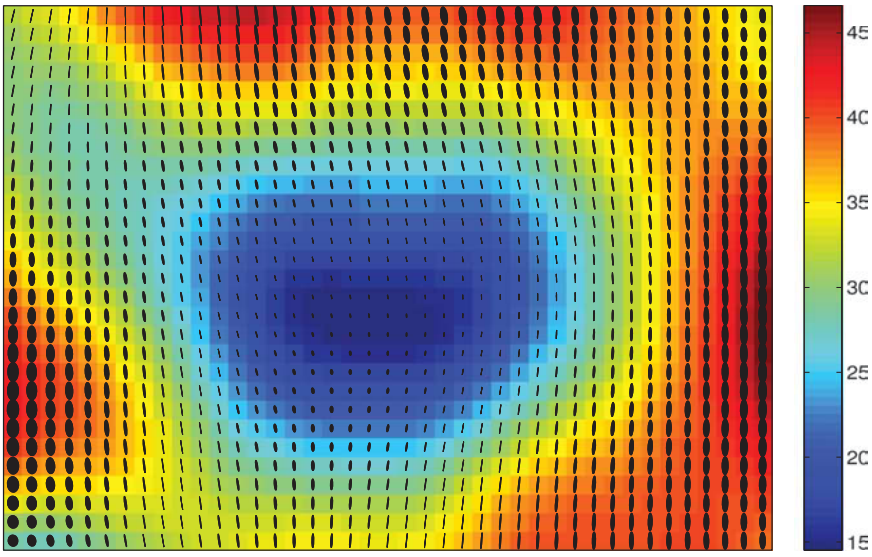


Fig. 7. Tensorial image overlaid on the axial strain mapped to colors for commercial CIRS phantom (tissue-mimicking commercial breast). The inclusion appears in blue, less axial strain meaning a harder tissue. In this area the ellipses (painted in black for compression) are smaller than at the surrounding material

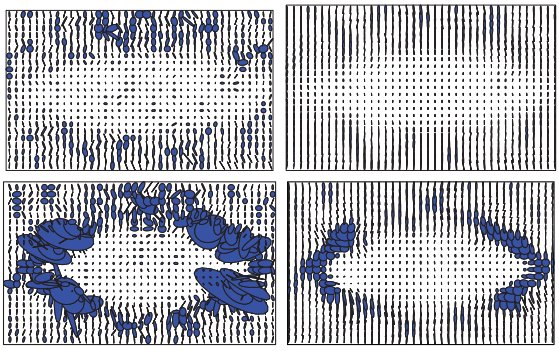


Fig. 8. Tensorial images filtering. *Top*: bounded case (malign); *bottom*: unbounded case (benign). *Left*: noisy data (US simulation); *right*: anisotropically filtered data after regularization

Although there are studies [7, 9] showing differences between malignant and benign tumors appreciated in axial strain and shear strain elastograms, the vorticity isolates the information about rotation in the deformation tensor and might contribute in the diagnosis and prognosis of in vivo cancers (e.g., breast and prostate), what have not been assessed, as far as we know, in other studies in elastography.

5.1 Real and Estimated Vorticity

In Fig. 9, we have included a representation of the ideal vorticity, meaning that we have computed the output of the FEA software without simulating the ultrasound. This image shows a ring similar to those found in shear elastograms [9, 29], presenting diagonal symmetry too. In our case, black or white represent clockwise or anti-clockwise rotation. This effect is due to the behavior of the material around the inclusion. As the background absorbs more relative displacement when compressed (softer material), it seems intuitive to consider that it will slip around the inclusion; clockwise at the right and anti-clockwise at the left, as if the inclusion stayed at the same position. However, at the lower part of the inclusion the rotation is opposed to the former. The Poisson's effect describes the lateral displacements due to the axial deformation. For this case of symmetric boundary conditions, the lateral stretching of the background to both sides of the target makes it easy to understand the direction of the rotations as they appear in the Fig. 10.

The vorticity image of the US simulated response of the phantom yields a similar result, this time with the inherent noise of the US. Diagonal symmetry can also be appreciated; the direction of the rotation can be assessed with predominant positive or negative values in each quadrant of the ring at the boundary of the inclusion.

5.2 Effect of Boundary Conditions

As related in Sect. 3, part of this research was motivated by the results in the displacement images such as Fig. 3, with nonsymmetric boundary conditions that make the inclusion rotate when it is not bounded to the background. When these boundary conditions are symmetric, in theory, the inclusion would

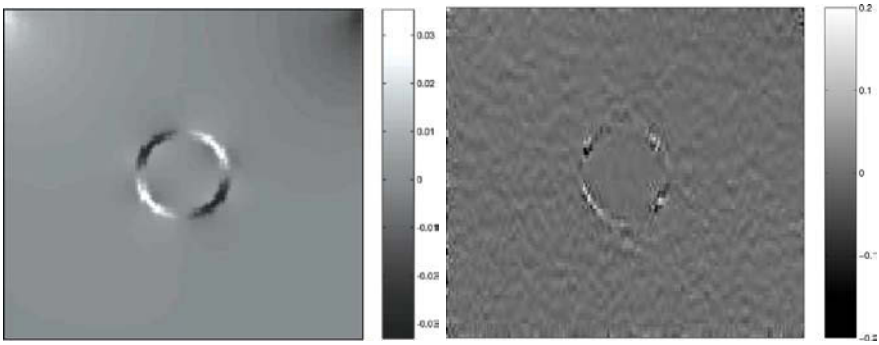


Fig. 9. Vorticity images for the unbounded case (benign) with symmetric boundary conditions. *Left*: ideal image. *right*: and the same image for the ultrasound simulation. One may observe the ring that also appears for the axial or the shear strain images, but this time isolated from the rest of the parameters

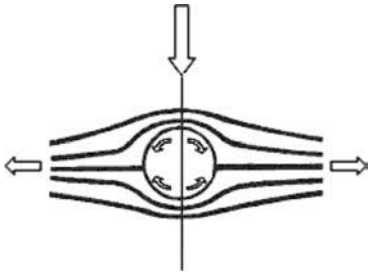


Fig. 10. Rotation scheme. The circle represents the inclusion, and the lines are analogous to a flow, in our case the stretching of the background due to Poisson's effect. The arrows inside the circle show the directions of the background's rotation with respect to the inclusion

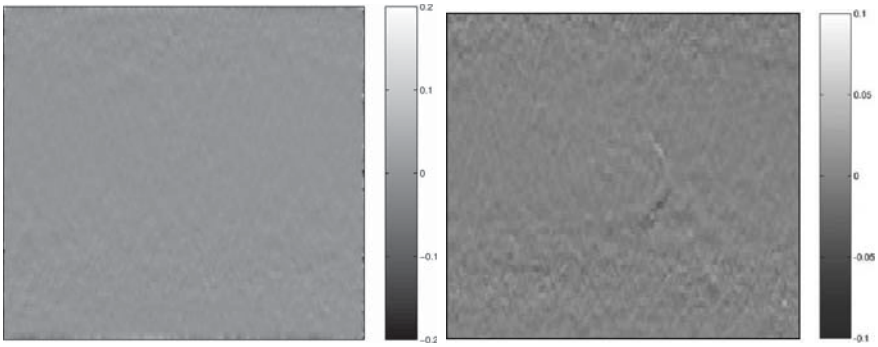


Fig. 11. Vorticity images. *Left*: bounded case; *right*: unbounded case with nonsymmetric boundary conditions. For the bounded case the vorticity is homogeneous (for both the inclusion and the background). For the unbounded case with nonsymmetric boundary conditions, it only appears as ring at the side where the target is free to move

not rotate due to the lateral equilibrium of the force system. However, vorticity keeps showing information, due to its infinitesimal considerations exposed in Sect. 2.

Comparing the Figs. 9 (right) and 11 (right), we can explain how the boundary conditions affect the vorticity images and other elastographic representations. The former figure has symmetric boundary conditions, being the target free to move in both laterals. On the contrary, at the latter image, at the left side of the target we have imposed lateral movement equal to zero. The applied compression is pure axial and symmetric, but this time, as the target cannot stretch to the left, the whole lateral deformation occurs towards the right. This yields, as it can be seen in Fig. 11 (right), higher values of vorticity on the right of the inclusion, whereas the diagonal symmetry encountered in the former image has disappeared.

5.3 Effect of Noise

We have introduced in the phantom study the effect of the noise produced by the acquisition equipment. The figures presented are the output of the simulated US response of the target after being processed with the elastography algorithm. We added to the US simulations random noise. Results are accurate enough to detect variations in the tensor representation as well as in vorticity, when the contrast ratio (9) between the inclusion and the background is 3 or higher. The effect of the contrast ratio is discussed in the next point. Regarding noise, in vivo elastograms are expected to behave similar to those presented here. Studies about the influence of noise in the detection of lesions using elastography can be found in [3].

$$\text{Contrast Ratio } = \frac{\text{Elastic Modulus of the Inclusion}}{\text{Elastic Modulus of the Background}}. \tag{9}$$

5.4 Effect of Contrast Ratio

We have executed the experiment for three different contrast ratios: 1.25, 2, and 3 (Fig. 12). The bounded case shows no differences for the three cases, and the visualization result is that shown in Fig. 9 (top left); an homogeneous image. In Fig. 13, it can be noticed that vorticity images with a contrast ratio lower than 3 are much noisier, making it difficult to distinguish between different rotational behaviors. However, as we present in Fig. 14, vorticity

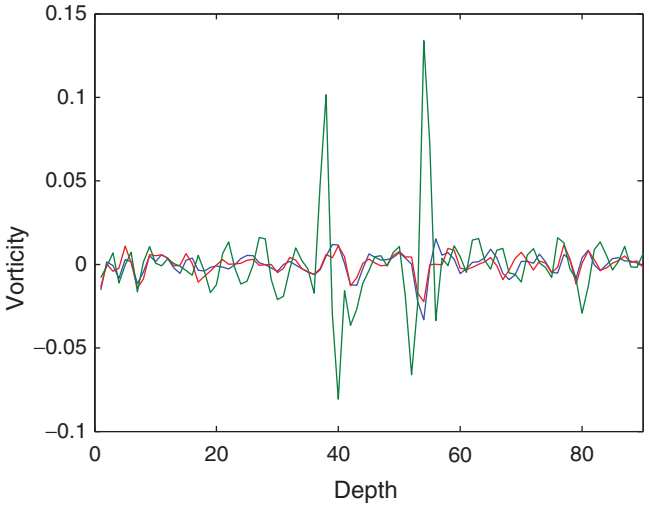


Fig. 12. Vorticity against depth (in pixels) for the same representative column (passing through the inclusion at its right) for the three different contrast ratios between the inclusion and the background. *Blue*: contrast ratio of 1.25; *red*: contrast ratio of 2; *green*: contrast ratio of 3

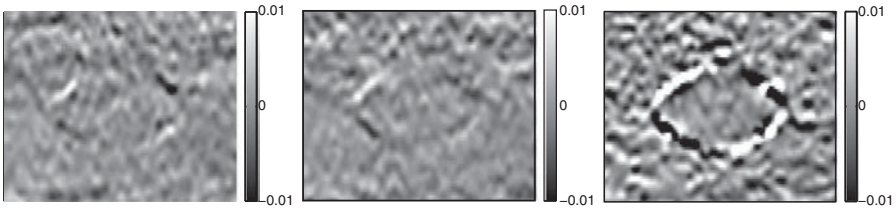


Fig. 13. Vorticity images for different contrast ratios between the inclusion and the background. *Left*: contrast ratio of 1.25; *center*: contrast ratio of 2; *right*: contrast ratio of 3. Note the different scales of the images in order to visualize each case better

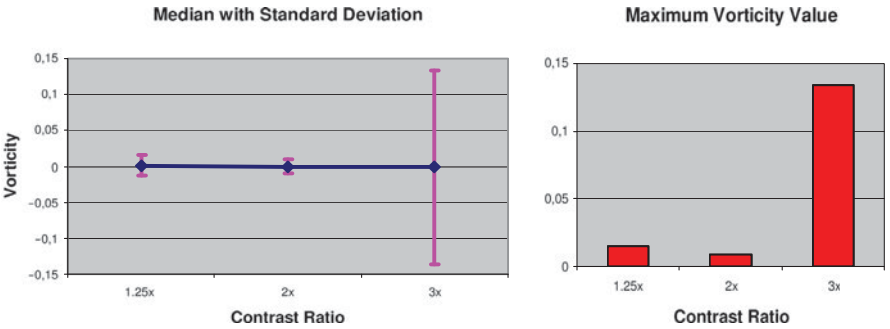


Fig. 14. Vorticity's median with its standard deviation and maximum value for the three different contrast ratios between the inclusion and the background along one representative column of the target

Table 1. One-column vorticity statistics

Contrast Ratio	1.25x	2x	3x
Mean	−0,0002	−0,0003	0,0002
Median	0,0009	0,0000	−0,0009
Deviation from Median	0,0144	0,0091	0,1348
Maximum Value	0,0153	0,0091	0,1339

has higher values at the boundary of the unbounded inclusion for the three contrast ratios. This figure shows the statistical values of the vorticity in one representative column of the image matrix, the same column for the three cases (Summarized in Table 1). The column passes through the inclusion where vorticity holds higher values at the boundary with the background. In the case where the inclusion is three times stiffer than the background, the deviation from the median, and therefore the difference from the background, is one order higher than in the other two cases. Appropriate processing of the image might yield results where this parameter can assess rotations at the target.

6 Conclusion and Ongoing Research

Other works visualize scalar values such as axial strain, shear strain, and Poisson's ratio, and have shown the usefulness of the information contained in them to assess the mobility of the tumor and therefore its malignancy [7, 9]. As the strain tensor integrates these parameters, its representation may provide new criterions and additional information that complements the scalar images. There might be some cases where the scalar images do not give enough information for the characterization of the tumor due to the complexity of this issue, which motivates our research. Vorticity isolates the information about rotation in the deformation tensor and may also contribute to the diagnosis and prognosis of in vivo cancers (e.g. breast and prostate), what have not been assessed, as far as we know, in other elastography studies.

Our on-going research deals with the use of new algorithms to estimate accurate strain and vorticity tensor fields and their filtering. Commercial and gelatin phantoms and clinical validation are under study, as well as other tensorial data representations that can be extracted from the mathematical approach presented here.

Acknowledgments

Funding was provided by the Spanish Ministry of Science and Technology (TEC-2004-06647-C03-02), the European NoE SIMILAR FP6-507609 and for the second and third author, cofunded by MEC and Social European Funds, (*Torres Quevedo* PTQ2004-1443 and PTQ2004-1444, respectively).

References

[1] Programme in medicinal chemistry. university of cambridge.
[2] P.E. Barbone and N.H. Gokhale, Elastic modulus imaging: on the uniqueness and nonuniqueness of the elastography inverse problem in two dimensions. *Institute of Physics Publishing*, 20:283–296, 2004.
[3] N. Belaid, I. Céspedes, J. Thijssen, and J Ophir, Lesion detection in simulated elastographic and ecographic images: A psycho-physical study. *Ultrasound in Medicine and Biology*, 20:877–891, 1994.
[4] M.M. Doyley, J.C. Bamber, F. Fuechsel, and N.L. Bush, A freehand elastographic imaging approach for clinical breast imaging: system development and performance evaluation. *Ultrasound in Medicine and Biology*, 27:1347–1357, 2001.
[5] M.M. Doyley, S. Srinivasan, S.A. Pendergrass, Z. Wu, and J. Ophir, Comparative evaluation of strain-based and model-based modulus elastography. *Ultrasound in Medicine and Biology*, 31(6):787–802, 2005.

[6] D.D. Duncan and S.J. Kirkpatrick, Processing algorithms for tracking speckle shifts in optical elastography of biological tissues. *Journl of Biomedical Optics*, 6(4):418–426, 2001.

[7] B.S. Garra, I. Céspedes, J. Ophir, S. Spratt, R.A. Zuurbier, C.M. Magnant, and M.F. Pennanen, Elastography of breast lesions: initial clinical results. *Radiology*, 202:79–86, 1997.

[8] K.M. Hiltawsky, M. Kruger, C. Starke, L. Heuser, H. Ermert, and A. Jensen, Freehand ultrasound elastography of breast lesions: clinical results. *Ultrasound in Medicine and Biology*, 27:1461–1469, 2001.

[9] E. Konofagou, T. Harrigan, and J. Ophir, Shear strain estimation and lesion mobility assessment in elastography. *Ultrasonics*, 38:400–404, 2000.

[10] E.E. Konofagou and J. Ophir, A new elastographic method for estimation and imaging of lateral displacements, lateral strains, corrected axial strains and poisson's ratios in tissues. *Ultrasound in Medicine and Biology*, 24(8):1183–1199, 1998.

[11] T.A. Krouskop, T.M. Wheeler, F. Kallel, B.S. Garra, and T. Hall, Elastic moduli of breast and prostate tissue under compression. *Ultrasonic Imaging*, 20:260–274, 1998.

[12] R.M. Lerner and K.J. Parker, Sonoelasticity imaging. In *Proceedings of the 16th International Acoustical Imaging Symposium (Plenum)*, volume 16, The Netherlands. Luxembourg, pp. 317–327, 1998.

[13] L.E. Malvern, *Introduction to the Mechanics of a Continuous Medium*. Prentice Hall, Englewood, NJ, 1969.

[14] R.L. Maurice, M. Daronat, J. Ohayon, E. Stoyanova1, F.S. Foster, and G. Cloutier, Non-invasive high-frequency vascular ultrasound elastography. *Physics in Medicine and Biology*, 50:1611–1628, 2005.

[15] J. Ophir, I. Céspedes, B. Garra, H. Ponnekanti, Y. Huang, and N. Maklad, Elastography: a quantitative method for imaging the elasticity of biological tissues. *Ultrasound Imaging*, 13:111–134, 1991.

[16] T. Otake, T. Kawano, T. Sugiyama, T. Mitake, and S. Umemura. High-quality/high-resolution digital ultrasound diagnostic scanner. *Hitachi Review*, 52(4), 2003.

[17] K.J. Parker, L.S. Taylor, and S. Gracewski, A unified view of imaging the elastic properties of tissue. *Acoustical Society of America*, 117(5):2705–2712, 2005.

[18] A. Pesavento, C. Perrey, M. Krueger, and H. Ermert, A time-efficient and accurate strain estimation concept for ultrasonic elastography using iterative phase zero estimation. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 46(5):1057–1067, 1999.

[19] E. Reissner, Note on the theorem of the symmetry of the stress tensor. *Journal of Mathematics and Physics*, 25, 1946.

[20] R. Righetti, J. Ophir, S. Srinivasan, and T. Krouskop, The feasibility of using elastography for imaging the poisson's ratio in porous media. *Ultrasound in Medicine and Biology*, 30(2):215–228, 2004.

[21] M.A. Rodriguez-Florido, C.-F. Westin, and J. Ruiz-Alzola, DT-MRI regularization using anisotroipic tensor field filtering. *IEEE International Symposium on Biomedical Imaging*, pp. 15–18, April 2004.

[22] P. Selskog, E. Heiberg, T. Ebbers, L. Wigstrom, and M. Karlsson, Kinematics of the heart: strain-rate imaging from time-resolved three-dimensional phase contrast mri. *IEEE Trans Med Imaging*, 21(9):1105–1109, 2002.

[23] M. Shling, M. Arigovindan, C. Jansen, P. Hunziker, and M. Unser, Myocardial motion analysis from b-mode echocardiograms. *IEEE Transactions on Image Processing*, 14(4), April 2005.

[24] D. Sosa-Cabrera, *Novel Processing Schemes and Visualization Methods for Elasticity Imaging*. PhD Dissertation, University of Las Palmas de Gran Canaria, 2008.

[25] D. Sosa-Cabrera, J. Ophir, T. Krouskop, A. Thitai-Kumar, and J Ruiz-Alzola, Study of the effect of boundary conditions and inclusion's position on the contrast transfer efficiency in elastography. In *IV International Conference on the Ultrasonic Measurement and Imaging of Tissue Elasticity*, p. 94, October 2005.

[26] D. Sosa-Cabrera, M.A. Rodriguez-Florido, E. Suarez-Santana, and J. Ruiz-Alzola, Vorticity visualization: Phantom study for a new discriminant parameter in elastography. In *Proceedings of SPIE in Medical Imaging*, volume 6513, 2007.

[27] S. Srinivasan, F. Kallel, R. Souchon, and J. Ophir, Analysis of an adaptive strain estimation technique in elastography. *Ultrasonic Imaging*, 24:109–118, 2002.

[28] S. Srinivasan, T. Krouskop, and J. Ophir, Comparing elastographic strain images with modulus images obtained using nano-indentation: Preliminary results using phantoms and tissue samples. *Ultrasound in Medicine and Biology*, 30(3):329–343, 2004.

[29] U. Techavipoo, Q. Chen, T. Varghese, and J. A. Zagzebski, Estimation of displacement vectors and strain tensors in elastography using angular insonifications. *IEEE Transactions on Medical Imaging*, 23(12), December 2004.

A Higher-Order Structure Tensor

Thomas Schultz[1], Joachim Weickert[2], and Hans-Peter Seidel[1]

[1] MPI Informatik, Campus E1 4, 66123 Saarbrücken, Germany
{schultz, hpseidel}@mpi-inf.mpg.de
[2] Mathematical Image Analysis Group, Faculty of Mathematics and Computer Science, Saarland University, 66041 Saarbrücken, Germany
weickert@mia.uni-saarland.de

Summary. Structure tensors are a common tool for orientation estimation in image processing and computer vision. We present a generalization of the traditional second-order model to a higher-order structure tensor (HOST), which is able to model more than one significant orientation, as found in corners, junctions, and multichannel images. We provide a theoretical analysis and a number of mathematical tools that facilitate practical use of the HOST, visualize it using a novel glyph for higher-order tensors, and demonstrate how it can be applied in an improved integrated edge, corner, and junction detector.

1 Introduction

The second-order structure tensor, formed as the outer product of the image gradient with itself, is a common tool for local orientation estimation. Since it was first introduced for edge and corner detection [1], it has been applied to a wide variety of problems in image processing and computer vision, including optic flow estimation [2], image diffusion [3], texture segmentation [4], image inpainting [5], and image compression [6].

Previous extensions include a generalization to vector- and tensor-valued images, which goes back to an idea of Di Zenzo [7], a modification to detect spiraling, cross-like, and parabolic shapes [8], and the introduction of nonlinear local averaging [9], which led to nonlinear structure tensors [10].

It is a known limitation of the traditional structure tensor that it can only represent a single dominant orientation. Recently, there have been attempts to overcome this: Arseneau and Cooperstock [11] have placed second-order structure tensors in discrete directional bins and derived parameters of multimodal *directional distribution functions* from them. Their work concentrates on lifting the constraint of antipodal symmetry (i.e., they treat direction $\mathbf{v}$ differently than direction $-\mathbf{v}$), a property that our approach preserves. Moreover, they use the structure tensors only as an intermediate representation, finally reducing them to two scalar parameters for each direction.

Herberthson et al. [12] have used outer products to handle pairs of orientations. However, their approach is specific to the case of two orientations: It neither generalizes to more than two directions, nor does it indicate cases in which representing a single orientation is sufficient.

In our present work, we present a generalization of the second-order structure tensor to a higher-order tensor model, which is able to capture the orientations of more complex neighborhoods, for example, corners, junctions, and multivalued images. The tensor order allows to specify the maximum complexity the structure tensor can represent and can be chosen based on the requirements of a given application.

This chapter is structured as follows: Section 2 introduces our new higher-order structure tensor (HOST). In Sect. 3, we present a novel glyph for higher-order tensors and use it to visualize first experimental results. A theoretical analysis and a number of mathematical tools that help to use the HOST in practice are presented in Sect. 4. They include an efficient representation of the structure tensor, an alternative representation as a truncated Fourier Series, a generalization of the matrix trace and the eigenvector decomposition, and an algorithm to extract contrast extrema from a higher-order tensor representation. Section 5 shows a proof-of-concept application, in which the HOST is used for an improved integrated edge and junction detection. Finally, Sect. 6 concludes this chapter and points out directions of future research.

2 A Higher-Order Structure Tensor

The standard second-order structure tensor $\mathbf{J}$ is given by the outer product of the image gradient ∇f with itself [1]:

$$\mathbf{J} := \nabla f \nabla f^T. \tag{1}$$

The structure tensor representation is independent of the sign of ∇f. Thus, it can be averaged over a neighborhood σ without canceling gradients that have the same direction, but opposite orientation. This spatially averaged matrix $\mathbf{J}_\sigma$ describes local image structure. Its principal eigenvector indicates the direction of largest contrast. However, the matrix representation is insufficient if there is more than one dominant direction in a neighborhood. For example, a structure tensor $\mathbf{J}_\sigma$ which describes two orthogonal, equally strong directions will have two equal eigenvalues and no longer indicate a principal direction.

This effect is avoided by a higher-order structure tensor $\mathcal{J}$, formed by repeating the outer product. Taking the outer product of a vector $\mathbf{v}$ with itself l times will be written $\mathbf{v}^{\otimes l}$. It yields an order-l tensor, indexed by $\{i_1, i_2, \ldots, i_l\}$:

$$\left(\mathbf{v}^{\otimes l}\right)_{i_1 i_2 \ldots i_l} := v_{i_1} \cdot v_{i_2} \cdots v_{i_l}. \tag{2}$$

To ensure antipodal symmetry of the resulting tensor, l is chosen to be even.

As an alternative to analyzing its eigenvectors, a structure tensor $\mathcal{J}$ can be interpreted through its induced homogeneous form $J(\mathbf{u})$, which specifies the local contrast in a given direction $\mathbf{u}$ and will thus be referred to as a *contrast function*. When $\mathbf{u}$ is a unit vector, $J(\mathbf{u})$ is defined by repeating the inner tensor–vector product of $\mathcal{J}$ and $\mathbf{u}$ until a scalar is left, that is, l times. In n dimensions, this can be written as

$$J(\mathbf{u}) := \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_l=1}^{n} (\mathcal{J})_{i_1 i_2 \ldots i_l} u_{i_1} u_{i_2} \cdots u_{i_l}. \tag{3}$$

For a second-order structure tensor, J is unimodal, which reflects the fact that it is suitable to model only one dominant direction. For higher orders, J can become multimodal, which allows for a more accurate representation of corners, junctions, and multivalued images.

We consider it a sensible requirement that the values of the contrast function should remain comparable, independent of the tensor order that we use. When evaluated in direction of the gradient, the contrast function yields the squared gradient magnitude in the second-order case. However, taking the outer product l times would raise the gradient magnitude to the lth power. We compensate this by scaling the gradient vector beforehand. An order-l structure tensor $\mathcal{J}$ that reduces to the well-known second-order tensor $\mathbf{J}$ for $l = 2$ is then given by

$$\mathcal{J} := \left(\frac{\nabla f}{|\nabla f|^{\frac{l-2}{l}}} \right)^{\otimes l}. \tag{4}$$

In some applications, it is beneficial to have a contrast function that gives the nonsquared gradient magnitude [13]. This can be achieved by replacing the exponent $\frac{l-2}{l}$ by $\frac{l-1}{l}$ in (4).

3 Glyphs for Higher-Order Tensors

The visualization of higher-order tensors has previously been addressed in the context of generalized diffusion tensor magnetic resonance imaging (DT-MRI) [14], where the induced homogeneous form describes a directionally dependent diffusivity coefficient. In this context, *generalized Reynolds glyphs* are so far the only glyph-based visualization technique [14, 15]. Let S be the unit sphere and J the homogeneous form as defined in (3). Then, these glyphs are formed by the set of points $\{J(\mathbf{u})\mathbf{u} \,|\, \mathbf{u} \in S\}$, which directly depicts the contrast profile of the tensor. However, these glyphs have a round shape around their maxima, which makes their exact direction difficult to see. To compensate this problem, Hlawitschka and Scheuermann [15] suggest to add arrows that point to the maxima.

3.1 Higher-Order Tensor Glyphs with Maximum Enhancement

The diffusion ellipsoid is accepted as the standard glyph for second-order diffusion tensors, but the Reynolds glyph does *not* reduce to it for $l = 2$. Since the tensor ellipsoid can be constructed by transforming the unit sphere under the linear mapping induced by the tensor, it is natural to generalize it by taking the inner tensor–vector product $l - 1$ times, until a vector is left. We denote the inner product $\mathcal{J} \bullet \mathbf{u}$, where

$$(\mathcal{J} \bullet \mathbf{u})_{i_1 i_2 \ldots i_{l-1}} := \sum_{i_l=1}^{n} (\mathcal{J})_{i_1 i_2 \ldots i_l} u_{i_l} \tag{5}$$

and use the shortcut notation $\mathcal{J} \overset{l-1}{\bullet} \mathbf{u}$ to indicate that we repeat it $l-1$ times. Then, the surface of our glyph is given by the points $\{\mathcal{J} \overset{l-1}{\bullet} \mathbf{u} \,|\, \mathbf{u} \in S\}$.

As the 2D examples in Fig. 1 illustrate, the extrema of these new glyphs coincide with the extrema of the Reynolds glyphs. However, they develop sharp features around the maxima, at the cost of a smoother shape around the minima. Consequently, we name them higher-order maximum enhanced (HOME) glyphs. In examples two and three, the HOME glyphs immediately make clear that the respective tensors are not axially symmetric, a fact which the Reynolds glyph may not reveal at first glance. Since we are generally more interested in the maxima than in the minima of the contrast function, we use the HOME glyphs in the remainder of this chapter.

3.2 Experimental Results

We now present some experiments to confirm that higher-order structure tensors indeed give a more accurate representation of junctions and multivalued images. Our first experiment uses simple junctions in synthetic grayscale images. Derivatives are calculated by convolution with a derivative-of-Gaussian filter ($\sigma = 0.7$). After HOSTs of different order l have been computed, their information is propagated to a local neighborhood by convolution with a Gaussian kernel ($\rho = 1.4$).

Figure 2 shows the test images, with the position of the displayed structure tensor marked by a cross. The results show that a HOST of order $l = 4$ is

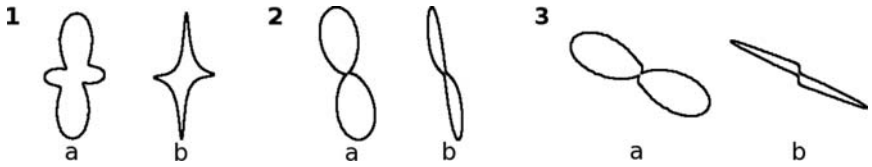
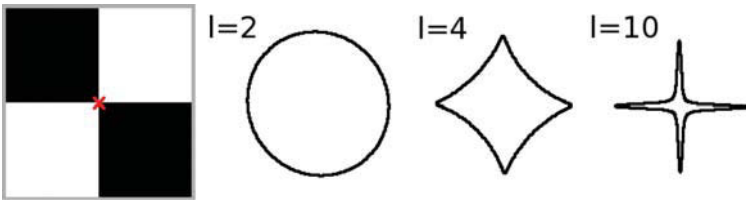
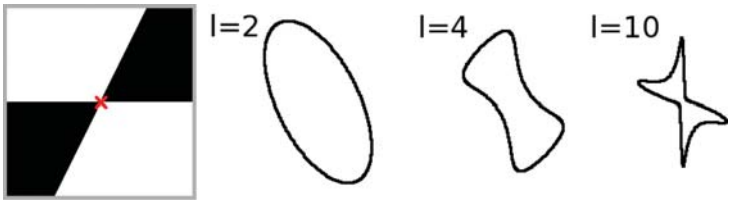


Fig. 1. Three tensors of order six, visualized with Reynolds glyphs (**a**) and our new HOME glyphs (**b**). In (**b**), maxima of the contrast function appear more localized



(a) Orthogonal edges are clearly distinguished with order $l = 4$.



(b) For non-orthogonal edges, higher orders result in more accurate representations.

Fig. 2. Two junctions in grayscale images and the corresponding structure tensors. For orders $l > 2$, the directions of the meeting edges can be represented

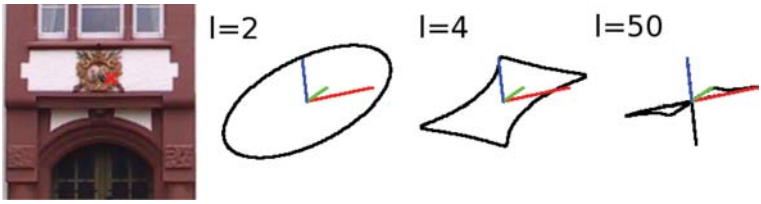


Fig. 3. In a color image, the channel-wise gradients may point into different directions. Higher-order structure tensors can be used to model this situation accurately

sufficient to represent two edges that cross orthogonally, while the traditional structure tensor ($l = 2$) does not distinguish any particular direction. In the nonorthogonal case, the traditional model indicates a principal direction, but it does not correspond to any gradient found in the image. In contrast, the HOME glyph of order four gives an impression of the involved directions and Sect. 4.4 will show that the generalized eigenvectors of the HOST give a good approximation of the gradient directions in this case. However, a clear separation of the maxima in the contrast profile requires higher orders.

The second experiment is based on a natural color image. Derivatives are now calculated channel-wise and according to the conventional generalization to multi-channel images, the HOSTs of the red, green, and blue color channels are added. In this case, we do not propagate the structure information ($\rho = 0$).

For comparison, Fig. 3 also shows the gradients of the individual color channels. Again, the structure tensor of order four gives a much better impression of the dominant directions than the traditional model. To demonstrate

the feasibility of going to very high tensor orders, we also present the representation with $l = 50$.

4 A Mathematical Toolbox

4.1 Efficient Representation

An order-l tensor in n dimensions has n^l tensor channels, which becomes impractical already for moderate l. However, higher-order structure tensors are totally symmetric, that is, invariant under permutation of their indices. This reduces the number of independent channels to $N = \binom{n+l-1}{l}$, which means merely linear growth for $n = 2$ ($N = l + 1$) and quadratic growth for $n = 3$.

With some additional notation, it is possible to evaluate $J(\mathbf{u})$ directly from this nonredundant representation: Call the ith nonredundant element $[\mathcal{J}]_i$, stored in a zero-based linear array $[\mathcal{J}]$. Let $\nu_{i,k} \in \{0, 1, \ldots, l\}$ denote the number of times $k \in \{1, 2, \ldots, n\}$ appears as an index of the ith element. The multiplicity of element i, denoted μ_i, is the number of times it appears as a channel of the original tensor. For $n = 2$, $\mu_i = \binom{l}{\nu_{i,1}}$, for $n = 3$, $\mu_i = \binom{l}{\nu_{i,1}}\binom{l-\nu_{i,1}}{\nu_{i,2}}$. Then, (3) can be rewritten as

$$J(\mathbf{u}) = \sum_{i=0}^{N-1} \mu_i [\mathcal{J}]_i u_1^{\nu_{i,1}} u_2^{\nu_{i,2}} \cdots u_n^{\nu_{i,n}}. \tag{6}$$

For $n = 2$, we chose indices such that $\nu_{i,1} = l - i$ (e.g., $[\mathcal{J}_{1111}, \mathcal{J}_{1112}, \ldots]$).

4.2 Relation to Truncated Fourier Series

From generalized DT-MRI, it is known that using a higher-order tensor model in 3D is equivalent to approximating the diffusivity profile with a truncated Laplace series [14]. We now show that the corresponding result in 2D is a relation of higher-order tensors to truncated Fourier Series. This fact will serve as the basis of the methods in Sects. 4.3 and 4.5.

Consider a Fourier Series, truncated after order l:

$$f(\phi) = \frac{1}{2}a_0 + \sum_{k=1}^{l} a_k \cos(k\phi) + \sum_{k=1}^{l} b_k \sin(k\phi). \tag{7}$$

Setting $a_k := b_k := 0$ for odd k leaves a $l + 1$-dimensional vector space of functions. For $n = 2$, (6) can be rewritten in polar coordinates:

$$J(\phi) = \sum_{i=0}^{l} [\mathcal{J}]_i \binom{l}{i} \cos^{l-i}\phi \sin^i \phi. \tag{8}$$

Let us regard $[\mathcal{J}]_i$ as coefficients and $\binom{l}{i}\cos^{l-i}\phi\sin^i\phi$ as basis functions. We now show that these basis functions span the same space as the truncated Fourier Series.

Proof by induction on order l. Let $\{\mathbf{f}_k\}$ denote the basis functions of a truncated Fourier Series in which only even multiples of ϕ are allowed:

$$\mathbf{f}_k := \begin{cases} 0.5 & \text{if } k = 0 \\ \cos((k+1)\phi) & \text{if } k \text{ odd} \\ \sin(k\phi) & \text{if } k \text{ even } (k \neq 0). \end{cases}$$

Likewise, $\mathbf{t}_k^l$ is the kth basis function of an order-l tensor:

$$\mathbf{t}_k^l := \binom{l}{k}\cos^{l-k}\phi\sin^k\phi.$$

For $l = 0$, both the Fourier Series and the tensor basis represent constant functions and $\mathbf{f}_0 = 0.5\mathbf{t}_0^0$. Assume that the functions that can be represented using $\{\mathbf{f}_k\}$ with $k \leq l$ are equivalent to the functions represented by $\{\mathbf{t}_k^l\}$. Further, assume that we know how to express the Fourier basis in terms of the tensor basis. Then, we can show that the same assumption also holds for $l + 2$: Observe that

$$\begin{aligned}\cos^{l-i}\phi\sin^i\phi &= \left(\cos^2\phi + \sin^2\phi\right)\cos^{l-i}\phi\sin^i\phi \\ &= \cos^{l+2-i}\phi\sin^i\phi + \cos^{l-i}\phi\sin^{i+2}\phi\end{aligned}$$

and that the latter functions are proportional to functions in $\{\mathbf{t}_k^{l+2}\}$. Thus, we can express the first $l + 1$ Fourier basis functions in terms of $\{\mathbf{t}_k^{l+2}\}$ by replacing each occurrence of $\mathbf{t}_k^l$ in their known representation by

$$\mathbf{t}_k^l = \frac{(l+2-k)(l+1-k)}{(l+2)(l+1)}\mathbf{t}_k^{l+2} + \frac{(k+2)(k+1)}{(l+2)(l+1)}\mathbf{t}_{k+2}^{l+2}.$$

It remains to be shown how to express $\mathbf{f}_{l+1}$ and $\mathbf{f}_{l+2}$ in terms of $\{\mathbf{t}_k^{l+2}\}$. For this, we use trigonometric identities for multiple angles:

$$\mathbf{f}_{l+1} = \cos((l+2)\phi) = \sum_{i=0}^{l/2+1}(-1)^i\binom{l+2}{2i}\cos^{l+2-2i}\phi\sin^{2i}\phi = \sum_{i=0}^{l/2+1}(-1)^i\mathbf{t}_{2i}^{l+2}$$

$$\mathbf{f}_{l+2} = \sin((l+2)\phi) = \sum_{i=0}^{l/2}(-1)^i\binom{l+2}{2i+1}\cos^{l+1-2i}\phi\sin^{2i+1}\phi = \sum_{i=0}^{l/2}(-1)^i\mathbf{t}_{2i+1}^{l+2}.$$

□

Our proof is constructive in the sense that it implies a recursive method to construct a change-of-basis matrix. For reference, Table 1 presents the relations for $l = 2$ and $l = 4$.

Table 1. Relation of Fourier coefficients and tensor components for orders $l = 2$ and $l = 4$

$l = 2$	$a_0 = [\mathcal{J}]_0 + [\mathcal{J}]_2$	$a_2 = \frac{1}{2}[\mathcal{J}]_0 - \frac{1}{2}[\mathcal{J}]_2$	$b_2 = [\mathcal{J}]_1$
$l = 4$	$a_0 = \frac{3}{4}[\mathcal{J}]_0 + \frac{3}{2}[\mathcal{J}]_2 + \frac{3}{4}[\mathcal{J}]_4$	$a_2 = \frac{1}{2}[\mathcal{J}]_0 - \frac{1}{2}[\mathcal{J}]_4$	$b_2 = [\mathcal{J}]_1 + [\mathcal{J}]_3$
		$a_4 = \frac{1}{8}[\mathcal{J}]_0 - \frac{3}{4}[\mathcal{J}]_2 + \frac{1}{8}[\mathcal{J}]_4$	$b_4 = \frac{1}{2}[\mathcal{J}]_1 - \frac{1}{2}[\mathcal{J}]_3$

A method to compute these relations for general l is given in the text

4.3 Generalized Tensor Trace

The second-order tensor trace has been used as a substitute of the squared gradient magnitude [16]. For the higher-order case, Özarslan et al. [17] have proposed a generalized trace operation "gentr" in 3D, which is based on integrating J over the unit hemisphere Ω and reduces to the standard matrix trace for $l = 2$:

$$\mathrm{gentr}(\mathcal{J}) := \frac{3}{2\pi} \int_{\Omega} J(\mathbf{u})\, d\mathbf{u}. \tag{9}$$

In the 2D case, Ω is one half of the unit circle and the normalization factor $\frac{3}{2\pi}$ is to be replaced with $\frac{2}{\pi}$. Since the generalized trace of an order-l 2D tensor equals its Fourier coefficient a_0^l, we can use the results from Sect. 4.2 to verify that

$$\mathrm{gentr}(\mathcal{J}) = a_0^l = 2 \sum_{i=0}^{l/2} [\mathcal{J}]_{2i} \frac{(l-1)!!}{(l-2i)!! \cdot (2i)!!}, \tag{10}$$

where $l!!$ is the *double factorial,* that is, the product of integers in steps of two.

In the definition of $\mathcal{J}$, we scaled the gradient magnitude such that the maximum value of J is invariant to the tensor order. However, maxima become narrower with increasing l, so the generalized trace decreases. It follows from (10) that the generalized trace of an order-l structure tensor equals

$$\mathrm{gentr}(\mathcal{J}) = 2 \frac{(l-1)!!}{l!!} |\nabla f|^2. \tag{11}$$

4.4 Generalized Eigenvector Decomposition

Many applications of the second-order structure tensor depend on its spectral decomposition into eigenvectors and eigenvalues (e.g., [3, 5, 6, 16, 18]). In this section, we introduce the CAND (Canonical Decomposition), which can be regarded as a generalized eigen decomposition for higher-order tensors and has first been studied by Hitchcock [19, 20]. A review in the context of higher-order statistics and some new results are given by Comon et al. [21, 22].

We concentrate on the symmetric CAND (sCAND), which decomposes a symmetric order-l tensor $\mathcal{J}$ into a sum of r outer powers $\mathbf{v}_i^{\otimes l}$ of unit vectors $\mathbf{v}_i$, $i \in \{1, 2, \ldots r\}$, scaled with λ_i:

$$\mathcal{J} = \sum_{i=1}^{r} \lambda_i \mathbf{v}_i^{\otimes l}. \tag{12}$$

For $l = 2$, (12) reduces to the spectral decomposition, where λ_i are the eigenvalues and $\mathbf{v}_i$ the eigenvectors. It can be shown that any symmetric higher-order tensor $\mathcal{J}$ has a sCAND [22]. In analogy to the matrix rank, the symmetric rank R_S of $\mathcal{J}$ is defined as the smallest number r for which a sCAND exists. In dimension $n = 2$, it holds that $R_S \leq l$ [21].

The CAND is a current research topic and some theoretical results have been obtained. However, practical algorithms for efficient computation are rare. Fortunately, Comon et al. [21] present an algorithm that works for $n = 2$ and thus can be applied to our HOSTs. It is outside the scope of this chapter to review the full theory required to derive the algorithm. For our experiments, we have reimplemented the MATLAB code given in [21] in C, using routines from LAPACK[3] and the numerical recipes [23].

The algorithm returns pairs of λ_i and $\mathbf{v}_i$, where the $\mathbf{v}_i$ are not normalized. While it appears trivial to convert this result to the canonical form, the algorithm proves numerically unstable for vectors $\mathbf{v}_i$, which are nearly aligned with the y-axis: In such cases, λ_i tends to zero, while the magnitude of $\mathbf{v}_i$ tends to infinity. We work around this problem by reconstructing a tensor $\mathcal{J}'$ only from those $\mathbf{v}_i$ which have a reasonable magnitude. Then, the residual $\tilde{\mathcal{J}} := \mathcal{J} - \mathcal{J}'$ can be rotated by 90° to obtain the remaining $\mathbf{v}_i$. Note that the tensor rotation only requires a simple permutation of its elements and some sign changes: In the array representation from Sect. 4.1, it is sufficient to reverse the array and to multiply all entries $[\mathcal{J}]_i$ with an odd index i by -1.

Figure 4 visually represents (12) for a particular HOST. Even though the gradient directions in the neighborhood of the considered pixel are too close to be resolved in the contrast profile of an order-four tensor, they are well approximated by the two largest generalized eigenvectors.

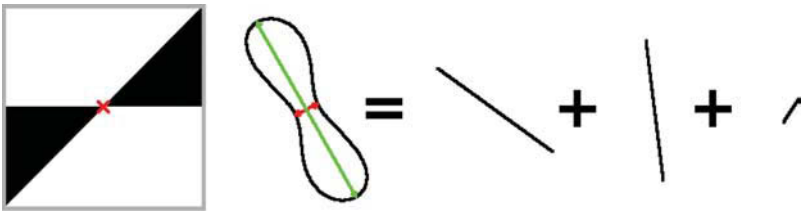


Fig. 4. Generalized eigenvectors can be used to recover individual directions from a higher-order structure tensor

[3] http://www.netlib.org/lapack/

Our prototype implementation found the sCand of 160,000 order six structure tensors from a natural color image in around 2.5 s on a 2 GHz Athlon 64.

4.5 Extrema of the Contrast Function

A frequent problem when dealing with higher-order structure tensors will be to find the angles at which the contrast function J attains an extremum. In the second-order case, maxima and minima are given by the directions of the major and minor eigenvectors, respectively. However, the generalized eigenvectors do not in general coincide with maxima in the contrast function.

To find the extrema of the contrast function, we use the Fourier Series representation from Sect. 4.2. This makes it easy to take derivatives of J, since they are again Fourier Series of the same order, whose coefficients are straightforward to compute.

4.5.1 Accelerating the Brute Force Method

Extracting extrema requires to find angles ϕ at which $J'(\phi) = 0$. The obvious method to find such points for an arbitrary differentiable function J is to sample its derivative with some resolution r to identify intervals in which it changes sign and to refine the result to a desired accuracy a by a binary search for the sign change on these intervals.

This method may miss pairs of extrema whose distance is less than the sampling resolution r. Fortunately, such pairs are usually only minor local variations in the contrast function, which are of no practical interest (cf. Fig. 5), so we found $r = 2°$ sufficient for structure tensors of order six. For higher orders, denser sampling will be necessary, as peaks in the contrast function become sharper.

The computational cost of this method is dominated by the cost of evaluating the derivative J'. Evaluating (7) directly involves l sines and l cosines. A recursive formulation exists, which is known alternatively as Clenshaw's algorithm or as the Goertzel–Watt algorithm and requires only a single sine and cosine. It is given by the recursion rule

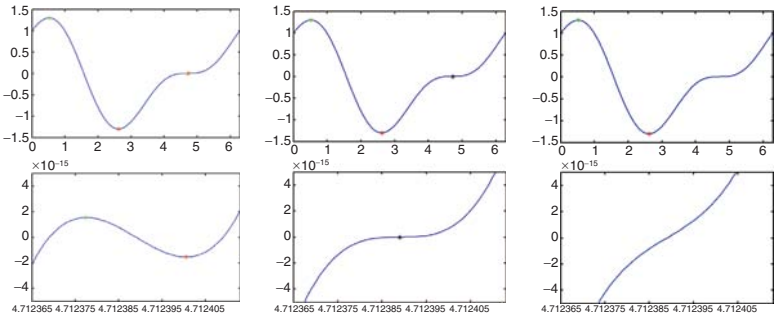
$$u_{l+1} = u_{l+2} = 0 \tag{13a}$$

$$u_r = f_r + 2u_{r+1}\cos\phi - u_{r+2} \quad \text{with } r = l, l-1, \ldots, 1, \tag{13b}$$

where $f_r := a_r$ if the sum of cosines is to be computed, $f_r := b_r$ for the sum of sines. From u_1 and u_2, the final result is determined as

$$\sum_{k=1}^{l} a_k \cos(k\phi) = u_1 \cos\phi - u_2 \quad \text{and} \quad \sum_{k=1}^{l} b_k \sin(k\phi) = u_1 \sin\phi. \tag{14}$$

Gentleman [24] has shown that this method magnifies roundoff errors when evaluated near $\phi = k\pi$ ($k \in \mathbb{Z}$) and can produce unusable results for large



(a) $\cos\phi + (\frac{1}{2} + \varepsilon)\sin 2\phi$ (b) $\cos\phi + \frac{1}{2}\sin 2\phi$ (c) $\cos\phi + (\frac{1}{2} - \varepsilon)\sin 2\phi$

Fig. 5. In a Fourier Series, a small ε (here, $\varepsilon = 10^{-10}$) can make the difference between a pair of extrema, a saddle, and no stationary point. The top row shows one full period, while the bottom row gives a close-up of the affected extrema

orders l. Newbery [25] suggests to adaptively perform a phase shift of $\pi/2$ to avoid such cases. However, an experimental comparison of direct evaluation, the original and the modified version of Clenshaw's algorithm indicates that the error is tolerable for the moderate values of l that occur in our context: For 160,000 structure tensors of order six from a natural color image, all methods produced identical results (with $r = 2°$ and $a = (2^{-7})°$, at single precision). Even with $l = 50$, all methods gave the same extrema, with identical results in more than 99% of the cases and a maximum angular deviation of $(2^{-7})°$ in the rest of them. Thus, we chose the unmodified Clenshaw algorithm, which gave a speedup factor of 2.2 for $l = 6$.

4.5.2 A Faster Method

We now present a more efficient algorithm which exploits the fact that even higher derivatives of J are easy to evaluate. The basic idea of the method is to expand J' into a Taylor series $\tilde{J}'$, which is terminated after degree three to obtain a polynomial that can be solved analytically in a numerically stable manner [23]. Then, the error bounds of the expansion define a corridor around the x-axis. For intervals in which $\tilde{J}'(\phi)$ is outside this corridor, we can be sure that $J'(\phi) \neq 0$, that is, they do not contain an extremum.

It is possible to recurse on the (now shorter) intervals in which $\tilde{J}'(\phi)$ lies within the corridor; this allows one to identify extrema which are so close that finding them with the brute force approach would be computationally infeasible. For example, finding the pair of extrema that is shown in Fig. 5(a) would require sampling at $r \approx (2^{-10})°$. In contrast, our algorithm needs only 10 Taylor expansions to identify all four extrema.

While it is nice to be able to find such small extrema, they are not important in our context, so we would prefer to find the major ones more efficiently.

As intervals get down to a certain width, the computational effort of reducing them further via the Taylor expansion exceeds the cost of the brute force method. A simple way to combine both methods is to use the Taylor expansion once to cut down search space, and to run the brute force method on the remaining intervals. In the above experiment with order six tensors ($r = 2°$, $a = (2^{-7})°$), this gave a speedup of factor two, reducing the total time for processing all 160,000 tensors on a 2 GHz Athlon 64 processor to 2.2 s, which is about as fast as a sCand on the same data. The speedup is a function of desired angular resolution and higher for smaller r (factor 7.5 for $r = 0.1°$).

4.5.3 Implementation of the Faster Method

Since we cannot expect a polynomial of degree three to reasonably approximate a sine of frequency l in an interval larger than π/l, we initially partition $[0, 2\pi)$ into $2l$ equal intervals for a Fourier series of order l. For each of these intervals, the Taylor expansion is performed by evaluating the higher derivatives at its center ϕ_0. The required Fourier coefficients are precomputed once. For a third-order approximation, the error bound Δ is

$$\Delta = \frac{J^{(5)}(\phi)}{4!}(\phi - \phi_0)^4 \tag{15}$$

for some ϕ within the interval. Taking the order-five derivative in this expression is appropriate, since we approximate J'. To obtain a simple, safe estimate of Δ, the Fourier coefficients a_k and b_k of $J^{(5)}(\phi)$ are used to state that

$$J^{(5)}(\phi) \leq \sum_{k=1}^{l} \sqrt{a_k^2 + b_k^2} \tag{16}$$

and half the interval length is taken for $(\phi - \phi_0)$.

Now, the roots of $\tilde{J}' - \Delta$ and $\tilde{J}' + \Delta$ give the intersections with the upper and lower error bound, respectively. Starting from the value of $\tilde{J}'$ at the left interval boundary, we can go through the sorted error bound intersections to determine the intervals in which $\tilde{J}'$ is within the error corridor. Within some intervals, $\tilde{J}'$ may lie fully inside or outside the corridor.

5 Integrated Edge and Junction Detection

Given a grayscale image, edges can be seen as lines across which local contrast is high. Then, corners are points in which edges bend sharply, and junctions are points in which two or more edges meet. Corner detection has been one of the first applications of the structure tensor [1]. More recently, Köthe [18] has presented an algorithm that uses the structure tensor for "integrated" edge and junction detection.

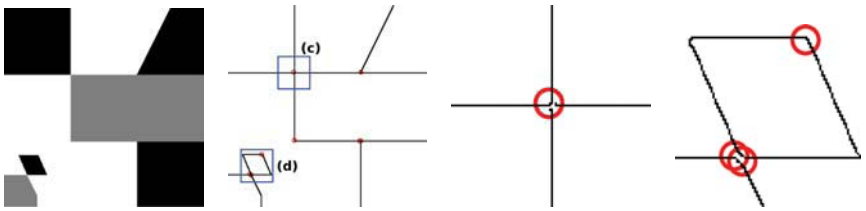


Fig. 6. Subfigure (**b**) shows edges and junctions extracted from (**a**) using second-order structure tensors. (**c**) and (**d**) illustrate that edges break down as the tensor becomes isotropic near junctions (marked with *circles*)

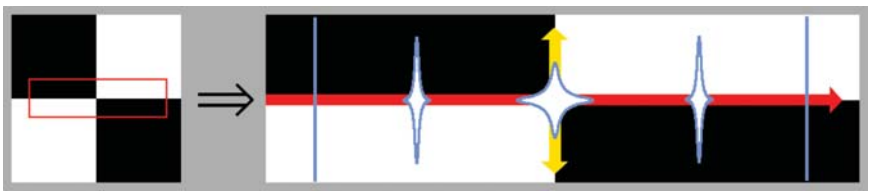


Fig. 7. Scheme of the tracking process that extracts edges and junctions simultaneously

While this method is a considerable improvement over previous approaches, it finally decomposes the structure tensor field into an "edge" and a "junction" part, which are processed by separate algorithms to produce the respective edge and junction maps. Because of this, edges break down near junctions, it is not always clear which edges are connected by a junction, and isolated or duplicate junctions can occur (cf. Fig. 6).

Unlike the second-order structure tensor, the HOST at a junction holds enough information to find the adjacent edges. This allows one to extract edges, corners, and junctions in a single, fully integrated process. The fundamental idea is similar to tracking lines in higher-order tensor fields [15], except that we assume that edges are *orthogonal* to contrast maxima.

Figure 7 illustrates the process of tracking edges in a HOST field: Starting from a seed with locally maximal generalized trace, we integrate edges orthogonal to the major contrast indicated by the tensor (red arrow). When a secondary peak in the contrast function attains a local maximum, we insert a junction and start new edges from it (yellow arrows).

While standard hyperstreamline integration techniques can be applied by interpolating the higher-order tensors channel-wise, we are typically not interested in edge maps that have a greater resolution than the structure tensor grid (note, however, that it is advisable to sample the structure tensors at twice the image resolution [18]).

Thus, we implemented a simple tracking algorithm that works on the given grid. For each edge, it outputs a list of sub-pixels that belong to it, and each

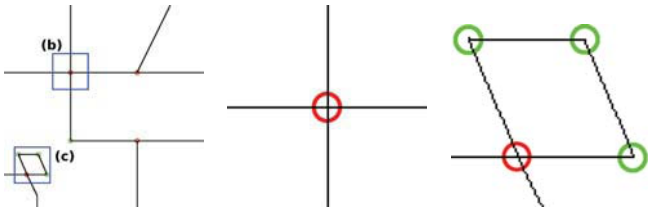


Fig. 8. With higher-order tensors, edges can be traced through and assigned to junctions, which allows easy classification of corners (*green*) and junctions (*red*)

junction holds references to the edges it connects. This information allows for an easy classification of corners and junctions, which are shown in different colors in Fig. 8, based on the number of incident edges.

Integration over a Gaussian neighborhood propagates the contrast from edges beyond corner points, which leads to short "phantom" edges that make corners appear as junctions. To avoid such artifacts, we reject edges along which the HOST indicates multiple directions in all sub-pixels. Also, edges may reach existing junctions through a cycle (like in Fig. 8(c)), and we need to connect them explicitly to such junctions.

Figure 8 presents a result of our method with HOSTs of order $l = 6$. The present algorithm gives superior results on synthetic images and demonstrates the applicability of the methods introduced in Sect. 4. However, a version that is robust enough for natural images has to be left for future research. In particular, we plan to explore the potential of using generalized eigenvectors instead of contrast maxima to steer the tracking.

Our implementation of Köthe's algorithm took 0.3 s for the shown example image (including his anisotropic averaging), while our method needed 0.9 s, again including computation and integration of the tensors. The original image size was 300×300 pixels, giving a 599×599 sub-pixel structure tensor field.

6 Conclusions and Future Work

In the present work, we have shown how higher-order tensors can be used to represent the average of orientations in greater detail than it is possible using traditional second-order structure tensors. We have introduced the notions, definitions, and mathematical tools required to work with such higher-order structure tensors efficiently and to visualize them appropriately.

While Sect. 5 demonstrates the advantages of the HOST for integrated edge and junction detection, it is intended as a proof of concept for the introduced methods, not as the ultimate goal of our research. Consequently, our next step will be to explore the potential of the HOST in several image processing and computer vision applications. We have already conducted

some promising experiments on using HOSTs for texture segmentation and for steering image diffusion.

All results in this chapter are in 2D. Some of the theory (Sect. 2) and practical methods (Sects. 3 and 4.1) easily carry over to three dimensions, or equivalents can be taken from the literature (Sects. 4.2 and 4.3). However, efficient methods for the sCand with $n = 3$ and for finding the maxima of 3D contrast functions are still missing and probably require substantial research.

Acknowledgements

We thank Holger Theisel, who is with the University of Magdeburg, for discussions at all stages of this project. Discussions with Torsten Langer, who is with the MPI Informatik, helped in developing parts of the "mathematical toolbox" in Sect. 4.

Our implementation uses the CImg library by David Tschumperlé, available from `http://cimg.sf.net/`.

This research has partially been funded by the Max Planck Center for Visual Computing and Communication (MPC-VCC).

References

[1] Förstner, W., Gülch, E. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: International Society for Photogrammetry and Remote Sensing (ISPRS) Intercomission Conference on Fast Processing of Photogrammetric Data, Interlaken (1987) pp. 281–305

[2] Bigün, J., Granlund, G., Wiklund, J. Multidimensional orientation estimation with applications to texture analysis and optical flow. IEEE Transactions on Pattern Analysis and Machine Intelligence **13** (1991) 775–790

[3] Weickert, J. Anisotropic Diffusion in Image Processing. Teubner, Stuttgart (1998)

[4] Rousson, M., Brox, T., Deriche, R. Active unsupervised texture segmentation on a diffusion based feature space. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Madison, Wisconsin, USA (2003) pp. 699–706

[5] Tschumperlé, D., Deriche, R. Vector-valued image regularization with PDE's: A common framework for different applications. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2003). (2003) pp. 651–656

[6] Galić, I., Weickert, J., Welk, M., Bruhn, A., Belyaev, A., Seidel, H.P. Image compression with anisotropic diffusion. Journal of Mathematical Imaging and Vision **31** (2008) 255–269

[7] Di Zenzo, S. A note on the gradient of a multi-image. Computer Vision, Graphics, and Image Processing **33** (1986) 116–125
[8] Bigün, J., Bigün, T., Nilsson, K. Recognition by symmetry derivatives and the generalized structure tensor. IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004) 1590–1605
[9] Weickert, J., Brox, T. Diffusion and regularization of vector- and matrix-valued images. In: Nashed, M., Scherzer, O., eds., Inverse Problems, Image Analysis, and Medical Imaging. Volume 313 of Contemporary Mathematics. AMS, Providence (2002) pp. 251–268
[10] Brox, T., Weickert, J., Burgeth, B., Mrázek, P. Nonlinear structure tensors. Image and Vision Computing **24** (2006) 41–55
[11] Arseneau, S., Cooperstock, J.R. An improved representation of junctions through asymmetric tensor diffusion. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Remagnino, P., Nefian, A.V., Gopi, M., Pascucci, V., Zara, J., Molineros, J., Theisel, H., Malzbender, T., eds., Advances in Visual Computing. Volume 4291 of Lecture Notes in Computer Science, Springer (2006) pp. 363–372
[12] Herberthson, M., Brun, A., Knutsson, H. Pairs of orientations in the plane. In: Proceedings of the SSBA Symposium on Image Analysis, Umeå, Sweden, SSBA (2006) pp. 97–100
[13] Brox, T., Weickert, J. A TV flow based local scale measure for texture discrimination. In: Pajdla, T., Matas, J., eds., Proc. 8th European Conference on Computer Vision (ECCV'04). Volume 3022 of LNCS., Springer (2004) pp. 578–590
[14] Özarslan, E., Mareci, T. Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. Magnetic Resonance in Medicine **50** (2003) 955–965
[15] Hlawitschka, M., Scheuermann, G. HOT-lines: Tracking lines in higher order tensor fields. In: Silva, C., Gröller, E., Rushmeier, H., eds., Proceedings of IEEE Visualization 2005, Minneapolis, MN, USA (2005) pp. 27–34
[16] Feddern, C., Weickert, J., Burgeth, B. Level-set methods for tensor-valued images. In: Faugeras, O.D., Paragios, N., eds., Proc. Second IEEE Workshop on Geometric and Level Set Methods in Computer Vision, Nice, France (2003) pp. 65–72
[17] Özarslan, E., Vemuri, B.C., Mareci, T.H. Generalized scalar measures for diffusion MRI using trace, variance, and entropy. Magnetic Resonance in Medicine **53** (2005) 866–876
[18] Köthe, U. Edge and junction detection with an improved structure tensor. In: Michaelis, B., Krell, G., eds., Pattern Recognition. 25th DAGM Symposium. Volume 2781 of Lecture Notes in Computer Science, Springer (2003) pp. 25–32
[19] Hitchcock, F.L. The expression of a tensor or a polyadic as a sum of products. Journal of Mathematics and Physics **6** (1927) 164–189

[20] Hitchcock, F.L. Multiple invariants and generalized rank of a p-way matrix or tensor. Journal of Mathematics and Physics **7** (1927) 39–79
[21] Comon, P., Mourrain, B. Decomposition of quantics in sums of powers of linear forms. Signal Processing **53** (1996) 96–107
[22] Comon, P., Golub, G., Lim, L.H., Mourrain, B. Symmetric tensors and symmetric tensor rank. Technical Report SCCM-06-02, Stanford Scientific Computing and Computational Mathematics (SCCM) (2006)
[23] Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P. Numerical Recipes in C++: The Art of Scientific Computing, Second Edition. Cambridge University Press (2002)
[24] Gentleman, W. An error analysis of Goertzel's (Watt's) method for computing Fourier coefficients. The Computer Journal **12** (1969) 160–164
[25] Newbery, A. Error analysis for Fourier series evaluation. Mathematics of Computation **27** (1973) 639–644

Monogenic Curvature Tensor as Image Model*

Gerald Sommer, Lennart Wietzke, and Di Zang

Department of Computer Science, Christian-Albrechts-University, Kiel, Germany
{gs,LW}@KS.iNFORMATiK.uni-kiel.de

Summary. In this chapter, a new rotation-invariant generalization of the analytic signal will be presented to analyze intrinsic 1D and 2D local image structures. By combining differential geometry and Clifford analysis, the monogenic curvature tensor can be derived to perform a split of identity and to enable simultaneous estimation of local amplitude, phase, main orientation, and angle of intersection in a monogenic scale-space framework.

1 Introduction

The processing and analysis of images and image sequences is a well established technology, although not fully satisfactory in some respect. Contemporary stated limitations have its reasons in the lack of a well founded and powerful theory of multidimensional signals. Because of the different topology of multidimensional signals in comparison to one-dimensional ones, serious consequences result with respect to formulation of a multidimensional signal theory. A signal theory should support the modeling of signal structures we are interested in and the operations we are applying to cope with certain tasks at hand. For both that signal theory should deliver useful representations. For instance, it is well known that the complex-valued 1D Fourier transform enables a global view on the parity symmetry decomposition of a 1D function. Less known is the fact that this fails in case of the 2D Fourier transform because the possible symmetries are partially covered in the real and imaginary parts of the spectrum [42]. The central problem of modeling is the so-called representation problem. That is the problem of giving a certain concept a useful representation form. To cope with that problem in science and engineering, algebra as a mathematical language often delivers the right structure of representations. Also analysis and geometry as other resources of modeling are tightly related to algebra. Let us give two examples: First,

* This work was supported by DFG grant So 320/4-2 (GS,LW) and DFG Graduiertenkolleg No. 357 (DZ)

the holomorphic extension of a real-valued 1D function to a complex one is a well-known method of complex analysis. But doing the same for a nD function requires another algebraic framework for analysis. The coupling of analysis with Clifford algebra establishes the Clifford analysis [7] as a useful approach to multidimensional functions. Second, the tight coupling of geometry with algebra is well-known since Felix Klein. Clifford algebra delivers algebraic structures for modeling any type of geometry, which is particularly interesting in the case of multiple dimensions. Hence, we have to take advantage of the achievements in math for handling multidimensional functions. There are different interesting mathematical sources available, which extend the representation of real-valued multidimensional signals with the result that their structure becomes accessible. A well established concept is tensor algebra as generalization of vector algebra or matrix algebra [10]. Tensors are well-known as useful representations of geometry [11]. Clifford algebra or geometric algebra [22] is another concept, which only recently has been considered in engineering [40]. These algebras constitute other generalizations of the vector algebra, namely with respect to the representation of higher-order directed numbers, called multivectors. Both tensor algebras and geometric algebras deliver rich subspace structures in comparison to vector algebra. An advantage of geometric algebra over tensor algebra is its easier interpretation with respect to geometric concepts. Some entities contributing to the formulation of a problem can immediately represent geometry while possessing algebraic properties. Vice versa, the advantage of tensor algebra over geometric algebra is its easier numerical realization. Therefore, in practice it may be advantageous to transform expressions from geometric algebra to tensor algebra, see, for example, [41].

Very important in signal theory is the use of complex numbers. But according to our experience, complex numbers are only adequate to model one-dimensional signals. As already mentioned earlier, in the case of multidimensional signals, the algebraic framework has to be extended accordingly, see, for example, [42]. While Clifford algebra or geometric algebra supports a global view onto signal structures, Clifford analysis is useful for a local approach to signal analysis. The Hilbert transform takes over the role of the Fourier transform in the case of locally expanding a real-valued one-dimensional function to a complex one. This corresponds to the holomorphic extension in complex analysis. In Clifford analysis we instead meet the Riesz transform, which delivers a Clifford valued expansion of a real-valued multidimensional function. The resulting representation is called a monogenic function. In both cases a real-valued function will be completed by a harmonic conjugate which is in quadrature phase relation to the original real function. This most useful property will play a leading role throughout this chapter.

In this contribution, we are fusing the concepts of differential geometry for local image modeling in a tensor representation with the Clifford analysis concept of monogenic functions. This delivers a representation, called monogenic curvature tensor, which will enable local image analysis from one single

coherent mathematical point of view. The evaluation of the monogenic curvature tensor [48] delivers two curvature related signal representations, which are specific for either intrinsically 1D or 2D structures. These signal representations are both generalizations of the well-known analytic signal [20]. From these Clifford valued signal representations local amplitude and local phase as local spectral representations as well as some geometric features can be computed. Because the monogenic curvature tensor is embedded in a monogenic scale-space [16], all features derived from it possess their own scale-space representation. Furthermore, the monogenic scale-space is the unifying framework for the scale related properties of all derived features.

In Sect. 2, we describe the required properties of the wanted signal model and the difficulties occurring in related work. In particular, we give a short view on tensor based image analysis. In Sect. 3, we derive the monogenic curvature tensor and will present its evaluation. Finally, we give a short summary and conclusion in Sect. 4. Note that we will not give an introduction to geometric algebra or Clifford analysis. Instead, the reader is advised to have a look at [42] for a short introduction that is specific to the topic of that chapter. Other necessary hints are given at several places of this chapter.

2 Related Work

2.1 Key Point Detectors and Local Image Features

There are two basic tasks in low-level vision, which are building a bottleneck in practice of image analysis, although plenty of work has been done over decades to overcome this situation. These tasks are the detection of points of interest and the analysis of their structure with respect to the neighborhood by a set of meaningful features. A detector is a local operator that has to fulfil two contradictory requirements: good recognition of the structure of interest and good localization accuracy. While the first one is located in a feature domain, the second one is located in spatial domain. In the case of LSI-operators and the frequency space as feature domain, the associated uncertainty principle is well-known [19]. Other problems with detectors are to model the structures of interest and to gain some invariance. While in 1D the number of different structures is quite limited, in 2D it is infinite. With respect to feature descriptors the problem is again to define a meaningful set of features and to gain certain invariance. Low-level image analysis is always model based. Because there is no satisfactory theory of multidimensional signal structures, the only chance is a comparison of different approaches out of a plenty of proposals, see, for example, [38] and [35]. The most prominent feature detector is SIFT [32]. Its model is very simple and the main bulk of work is shifted to classification in a high dimensional feature space. Other types of detectors are based on second order tensors (or matrices), for example, the structure tensor [17]. This detector in essence represents Gaussian smoothed partial derivatives of

first order. Another kind of detector is based on the Hessian matrix, which is related to the curvature tensor of differential geometry [2], see, for example, [1]. These tensors are using derivatives of first and second order as elements. Therefore, they either represent gradient or curvature information. Because they are real-valued, we propose in Sect. 3 to unify the differential geometric point of view with the representation power of Clifford analysis.

2.2 Required Invariance

The image model we want to formulate should have some invariance properties. In general, image signals have no invariants in the strong sense of a quantity that does not change under a specified group of transformations. But there are some features that only slightly change by moderate changes of the conditions an image was recorded at. This is called weak invariance. We use the term invariance instead in such a way that the image representation we are looking for can cope with all variations of the features of interest and thus is complete with respect to that concepts of structure descriptions. This includes another type of invariance, called relative invariance or equivariance. In the case of equivariance, a systematic change of input data will cause another systematic change of output data. For instance, it is known in signal processing that a shift of a function in input space will result in an equivalent phase change in Fourier domain.

2.2.1 Invariance with Respect to Intrinsic Dimension

An image is locally composed of structures of different intrinsic dimension [50]. If the signal embedding dimension is two (a normal image signal), then a locally constant signal has intrinsic dimension zero (i0D), a nonbent edge or line has intrinsic dimension one (i1D), and all other bent structures including corners, junctions, and end stopping points are of intrinsic dimension two (i2D). In Fig. 1 these types of intrinsic dimensions are demonstrated for some simple cases. From left to right follow a constant signal (i0D), an i1D signal which is always a rotated 1D signal and three different examples of i2D signals. The intrinsic dimension obviously corresponds to the number of degrees of freedom necessary to model a function. Responding to either edges/lines or corners/junctions makes a difference between detectors, which are specific to intrinsic dimension. The eigenvalue analysis of the structure tensor is a well-known example. A special nonlinear detector based on a Volterra series approach which is specific to i2D has been proposed in [29]. On the other hand, the monogenic signal [14], see also Sect. 2.2, is specific to i1D structures. There is no detector available that responds to either i1D or i2D structure with a meaningful and rich set of features. It is well-known that the feature space spanned by mean and Gaussian curvature enables the classification with respect to intrinsic dimension [2], see also Sect. 2.4. We use that fact because we require completeness with respect to that concept.

Fig. 1. Example functions of different intrinsic dimension

2.2.2 Invariance with Respect to Parity Symmetry

Parity symmetry locally indicates either symmetry or antisymmetry, respectively, even or odd symmetry, of a structure in the case of reflection at a certain location. This enables, for example, classification as edge-like or line-like structure in i1D case and similarly in i2D case. Detectors should respond to both types of symmetry in the same manner but distinguishing. A counterexample is the structure tensor, which is sensible to edge-like structures because it is gradient based. Because the first-order derivatives operator represents an odd function, it can only respond to odd structures in the sense of a matching operator. Because a second order derivative operator represents an even function, it responds to even structures. Hence, the combinations of both should deliver the required invariance. But this method has some drawbacks. Instead, most easily, parity symmetry can be decided from computing the local phase [42]. This can be achieved by so-called quadrature filters. These detectors consist of two components which respond to either even or odd symmetry and which are in quadrature phase relation. That means, their components differ only by a phase shift of $\frac{\pi}{2}$. While this is an easy task for 1D signals, in 2D case this is not true. Nevertheless, the Gabor filter [19] is a popular candidate. Only in the framework of Clifford analysis certain concepts of multidimensional local phase can be reasonably formulated, because the topological situation can be modeled with sufficient degrees of freedom and with respect to the relations existing between these.

2.2.3 Invariance with Respect to Scale

Local image structure is restricted to a certain range of scale, called the intrinsic scale. Changing the scale will possibly change all other features. That is, intrinsic scale is a feature too [30] and a scale-adaptive scheme is advantageous in some cases. Traditionally, the Gaussian scale-space embedding of feature detection is used. Regrettably, the only feature that is intrinsic to a Gaussian scale-space is signal intensity. Only recently the monogenic scale-space has been proposed [16] as an alternative scale-space concept where local spectral representations, local orientation, and other geometric features as angle of intersection become features of one single scale-space theory. This result from Clifford analysis will be adopted here.

2.2.4 Invariance with Respect to Rotation

Rotation invariance of detectors is an important requirement. Regrettably, the design of rotation invariant detectors is a nontrivial task. While the structure tensor and the Laplacian are rotation invariant, all detectors in the past based on quadrature filters in multiple dimensions were not [28]. The problem was that in the complex domain there could not be designed an isotropic odd filter, which is in quadrature phase relation to its even counterpart. Only the Riesz transform as generalized Hilbert transform turned out to solve that problem in a Clifford valued domain [14]. This approach will also be used to formulate the harmonic conjugate part of the monogenic curvature tensor. Orientation is an important local feature of a structure, which should be estimated in a rotation invariant way. In case of an i1D structure there are different methods available for orientation estimation, for example, from the eigenvector analysis of the structure tensor, respectively, orientation tensor [3] or by using steerable filters [18]. These approaches do not give access to the other features mentioned in this section. Therefore, the monogenic signal is the method of choice because it delivers besides local energy and local phase also local orientation in a rotationally invariant manner. More complicated is the situation in the case of i2D structures, which are related to multiple orientations meeting in a keypoint. Steerable filters can cope with that situation as well [34], while the structure tensor analysis delivers main and minor orientations, which must not coincide with actual orientations of involved i1D structures. The same problems occur with the analysis of the monogenic curvature tensor. In [37], [21], and [36] different approaches for parametric modeling of corners have been proposed to cope with multiple orientations. Most interesting is the generalization of the structure tensor with respect to a multiple orientation model in [43].

2.2.5 Invariance with Respect to Angle of Intersection

Certain models of i2D structures as corners or junctions are described by superposition of i1D structures, which meet in a keypoint. Hence, there are several geometric features describing such model. The most intuitive one besides orientation is the angle of intersection or apex angle. The detector of i2D structures should respond independent of the angle of intersection. The proposals in [36] and [43] are invariant with respect to the angle of intersection and rotationally invariant. In the framework of the monogenic signal a generalization has been proposed, which is called structure multivector [15]. The involved model is two perpendicularly superimposed i1D structures. This rigid model has been generalized by the operator, which delivers the monogenic curvature tensor. Another problem is estimating the angles of intersection in the case of multiple superimposed i1D structures. All mentioned methods need to know the number of superimposed patterns. For the assumption of two

lines/edges with arbitrary but same phase and amplitude, we present a phase-based method in Sect. 3, which results from the evaluation of the monogenic curvature tensor.

2.3 Monogenic Signal

In this section, we give a short overview on the basic ideas of the monogenic signal as generalization of the analytic signal. More details can be found in the thesis [12] and in the papers [14, 42]. In the case of a real valued 1D signal $f : \mathbb{R} \to \mathbb{R}$, the extension to a complex valued signal $f_A : \mathbb{R} \to \mathbb{C}$, called analytic signal [19], is written in spatial domain as

$$f_A(x) = f(x) + j f_H(x) \; . \tag{1}$$

The components f and f_H are in quadrature phase relation, that is they are phase shifted by $\left|\frac{\pi}{2}\right|$. The imaginary completion f_H is computed from the real signal f by convolution with the Hilbert transform kernel

$$h(x) = \frac{1}{\pi x} \; . \tag{2}$$

Because the Hilbert transform is an all-pass operator, the use of quadrature filters h_q in practice is preferred,

$$h_q(x) = h_e(x) + j h_o(x) \; . \tag{3}$$

This pair of even (h_e) and odd (h_o) operators is in quadrature phase relation within a chosen passband. The best known quadrature filter is the Gabor filter [19], which is also widely used in image processing, that is in case of 2D signals, as oriented quadrature filter [20]. Convolution of $f(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^n$, with $h_q(\boldsymbol{x})$ results in a separation of the output function, $g(\boldsymbol{x})$, with respect to symmetry,

$$g(\boldsymbol{x}) = g_e(\boldsymbol{x}) + j g_o(\boldsymbol{x}), \tag{4}$$

from which the local energy, $e(\boldsymbol{x})$, and the local phase, $\varphi(\boldsymbol{x})$, can be computed,

$$e(\boldsymbol{x}) = g_e^2(\boldsymbol{x}) + g_o^2(\boldsymbol{x}) \tag{5}$$

$$\varphi(\boldsymbol{x}) = \arg g(\boldsymbol{x}) \; . \tag{6}$$

But the lack of rotation invariance of the Gabor filter results in wrong estimates of e and φ in most cases. The need of a rotation invariant generalization of the Hilbert transform in multiple dimensions can be established within Clifford analysis [7]. We assume that an n-dimensional function $f(\boldsymbol{x})$ is embedded into an $(n+1)$-dimensional space as vector field $\boldsymbol{f}(\boldsymbol{x}, x_{n+1}) = f(\boldsymbol{x})\boldsymbol{e}_{n+1}$ with $\boldsymbol{e}_i,\ i = 1, \ldots, n+1$ being unit vectors. Besides, we assume a geometric algebra $\mathbb{R}_{n+1}$ over the vector space $\mathbb{R}^{n+1}$. Such geometric algebra is a linear space with a total number of 2^{n+1} well distinguishable subspaces of different

grade. The unit subspace of highest grade is the so-called unit pseudoscalar $I_{n+1} = \boldsymbol{e}_1\boldsymbol{e}_2 \cdots \boldsymbol{e}_{n+1}$. Then the Riesz transform kernel in either spatial or Fourier domain

$$r_n(\boldsymbol{x}) = \frac{a_n \boldsymbol{x}\boldsymbol{e}_{n+1}}{|\boldsymbol{x}|^{n+1}} \quad \text{and} \quad R_n(\boldsymbol{u}) = \frac{\boldsymbol{u}}{|\boldsymbol{u}|} I_{n+1}^{-1} \tag{7}$$

results as solution of the Dirac equation for $x_{n+1} = 0$. It is an isotropic all-pass operator in any signal dimension n. Note that the geometric product of a geometric algebra is written by juxtaposition of the factors. Hence, in (7) the product $\boldsymbol{x}\boldsymbol{e}_{n+1}$ between the two vectors results in a bivector, belonging to another subspace as the original vectors. In addition, $a_n = \pi^{-\frac{n+1}{2}} \Gamma(\frac{n+1}{2})$ is a constant. If we restrict to the case $n = 2$, then the monogenic signal [14] as generalized analytic signal is represented by the vector field

$$\boldsymbol{f}_M(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{f}^r(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}) + (r_2 * \boldsymbol{f})(\boldsymbol{x}), \tag{8}$$

with $\boldsymbol{f}(\boldsymbol{x}) = f(\boldsymbol{x})\boldsymbol{e}_3$ and

$$\boldsymbol{f}^r(\boldsymbol{x}) = f^r_{(1)}(\boldsymbol{x})\boldsymbol{e}_1 + f^r_{(2)}(\boldsymbol{x})\boldsymbol{e}_2 \; . \tag{9}$$

As we see in (9), the convolution of the original vector field with the operator r_2 results in additional components which are laying in the image plane. A comparison with the analytic signal supports the interpretation that by the monogenic signal three orthogonal components are represented and that the complex domain is generalized in a certain way. In fact, it is useful to imagine a complex plane, the phase plane, oriented by the angle θ in $\mathbb{R}^3$ and spanned by the original signal $\boldsymbol{f}$ and the monogenic signal $\boldsymbol{f}_M$. According to (6), any phase angle indicates a certain local symmetry. The Riesz transform is a rotationally invariant and spinor valued operator, which is identical with the first order circular harmonic. This operator rotates the original signal $\boldsymbol{f}(\boldsymbol{x}) = f(\boldsymbol{x})\boldsymbol{e}_3$ to $\boldsymbol{f}_M(\boldsymbol{x})$ by introducing the additional components $\boldsymbol{f}^r_{(1)}$ and $\boldsymbol{f}^r_{(2)}$. All three components constitute a Riesz triple, that is they are in quadrature phase. Figure 2 visualizes the effect of the Riesz transform.

It delivers not only the right local spectral representations (in contrast to the Gabor filter) but in addition the orientation angle of the local structure. This is only half of the story, because the complete solution of the Dirac equation for the open half space $x_{n+1} > 0$ gives rise to two additional transformations, represented by the scalar valued Poisson kernel

$$p(\boldsymbol{x};\sigma) = \frac{a_n \sigma}{|\boldsymbol{x} + \sigma \boldsymbol{e}_{n+1}|^{n+1}} \quad \text{and} \quad P(\boldsymbol{u};\sigma) = \exp(-2\pi|\boldsymbol{u}|\sigma) \tag{10}$$

and by the bivector valued conjugate Poisson kernel,

$$q(\boldsymbol{x};\sigma) = \frac{a_n \boldsymbol{x}\boldsymbol{e}_{n+1}}{|\boldsymbol{x} + \sigma e_{n+1}|^{n+1}} \quad \text{and} \quad Q(\boldsymbol{u};\sigma) = R_n(\boldsymbol{u})P(\boldsymbol{u},\sigma), \tag{11}$$

with $|\boldsymbol{u}|$ being the absolute frequency value and $\sigma \equiv x_{n+1}$ being a scale parameter from which both kernels are parametrically depending on, as indicated

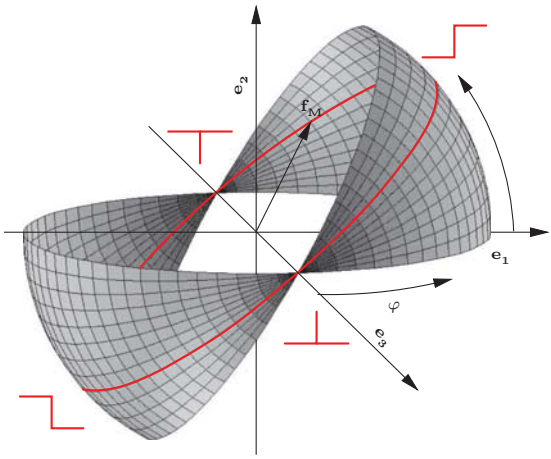


Fig. 2. The embedding of the monogenic signal in $\mathbb{R}^3$

by the semicolon in (10) and (11). Hence, $(\boldsymbol{x}, \sigma \boldsymbol{e}_{n+1})$ is spanning a linear and isotropic scale-space. The representation

$$\boldsymbol{f}_M(\boldsymbol{x}, \sigma) = \boldsymbol{f}_p(\boldsymbol{x}, \sigma) + \boldsymbol{f}_q(\boldsymbol{x}, \sigma), \tag{12}$$

with

$$\boldsymbol{f}_p(\boldsymbol{x}, \sigma) = (p(\sigma) * \boldsymbol{f})(\boldsymbol{x}) \tag{13}$$

and

$$\boldsymbol{f}_q(\boldsymbol{x}, \sigma) = (q(\sigma) * \boldsymbol{f})(\boldsymbol{x}) \tag{14}$$

is the monogenic scale-space representation [16] of the original real-valued signal $\boldsymbol{f}(\boldsymbol{x}) = f(\boldsymbol{x})\boldsymbol{e}_{n+1}$. As a consequence, the local spectral representations as well as the local orientation are getting scale-space representations too. Regrettably, the monogenic signal and the monogenic scale-space are only adequate multidimensional generalizations of the analytic signal with respect to i1D structures. Obviously, an enrichment of the used Clifford analysis framework with additional geometric modeling resources is needed. Therefore, we are adopting ideas from differential geometry and its tensor representations in a Clifford analysis framework as outlined in Sect. 3.

2.4 Basics of Differential Geometry

Differential geometry is applied in image modeling since 1980, pioneered by Koenderink and van Doorn [24, 25], and Besl and Jain [2]. An image is assumed to be a smooth surface S embedded in $\mathbb{R}^3$ in an explicit parametric form with respect to a known coordinate system,

$$S(f) = \{(x, y, z) : x = d(u, v), y = e(u, v), z = f(u, v); (u, v) \in \mathbb{R}^2\}. \tag{15}$$

Besides the parametric representation this is similar to the vector field embedding used in Clifford analysis. Also smoothness is required in both approaches. For the sake of simplicity often a Monge patch representation [6] is used. This considerably simplifies the metric tensor $M(\boldsymbol{x})$ of the first fundamental form and the curvature tensor $T(\boldsymbol{x})$ of the second fundamental form [2] for the scalar valued image function $f(\boldsymbol{x})$,

$$M(\boldsymbol{x}) = \begin{pmatrix} 1+f_x^2 & f_x f_y \\ f_x f_y & 1+f_y^2 \end{pmatrix}(\boldsymbol{x}), \quad T(\boldsymbol{x}) = (1+f_x^2+f_y^2)^{-\frac{1}{2}} B(\boldsymbol{x}), \tag{16}$$

with the Hessian matrix

$$B(\boldsymbol{x}) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix}(\boldsymbol{x}). \tag{17}$$

Although only up to second order derivatives are used, the associated tensor structure delivers rich insight with respect to local geometry. There are two curvature measures, the Gaussian curvature $\kappa(\boldsymbol{x})$ and the mean curvature $\mu(\boldsymbol{x})$, which are derived from the Weingarten mapping matrix $W(\boldsymbol{x}) = (M^{-1})T(\boldsymbol{x})$ according to

$$\kappa(\boldsymbol{x}) = \det(W)(\boldsymbol{x}) = \det(M^{-1})\det(T)(\boldsymbol{x}) \tag{18}$$

$$\mu(\boldsymbol{x}) = \frac{1}{2}\mathrm{trace}(W)(\boldsymbol{x}) = \frac{1}{2}\mathrm{trace}(M^{-1}T)(\boldsymbol{x}). \tag{19}$$

The determinant operator and the trace operator deliver the algebraic main invariants of second order tensors. In the Monge patch representation these curvature measures are written as

$$\kappa(\boldsymbol{x}) = \frac{f_{xx}f_{yy} - f_{xy}^2}{(1+f_x^2+f_y^2)^2} \tag{20}$$

$$\mu(\boldsymbol{x}) = \frac{1}{2}\frac{f_{xx}(1+f_y^2) + f_{yy}(1+f_x^2) - 2f_{xy}f_x f_y}{(1+f_x^2+f_y^2)^{\frac{3}{2}}}. \tag{21}$$

2.5 Alternative Recent Tensor Representations

In recent years tensor representations became an attractive tool for different purposes of image analysis, for instance for the analysis of range images in a differential geometric framework [2] or for the representation of 3D surface orientations in the framework of quadrature filters [23]. In [44] the orientations of flow fields in image sequences are estimated. An overview on the use of tensors for analyzing orientated patterns is given in [3]. Finally, [33] present a linear tensor voting technique for salient feature inference. Regrettably, all these proposals are lacking the advantages resulting from a monogenic signal representation. In the following, we summarize some key features of recent

tensor proposals with respect to the invariances mentioned in Sect. 2.2. As we show in Sect. 3, all these invariances are fulfilled in the case of the monogenic curvature tensor. The structure tensor [17] delivers energy and orientation in a rotation invariant manner. But it responds only reasonably to odd symmetric patterns. Thus, phase invariance is missed and no phase information is contained. The orientation tensor [4] can be also interpreted as a structure tensor, although a set of directed quadrature filters constitutes its elements. Nevertheless, it delivers no phase information. It is restricted to i1D structures but is phase invariant. The energy tensor [13] represents products of first- and second-order derivatives of a bandpass filtered image representation. Although it is invariant with respect to intrinsic dimension its phase invariance is restricted [27]. It represents no phase information. The gradient energy tensor [27] is composed by an even and an odd part. It delivers an energy measure that is invariant with respect to the intrinsic dimension but is not phase invariant. Most interestingly in our context is the boundary tensor [26]. As has been shown in [27], this tensor has the interesting property that its odd part is the Riesz transform of the even part. Nevertheless, although from that construction several invariances result (intrinsic dimension, phase, rotation), it represents an energy measure and thus fails to represent phase.

3 Monogenic Curvature Tensor and its Evaluation

In this section, we derive the monogenic curvature tensor by merging the concept of the monogenic signal with the Hessian matrix from differential geometry. We evaluate the monogenic curvature tensor with respect to its determinant and trace. While the trace results in a representation of i1D signals, which is identical to the monogenic signal, the determinant delivers a novel signal representation for i2D signals. We evaluate that signal representation with respect to local spectral representations and geometric features. This set of features will posses all invariance requirements formulated in Sect. 2.2. Because the monogenic curvature tensor will be embedded in a monogenic scale-space, all derived features will get the same scale-space embedding.

3.1 Monogenic Curvature Tensor

Our approach of deriving the monogenic curvature tensor [48, 49] is based on lifting up a monogenic signal $\boldsymbol{f}_M \in \mathbb{R}_3$ into a tensor representation associated to the curvature tensor of differential geometry. Instead of taking the complete curvature tensor we restrict ourselves to the Hessian matrix, (17). We call the resulting tensor, $T(\boldsymbol{x}) : \mathbb{R}^2 \rightarrow M(2, \mathbb{R}_3)$, the monogenic curvature tensor although it is different from the curvature tensor in (16). The matrix geometric algebra $M(2, \mathbb{R}_3)$, see [39], is a much more powerful algebraic framework than the Euclidean geometric algebra $\mathbb{R}_3$, which is used in Sect. 2.3. This can be concluded from the isomorphism $M(2, \mathbb{R}_{p,q}) \cong \mathbb{R}_{p+1,q+1}$ [31]; hence,

$M(2,\mathbb{R}_3) \cong \mathbb{R}_{4,1}$. The Hessian matrix with elements in $\mathbb{R}_3$, $B(\boldsymbol{x}) \in M(2,\mathbb{R}_3)$, applied to the original signal $\boldsymbol{f}(x)$ reads

$$B(\boldsymbol{x}) = \begin{pmatrix} f_{xx}\boldsymbol{e}_3 & -f_{xy}I_3 \\ f_{xy}I_3 & f_{yy}\boldsymbol{e}_3 \end{pmatrix}, \tag{22}$$

with $I_3 = \boldsymbol{e}_1\boldsymbol{e}_2\boldsymbol{e}_3$ being the unit pseudoscalar of $\mathbb{R}_3$. This results from a tensorial convolution, $*_\tau$, of the signal $\boldsymbol{f}(\boldsymbol{x}) = f(\boldsymbol{x})\boldsymbol{e}_3$ with an even symmetric Hessian operator $\mathcal{H}_e \in M(2,\mathbb{R}_3^+)$, written as

$$h_e(\boldsymbol{x}) = \begin{pmatrix} \partial_{xx} & -\partial_{xy}\boldsymbol{e}_{12} \\ \partial_{xy}\boldsymbol{e}_{12} & \partial_{yy} \end{pmatrix}(\boldsymbol{x}), \tag{23}$$

and the definition of the first order derivative of the vector field $\boldsymbol{f}$, $\partial\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{e}_1\partial_x f(x,y)\boldsymbol{e}_3 + \boldsymbol{e}_2\partial_y f(x,y)\boldsymbol{e}_3 = f_x\boldsymbol{e}_{13} + f_y\boldsymbol{e}_{23}$. The minus sign in this equation results from the noncommutativity of the geometric product, hence, $\boldsymbol{e}_{21} = -\boldsymbol{e}_{12}$. Obviously, the elements of the Hessian operator are either scalars or bivectors. Both multivector types belong to the so-called even subalgebra of $\mathbb{R}_3$, which possesses only even-grade elements, as is indicated by $\mathbb{R}_3^+$. The term even symmetric as characterization of the Hessian operator follows from the fact that second order derivative operators are of even symmetry (in contrast to first order derivative operators which are of odd symmetry) and thus are responding to even-symmetric structures as lines. In case of odd symmetric structures their response would be zero. But we want to apply that operator on the monogenic signal $\boldsymbol{f}_M$ instead, and the result will be called the monogenic curvature tensor, T_M. Then it follows

$$T_M = \mathcal{H}_e\boldsymbol{f}_M = \mathcal{H}_e(\boldsymbol{f} + \boldsymbol{f}^r) = \mathcal{H}_e(\mathcal{I} + \mathcal{R})\boldsymbol{f} \tag{24}$$

with the identity operator $\mathcal{I}$ and the operator of the Riesz transform $\mathcal{R}$. According to (9), the function $\boldsymbol{f}^r$ indicates the monogenic completion of the function $\boldsymbol{f}$ resulting from the convolution of $\boldsymbol{f}$ with the Riesz kernel r_2, see (7). Because $\boldsymbol{f}_M \in C^\infty(\Omega)$ with Ω as an open region in $\mathbb{R}^2$, the partial derivatives of $\mathcal{H}_e$ will be applied without problems. In addition, suppose $\boldsymbol{f} \in L_2(\mathbb{R}^2,\mathbb{R}_3)$, then the commutativity $\mathcal{H}_e\mathcal{R} = \mathcal{R}\mathcal{H}_e$ will follow, see [9] for more details. So we can also formulate an odd Hessian operator $\mathcal{H}_o = \mathcal{R}\mathcal{H}_e$ and the monogenic curvature tensor will be composed by an even and an odd part,

$$T_M = (\mathcal{H}_e + \mathcal{H}_o)\boldsymbol{f} = T_e + T_o = T_e + \mathcal{R}T_e. \tag{25}$$

To get these relations more explicitly, we are going to the Fourier domain and express the operator $\mathcal{H}_e$ by polar coordinates $\boldsymbol{u} = (\varrho,\alpha)$. Then the even Hessian operator is separable,

$$H_e(\varrho,\alpha) = -\frac{4\pi^2\varrho^2}{2}\begin{pmatrix} 1+\cos(2\alpha) & -\sin(2\alpha)\boldsymbol{e}_{12} \\ \sin(2\alpha)\boldsymbol{e}_{12} & 1-\cos(2\alpha) \end{pmatrix}. \tag{26}$$

While the radial part $H_e(\varrho)$ expresses the well known highpass characteristics according to the derivative theorem of Fourier theory, the angular part

$H_e(\alpha)$ enables the investigation of local geometry. The angular part is written in terms of trigonometric functions according to the angular components of derivatives written in polar coordinates in spectral domain. These matrix entries can be related to the Fourier representation of circular harmonics of order $m \geq 0$,

$$C_m(\alpha) = \exp(m\alpha \boldsymbol{e}_{12}) = \cos(m\alpha) + \sin(m\alpha)\boldsymbol{e}_{12}. \tag{27}$$

Their radial part is constant, $C_m(\varrho) = \text{const}$. We want to express the Hessian operator in terms of regularized derivatives. This means to convert the highpass characteristic of $H_e(\varrho)$ to a bandpass characteristic. Because we are operating our signal analysis in a monogenic scale-space, the Poisson kernel, (10), can be used to define a difference of Poisson (DOP) kernel [16]

$$H_{DOP}(\varrho; \sigma_{BP}) = P(\varrho; \sigma_f) - P(\varrho; \sigma_c), \tag{28}$$

with σ_f being a fine scale and σ_c being a coarse scale so that H_{DOP} will have its maximum at $\varrho_{BP} = \frac{1}{\sigma_{BP}}$. As a consequence, we are considering damped circular harmonic functions of order m according to

$$C_m^P(\varrho, \alpha; \sigma) = H_{DOP}(\varrho; \sigma) C_m(\alpha)\ . \tag{29}$$

From this follows for the even Hessian operator $H_e(\varrho, \alpha; \sigma) = H_e(\varrho; \sigma) H_e(\alpha)$, represented in terms of circular harmonics as

$$H_e(\varrho, \alpha; \sigma) = \frac{1}{2} \begin{pmatrix} C_0^P + \langle C_2^P \rangle_0 & -\langle C_2^P \rangle_2 \\ \langle C_2^P \rangle_2 & C_0^P - \langle C_2^P \rangle_0 \end{pmatrix} (\varrho, \alpha; \sigma). \tag{30}$$

Here $\langle\ \ \rangle_0$ means the scalar part and $\langle\ \ \rangle_2$ means the bivector part of C_2^P. The odd Hessian operator follows simply as the Riesz kernel transformed even one and by remembering that the Riesz transform corresponds to a circular harmonic of first order. Hence,

$$H_o(\varrho, \alpha; \sigma) = (C_1^P \times_\tau H_e)(\varrho, \alpha; \sigma), \tag{31}$$

with X_τ indicating the pointwise product of the tensor elements.

We are calling the operator $\mathcal{H}_M = \mathcal{H}_e + \mathcal{H}_o$ the monogenic Hessian operator, which applied to the original signal $\boldsymbol{f}$ results in the monogenic curvature tensor. What remains open is the discussion of its angular part $H_M(\alpha)$. As discussed in detail in [48, 49], it can be interpreted as a rotation invariant detector of i1D structures superimposed with arbitrary angles of intersection. This can be seen from rewriting the angular part of H_e as

$$H_e(\alpha) = \begin{pmatrix} \cos^2(\alpha) & -\frac{1}{2}\sin(2\alpha)\boldsymbol{e}_{12} \\ \frac{1}{2}\sin(2\alpha)\boldsymbol{e}_{12} & \sin^2(\alpha) \end{pmatrix}. \tag{32}$$

The two functions on the diagonal and that one on the anti-diagonal, $\sin(2\alpha) = \cos^2(\alpha - \frac{\pi}{4}) - \sin^2(\alpha - \frac{\pi}{4})$, constitute four basis functions to steer a detector

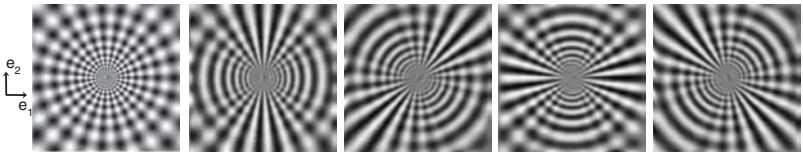


Fig. 3. Action of the even angular windowing functions on the left most test image

for even symmetric i1D structures. In Fig. 3 their effect on the pattern to the left is shown. The same arguments lead to the interpretation of the entries of $H_o(\alpha)$ as rotationally invariant detector of odd symmetric i1D structures.

From this interpretation of the monogenic Hessian operator follows that it analyzes i2D structures as superposition of i1D structures.

3.2 Evaluation of the Monogenic Curvature Tensor

In this section, we give a short sketch on the ways the structural information of the monogenic curvature tensor can be evaluated. More details can be found in [48]. As a result of the preceding section, we can state the invariance of this representation with respect to scale, parity symmetry, and rotation. Here we show the invariance with respect to intrinsic dimension and the derivation of local spectral features as well as geometric features. An alternative analysis of the monogenic curvature tensor is given in the report [51].

3.2.1 Local Representations for i1D and i2D Structures

As is known from real valued differential geometry [2], structures of different intrinsic dimension can be classified in a space spanned by Gaussian curvature, κ, and mean curvature, μ. Both are computed from the curvature tensor by applying either the determinant or the trace operator. If we are doing the same in our algebraic framework $M(2, \mathbb{R}_3)$, we get two different signal representations, which are specific to the intrinsic dimension. The first one is a vector field specific to i1D structures,

$$\boldsymbol{f}_{i1D}(\boldsymbol{x}) = \boldsymbol{t}_e(\boldsymbol{x}) + \boldsymbol{t}_o(\boldsymbol{x}) = \mathrm{trace}(T_e)(\boldsymbol{x}) + \mathrm{trace}(T_o)\boldsymbol{e}_2(\boldsymbol{x}) \tag{33}$$

$$= \boldsymbol{f}(\boldsymbol{x}) + (c_1 * \boldsymbol{f})(\boldsymbol{x}) \equiv \boldsymbol{f}_M(\boldsymbol{x}). \tag{34}$$

That is, the trace operation reconstructs the monogenic signal from the monogenic curvature tensor. Here c_1 is the first-order circular harmonic, which is identical to the Riesz transform. By computing the determinant of $T(\boldsymbol{x})$ another vector field specific to i2D structures will result.

$$\boldsymbol{f}_{i2D}(\boldsymbol{x}) = \boldsymbol{d}_e(\boldsymbol{x}) + \boldsymbol{d}_o(\boldsymbol{x}) = \det(T_e)\boldsymbol{e}_3(\boldsymbol{x}) + \boldsymbol{e}_1\det(T_o)(\boldsymbol{x}) \tag{35}$$

$$= \boldsymbol{d}_e(\boldsymbol{x}) + (\boldsymbol{e}_1 c_2 \boldsymbol{e}_3 * \boldsymbol{d}_e)(\boldsymbol{x}) \equiv \boldsymbol{f}_{MC}(\boldsymbol{x}). \tag{36}$$

Note that the computation of the determinant in the algebraic framework $M(2,\mathbb{R}_3)$ is in general rather expensive, see [39]. But a thorough analysis of our case [48] did yield the same computation as in $M(2,\mathbb{R})$. This signal representation is called "generalized monogenic" curvature signal because its conjugate harmonic part results from the real part by applying a second-order circular harmonic as another generalized Hilbert transform [8]. While the structure of both signal representations is the same, they are coding quite different properties of a local signal structure. They enable classification with respect to the intrinsic dimension.

3.2.2 Local Spectral Representations

Both signal representations can be interpreted as the result of a spinor valued operator, $\mathcal{S}$, which rotates and scales the original vector field $\boldsymbol{f}(\boldsymbol{x}) = f(x,y)\boldsymbol{e}_3$ so that it will be supplemented by a conjugate harmonic component, which projects to the plane $\boldsymbol{e}_1 \wedge \boldsymbol{e}_2$ and fulfills the conditions $\boldsymbol{t}_e^2 = \boldsymbol{t}_o^2$ and $\boldsymbol{d}_e^2 = \boldsymbol{d}_o^2$. The scaling-rotation is performed in the "phase plane" expressed by the outer product $\boldsymbol{f}_s(\boldsymbol{x}) \wedge \boldsymbol{e}_3 = \langle \boldsymbol{e}_3 \boldsymbol{f}_s(\boldsymbol{x})\rangle_2$, with $s(\boldsymbol{x}) = \boldsymbol{e}_3\boldsymbol{f}_s(\boldsymbol{x})$ being the respective spinor and $\boldsymbol{f}_s \equiv \boldsymbol{f}_{i1D}$ or $\boldsymbol{f}_s \equiv \boldsymbol{f}_{i2D}$. That is, for both $\boldsymbol{f}_{i1D}$ and $\boldsymbol{f}_{i2D}$ a similar model for the monogenic extension of the real valued function $\boldsymbol{f}$ is assumed. Only the involved generalized Hilbert transform differs. Therefore, also the computation of the local features in both cases is the same. By evaluating the exponential representation of s with respect to the $\mathbb{R}_3^+$-logarithm, see [12] and [42] for more details, the local features can be computed. These are the "local amplitude"

$$a(\boldsymbol{x}) = |\boldsymbol{f}_s(\boldsymbol{x})| = \exp(\langle \log(\boldsymbol{e}_3\boldsymbol{f}_s(\boldsymbol{x}))\rangle_0) \tag{37}$$

and the (generalized) monogenic "local phase" bivector

$$\Phi(\boldsymbol{x}) = \arg(\boldsymbol{f}_s(\boldsymbol{x})) = \langle \log(\boldsymbol{e}_3\boldsymbol{f}_s(\boldsymbol{x}))\rangle_2. \tag{38}$$

From $\Phi(\boldsymbol{x})$ follow the "local phase" $\varphi(\boldsymbol{x})$ as rotation angle within the phase plane,

$$\varphi(\boldsymbol{x}) = |(\Phi(\boldsymbol{x}))^\star| = \arctan\left(\frac{|\langle \boldsymbol{e}_3\boldsymbol{f}_s(\boldsymbol{x})\rangle_2|}{|\langle \boldsymbol{e}_3\boldsymbol{f}_s(\boldsymbol{x})\rangle_0|}\right), \tag{39}$$

and the orientation angle of the phase plane

$$\theta(\boldsymbol{x}) = \frac{\langle \boldsymbol{e}_3\boldsymbol{f}_s(\boldsymbol{x})\rangle_2}{|\langle \boldsymbol{e}_3\boldsymbol{f}_s(\boldsymbol{x})\rangle_2|}\,. \tag{40}$$

In fact, the terms in quote marks corresponds to the local spectral representation only in case of f_{i1D}. For f_{i2D} this is not the case because of a missing signal model. Nevertheless, the features obey the interesting invariance properties.

Here the star in (39) indicates the duality operation, which converts the local phase bivector $\Phi(\boldsymbol{x})$ into the local rotation vector $(\Phi(\boldsymbol{x}))^\star$, the magnitude of which is the local phase $\varphi(\boldsymbol{x})$. The interpretation of the local spectral

representations and of the local orientation derived from the monogenic signal is well known for i1D structures. I2D structures will be also seen from the monogenic signal as i1D structures. Hence, it delivers mean orientation for any phases of the contributing i1D structures and mean phase only in case of equal phases of the contributing i1D structures. But this is no serious restriction in reality. The interpretation of the features derived from the generalized monogenic curvature signal is not so obvious. As has been discussed in Sect. 3.1, a first hint results from the interpretation of the monogenic Hessian operator in Fourier domain. That is, i2D patterns are seen by the operator as superposition of i1D patterns and each of them contributes to the monogenic curvature tensor representation. Hence, the i1D-view on the structure as known from the monogenic signal is not left in case of the monogenic curvature tensor representation. But a more detailed analysis of this fact in Fourier domain has not been done. Instead, a careful analysis of the monogenic curvature tensor in the Radon domain [47] has been performed. The results confirm that the monogenic curvature tensor in the case of superimposed i1D structures delivers mean phase and mean orientation, just as the monogenic signal itself. Therefore, the proposed signal representation seems to be not a sufficient extension for the extraction of local spectral features of i2D structures. Indeed, the alternative method of analyzing the monogenic curvature tensor, reported in [51], overcomes the interpretation problem. Nevertheless, several applications confirm the usefulness of the proposed signal representation, for example, as corner detector or as phase-based constraint in image sequence analysis. It can be shown that the determinant of the even part of the monogenic curvature tensor equals zero iff the underlying signal structure is of intrinsic dimension one. These results are illustrated in Fig. 4 and can be used as a rotational invariant corner detector.

For further results the reader is advised to [51] and to have a look on our website `http://www.ks.informatik.uni-kiel.de/`.

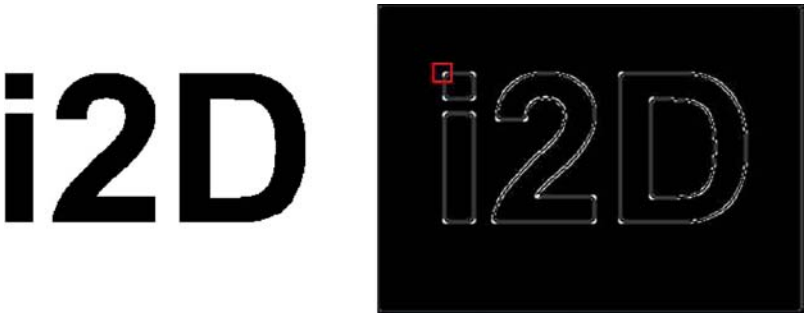


Fig. 4. Original image and local information by the determinant of the even part of the monogenic curvature tensor. One representative i2D corner is marked by the small rectangle

3.2.3 The Angle of Intersection

The monogenic curvature tensor enables to compute the angle of intersection or apex angle of two superimposed i1D signals. To calculate the angle of intersection or apex angle $\alpha \in \left[0, \frac{\pi}{2}\right]$ of two superimposed i1D signals where each one of them can have arbitrary phase in terms of the monogenic signal, the following trick is to consider the resulting image structure as one locally intrinsic i2D hyperbolic saddle point model with *absolute* main curvature values. This model is completely described by its two main curvatures κ_1 and κ_2, which are related to the Gaussian curvature $\kappa = \kappa_1\kappa_2$ and the mean curvature $\mu = \frac{1}{2}(\kappa_1 + \kappa_2)$. The two main curvatures lie on the two orthogonal bisectors of the superimposed i1D signals. Only in direction of the two i1D signals the curvature $\kappa\left(\frac{\alpha}{2}\right)$ of the normal cut is zero. Now the surface theoretical results of Euler's and Meusnier's theorems [5]

$$\kappa\left(\frac{\alpha}{2}\right) = \lim_{\nu \to 0} \kappa_1 \cos^2\frac{\alpha}{2} + 2\nu \sin\frac{\alpha}{2}\cos\frac{\alpha}{2} + \kappa_2 \sin^2\frac{\alpha}{2} \tag{41}$$

can be used to determine the apex angle of our assumed model

$$\kappa\left(\frac{\alpha}{2}\right) = 0 \Rightarrow \alpha = 2\tan^{-1}\sqrt{\frac{|\kappa_1|}{|\kappa_2|}}, \tag{42}$$

where $\frac{\alpha}{2}$ is the angle relatively to the orientation with main curvature κ_1. Merging i1D signal theory and differential geometry delivers the exact apex angle

$$\alpha = 2\tan^{-1}\frac{|\varphi_1'|}{|\varphi_2'|}. \tag{43}$$

Here φ_i' is the i1D phase change at the point of interest in the direction of the main curvatures κ_i $\forall i \in \{1, 2\}$, which have the same orientation as the

Fig. 5. Original image and local angle of intersection information. The monogenic curvature tensor is able to differ between inner and outer corners. Both the black corner with intersection angle of 270° as well as the white corner with intersection angle of 90° are marked by a rectangle

two orthogonal main orientations of the i2D image structure. Applying this result to the monogenic curvature tensor in scale-space, the apex angle can be computed by

$$\alpha(\boldsymbol{x}) = 2\tan^{-1}\sqrt{\left|\frac{\boldsymbol{t}_e(\boldsymbol{x}) - \sqrt{\boldsymbol{t}_e^2(\boldsymbol{x}) - \boldsymbol{d}_e(\boldsymbol{x})}}{\boldsymbol{t}_e(\boldsymbol{x}) + \sqrt{\boldsymbol{t}_e^2(\boldsymbol{x}) - \boldsymbol{d}_e(\boldsymbol{x})}}\right|}. \tag{44}$$

Each i2D corner can be locally modeled by two superimposed i1D structures in scale space. The computation of the angle of intersection is illustrated in Fig. 5 for each test point within the image. At points of intrinsic dimension one the angle of intersection modulo 180° is zero. Therefore, the intrinsic dimension can be naturally determined by the angle of intersection. Note that in Fig. 5, inner and outer corners can also be separated and detected by our approach in a rotational invariant way.

4 Conclusions

This chapter presents a novel approach for local analysis of images. The proposed monogenic curvature tensor results from combining differential geometry and Clifford analysis in the setting of image analysis. The monogenic extension of a two-dimensional signal is lifted up to a new generalization of the analytic signal in a nonlinear way. The generalized monogenic curvature signal can be evaluated just in the same way as the monogenic signal. But as has been shown in other work, the extracted features do no leave the i1D view on the signal. Therefore, future studies are required to cope with i2D structures and to derive meaningful local features. One way could be to include the Weingarten mapping in the signal representation. Another way could be the extension of the Radon transform for curved lines and to generalize the monogenic signal in that way. We followed that way and recently we proposed a conformal monogenic signal [46], [45], in which lines as i1D structures and circles as i2D structures are handled in the same framework.

References

[1] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speed up robust features. In A. Leonardis, M. Bischof, and A. Pinz, editors, *ECCV*, volume 3951 of *LNCS*, Springer-Verlag, Berlin, Heidelberg, New York, pp. 404–417, 2006.

[2] P. J. Besl and R. C. Jain. Invariant surface characteristics for 3D object recognition in range images. *Computer Vision, Graphics, and Image Processing*, 33(1):33–80, 1986.

[3] J. Bigun. *Vision with Direction*. Springer-Verlag, Berlin, Heidelberg, New York, 2006.

[4] J. Bigun and G. H. Granlund. Optimal detection of linear symmetry. In *Proceedings of IEEE First International Conference on Computer Vision*, pp. 433–438, 1984.
[5] W. Blaschke and K. Leichtwei. *Elementare Differentialgeometrie*. Springer-Verlag, Berlin, Heidelberg, New York, 1973.
[6] E. D. Bloch. *A First Course in Geometric Topology and Differential Geometry*. Birkhäuser, Boston, 1997.
[7] F. Brackx, R. Delanghe, and F. Sommen. *Clifford Analysis*. Pittman, Boston, 1982.
[8] F. Brackx, B. De Knock, and H. De Schepper. Generalized multidimensional Hilbert transforms in Clifford analysis. *International Journal of Mathematics and Mathematical Sciences*, 2006.
[9] F. Brackx and H. De Schepper. Hilbert-Dirac operators in Clifford analysis. *Chinese Annals of Mathematics*, 26(1):1–14, 2005.
[10] D.A. Danielson. *Vectors and Tensors in Engineering and Physics*. Addison-Wesley Publ. Company, Redwood City, 1992.
[11] C.T.J. Dodson and T. Poston. *Tensor Geometry*. Springer-Verlag, Berlin, Heidelberg, New York, 1997.
[12] M. Felsberg. *Low-Level Image Processing with the Structure Multivector*. PhD thesis, Technical Report No. 2016, Christian-Albrechts-Universität zu Kiel, Institut für Informatik und Praktische Mathematik, 2002.
[13] M. Felsberg and G. H. Granlund. POI detection using channel clustering and the 2D energy tensor. In *Pattern Recognition*, volume 3175 of *LNCS*, Springer-Verlag Berlin, Heidelberg, New York, pp. 103–110, 2004.
[14] M. Felsberg and G. Sommer. The monogenic signal. *IEEE Transactions on Signal Processing*, 49(12):3136–3144, 2001.
[15] M. Felsberg and G. Sommer. The structure multivector. In L. Dorst, C. Doran, and J. Lasenby, editors, *Applications of Geometric Algebra in Computer Science and Engineering*, Birkhäuser, Boston, pp. 437–448, 2002.
[16] M. Felsberg and G. Sommer. The monogenic scale-space: A unifying approach to phase-based image processing in scale-space. *Journal of Mathematical Imaging and Vision*, 21:5–26, 2004.
[17] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Workshop*, pp. 149–155, Interlaken, 1987.
[18] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, 1991.
[19] D. Gabor. Theory of communication. *Journal IEEE*, 93(26):429–457, 1946.
[20] G. H. Granlund and H. Knutsson. *Signal Processing for Computer Vision*. Kluwer Academic Publishers, Dordrecht, 1995.
[21] A. Guiducci. Corner characterization by differential geometry techniques. *Pattern Recognition Letters*, 8:311–318, 1988.

[22] D. Hestense and G. Sobczyk. *Clifford Algebra to Geometric Calculus*. Reidel, Dordrecht, 1984.
[23] H. Knutsson. Representing local structure using tensors. In *Proc. of 6th Scandinavian Conference on Image Analysis*, pp. 244–251, 1989.
[24] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
[25] J. J. Koenderink and A. J. van Doorn. The structure of two-dimensional scalar fields with applications to vision. *Biological Cybernetics*, 33:151–158, 1979.
[26] U. Köthe. Integrated edge and junction detection with the boundary tensor. In *Proceeding of 9th International Conference on Computer Vision*, volume 1, pp. 424–431, 2003.
[27] U. Köthe and M. Felsberg. Riesz transform vs. derivatives: On the relationship between the boundary tensor and the energy tensor. In R. Kimmel, N. Sochen, and J. Weickert, editors, *Scale Space and PDE Methods in Computer Vision*, volume 3459 of *LNCS*, Springer-Verlag, Berlin, Heidelberg, New York, pp. 179–191, 2005.
[28] P. Kovesi. *Invariant Measures of Image Features from Phase Information*. PhD thesis, University of Western Australia, Institute of Computer Science, Perth, 1996.
[29] G. Krieger and C. Zetzsche. Nonlinear image operators for the evaluation of local intrinsic dimensionality. *IEEE Transactions on Image Processing*, 5(6):1026–1042, 1996.
[30] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.
[31] P. Lounesto. *Clifford Algebras and Spinors*. London Mathematical Society Lecture Note Series. Cambridge University Press, 1997.
[32] D. Lowe. Distinctive image features from scale-invariant keypoints, cascade filtering approach. *International Journal of Computer Vision*, 60:91–110, 2004.
[33] G. Medioni, M. S. Lee, and C. K. Tang. *A Computational Framework for Segmentation and Grouping*. Elsevier, 2000.
[34] M. Michaelis and G. Sommer. Junction classification by multiple orientation detection. In *ECCV*, volume 801 of *LNCS*, Springer-Verlag, Berlin, Heidelberg, New York, pp. 101–108, 1994.
[35] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1615–1630, 2005.
[36] K. Rangarajan, M. Shah, and D. Van Brackle. Optimal corner detector. *Computer Vision, Graphics and Image Processing*, 48:230–245, 1989.
[37] K. Rohr. Recognizing corners by fitting parametric models. *International Journal of Computer Vision*, 9(3):213–230, 1992.
[38] K. Rohr. Localization properties of direct corner detectors. *Journal of Mathematical Imaging and Vision*, 4:133–150, 1994.

[39] G. Sobczyk and G. Erlebacher. Hybrid matrix geometric algebra. In H. Li, P. J. Olver, and G. Sommer, editors, *Computer Algebra and Geometric Algebra with Applications*, volume 3519 of *LNCS*, Springer-Verlag, Berlin, Heidelberg, New York, pp. 191–206, 2005.

[40] G. Sommer. *Geometric Computing with Clifford Algebras*. Springer-Verlag, Berlin, Heidelberg, New York, 2001.

[41] G. Sommer and C. Gebken. Robot vision in the language of geometric algebra. In G. Obinata and A. Dutta, editors, *Vision Systems: Applications*, I-Tech Education and Publishing, Vienna, pp. 459–486, 2007.

[42] G. Sommer and D. Zang. Parity symmetry in multi-dimensional signals. *Communications in Pure and Applied Analysis*, 6(3):829–852, 2007.

[43] I. Stuke, T. Aach, E. Barth, and C. Mota. Analysing superimposed oriented patterns. In *6th IEEE Southwest Symposium on Image Analysis and Interpretation, Lake Tahoe, NV*, pp. 133–137, 2004.

[44] C. F. Westin. *A Tensor Framework for Multidimensional Signal Processing*. PhD thesis, Linköping University, Sweden, SE-581 83 Linköping, Sweden, 1994. Dissertation No 348, ISBN 91-7871-421-4.

[45] L. Wietzke, O. Fleischmann, and G. Sommer. 2D image analysis by generalized Hilbert transforms in conformal space. In *ECCV*, volume 5303 of *LNCS*, Springer-Verlag, Berlin, Heidelberg, New York, pp. 638–649, 2008.

[46] L. Wietzke and G. Sommer. The conformal monogenic signal. In G. Rigoll, editor, *Pattern Recognition*, volume 5096 of *LNCS*, Springer-Verlag, Berlin, Heidelberg, New York, pp. 527–536, 2008.

[47] L. Wietzke, G. Sommer, C. Schmaltz, and J. Weickert. Differential geometry of monogenic signal representations. In G. Sommer and R. Klette, editors, *Robot Vision*, volume 4931 of *LNCS*, Springer-Verlag, Berlin, Heidelberg, New York, pp. 454–465, 2008.

[48] D. Zang. *Signal Modeling for Two-dimensional Image Structures and Scale-Space Based Image Analysis*. PhD thesis, Technical Report No. 0705, Christian-Albrechts-Universität zu Kiel, Institut für Informatik, 2007.

[49] D. Zang and G. Sommer. Signal modeling for two-dimensional image structures. *Journal of Visual Communication and Image Representation*, 18(1):81–99, 2007.

[50] C. Zetzsche and E. Barth. Fundamental limits of linear filters in the visual processing of two-dimensional signals. *Vision Research*, 30:1111–1117, 1990.

[51] L. Wietzke and G. Sommer. The 2D Analytic signal. Technical Report No. 0802, Christian-Albrechts-Universität zu Kiel, Institut für Informatik, 2008.

A General Structure Tensor Concept and Coherence-Enhancing Diffusion Filtering for Matrix Fields

Bernhard Burgeth, Stephan Didas, and Joachim Weickert

Saarland University, Mathematical Image Analysis Group,
Faculty of Mathematics and Computer Science, Bldg. E1 1,
66041 Saarbrücken, Germany
{burgeth,didas,weickert}@mia.uni-saarland.de

Summary. Coherence-enhancing diffusion filtering is a striking application of the structure tensor concept in image processing. The technique deals with the problem of completion of interrupted lines and enhancement of flow-like features in images. The completion of line-like structures is also a major concern in diffusion tensor magnetic resonance imaging (DT-MRI). This medical image acquisition technique outputs a 3D matrix field of symmetric 3×3-matrices, and it helps to visualize, for example, the nerve fibers in brain tissue. As any physical measurement, DT-MRI is subjected to errors causing faulty representations of the tissue corrupted by noise and with visually interrupted lines or fibers.

In this chapter, we address that problem by proposing a coherence-enhancing diffusion filtering methodology for matrix fields. The approach is based on a generic structure tensor concept for matrix fields that relies on the operator-algebraic properties of symmetric matrices, rather than their channel-wise treatment of earlier proposals.

Numerical experiments with artificial and real DT-MRI data confirm the gap-closing and flow-enhancing qualities of the technique presented.

1 Introduction

Coherence-enhancing diffusion (CED) filtering has been introduced in [25, 26] as an image restoration technique that enhances flow-like structures in scalar and vector-valued images. It regularizes images polluted by noise, and it is capable of closing gaps in line-like structures. CED-filtering of a scalar image f defined on the image domain $\Omega \subset \mathbb{R}^d$ produces simplified versions $u(\cdot, t)$ of f as solutions of the partial differential equation (PDE)

$$\begin{aligned} \partial_t u - \operatorname{div}\,(D \cdot \nabla u) &= 0 \qquad \text{in } \Omega \times I, \\ \partial_n u &= 0 \qquad \text{on } \partial\Omega \times I, \\ u(x,0) &= f(x) \qquad \text{in } \Omega, \end{aligned} \tag{1}$$

where $I = [0, T]$ is a potentially unbounded time interval, and $\partial_n u$ denotes the outer normal derivative of u at the boundary $\partial\Omega \times I$.

In view of the qualities mentioned, it would be desirable to have this method at our disposal for matrix-valued images, or *matrix fields* in short. For example, such a filter could serve as a pre-processing step for the so-called fiber tracking in diffusion tensor magnetic resonance imaging (DT-MRI) (see [28], Chapters by Cook et al. and Yushkevich et al. as well as the literature cited therein). In parts of the brain, e.g. the corpus callosum, nerve fibers form bundles with a coherent structure. In principle, the matrix field produced by DT-MRI allows for the application of fiber tracking techniques and hence for the accurate visualization of the nerve fibers. However, ubiquitous measurement errors during acquisition cause gaps and interruptions in the fibers' visualization. Therefore, a fiber tracking technique would benefit from a prior filtering by a matrix-valued counterpart of CED. It is the goal of this chapter to extend CED and the underlying structure tensor concept to matrix-valued images.

Promising proposals to generalize nonlinear regularization methods and related diffusion filters for scalar images to matrix fields have been made in [7, 8]. These approaches are based on a basic differential calculus for matrix fields. These concepts had direct implications for the chapters by Steidl et al. and Lie et al., and they will be useful in the context of this chapter as well. Edge-enhancing diffusion, EED, a concept related to CED has been extended to the setting of matrix fields in [5]. Other PDE-based methods for the regularization of matrix fields have been proposed in [10, 11, 13, 19, 22, 23, 27]. Approaches to tensor field regularization with a more differential geometric background are explored in [2, 18, 20], where the set of positive definite matrices is endowed with a Riemannian metric stemming from the DT-MRI field. Diffusion over tensor fields based on Lie groups is considered in [15] and also in the chapter by Gur and Sochen.

The essential ingredient in (1) underlying CED is the *diffusion tensor* D of the scalar image u, which steers the diffusion process: It amplifies diffusion along flow-like structures, and hinders diffusion perpendicular to those patterns. By postponing the detailed construction of D to the subsequent Sect. 2, for now we only remark that it is a function of the structure tensor [14], which is given by

$$S_\rho(u(x)) := G_\rho * \left(\nabla u(x) \cdot (\nabla u(x))^\top\right) = \left(G_\rho * \left(\partial_{x_i} u(x) \cdot \partial_{x_j} u(x)\right)\right)_{i,j=1,\ldots,d} .$$

Here $G_\rho *$ indicates a convolution with a Gaussian of standard deviation ρ, however, more general averaging procedures can be used. If $\nabla u(x) \neq 0$ the matrix

$$\left(\nabla u(x) \cdot (\nabla u(x))^\top\right)$$

has rank one, the eigenvector $\nabla u(x)$ belongs to the only non-zero eigenvalue $|\nabla u(x)|^2$. The eigenvalues represent the contrast in the directions of the eigenspaces. The averaging process then creates a matrix with full rank,

which contains valuable directional information. Note that the averaging of the structure tensor avoids cancellation of directional information. If one would average the gradients instead, neutralization of vectors with opposite sign would occur. In many applications, it is advantageous to use a pre-smoothed image $u_\sigma := G_\sigma * u$ instead of u to reduce the influence of noise for better numerical results. The structure tensor is a classical tool in image processing to extract directional information from an image, going back to [17], for more details the reader is referred to [3] and the literature cited there.

It is not straightforward to generalize both the structure and the diffusion tensor concept to the setting of matrix-valued images.

To fix notation in this work, matrix-valued images or matrix fields $M(x)$ are considered as mappings from $\mathbb{R}^d$ into the set $\mathrm{Sym}_n(\mathbb{R})$ of symmetric $n \times n$-matrices

$$M : \; x \mapsto M = \big(m_{i,j}(x)\big)_{i,j=1,\ldots,n} \in \mathrm{Sym}_n(\mathbb{R})\,.$$

and denoted by capital letters, while indexed lower case letters indicate their components.

Di Zenzo's approach [12] to build a structure tensor for multi-channel images has been generalized in a straightforward manner as follows: each channel considered as independent scalar image gives rise to a structure tensor, and then these structure tensors are summed up to give a standard structure tensor [5, 27]:

$$J_\rho(U(x)) := \sum_{i,j=1}^{n} S_\rho(u_{i,j}(x))\,.$$

This construction has been refined to a customizable structure tensor in [21]. There the resulting structure tensor is a weighted sum of tensors of scalar quantities that are now not just the channels, but other meaningful scalar quantities derived from the matrix field. The weights are provided by the user, and depending on the choice of weights the emerging structure tensor has a sensitivity for certain features of the matrix field. A special constellation of the weights turns the customizable structure tensor into the standard structure tensor from [5, 27]. It is important to mention that in case of a 3D matrix field of 3×3 symmetric matrices, these concepts yield also a 3×3 structure tensor, the very same order as a 3D scalar image.

Here we opt for a different approach: We assume an operator-algebraic view on symmetric matrices as finite dimensional instances of selfadjoint Hilbert space operators. The exploitation of the algebraic properties of matrices, ensures proper interaction between the different matrix channels. This is a decisive advantage over the standard component-wise treatment of vector-valued images.

In this chapter, we will present a general concept for a large size, second order structure tensor, a $nd \times nd$-matrix that carries all the directional information of the matrix field. We will show how this information can be deduced from this full tensor by a reduction process. This reduction process illuminates also its connection to the structure tensors mentioned above.

The chapter is structured as follows: Sect. 2 is devoted to a brief review of coherence enhancing diffusion filtering of scalar images. Notions necessary to construct the diffusion tensor and a basic differential calculus for matrix fields necessary to construct the diffusion tensor are provided in Sect. 3. In Sect. 4 we propose a novel structure tensor concept for matrix fields, study some of its properties by investigating the connection to already known structure tensors for matrix-valued data. We then feature the potential of this concept by proposing a coherence enhancing diffusion for matrix fields in Sect. 5. We report on the results of our experiments with matrix-valued coherence enhancing diffusion applied to synthetic data and real DT-MRI images in Sect. 6. Sect. 7 is made up by concluding remarks.

2 Synopsis of Coherence Enhancing Diffusion

The rationale behind the construction of the diffusion tensor D for grey value images as proposed in [25] is as follows: The matrix $S_\rho(u)$ as the positive average of different symmetric positive semidefinite matrices has the very same property. Hence $S_\rho(u)$ has an orthonormal system $\{w_1, \ldots, w_d\}$ of eigenvectors corresponding to the non-negative eigenvalues $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_d \geq 0$ indicating the contrast in each direction. In the line defined by w_d, the *coherence direction*, the contrast is the least compared to other orientations, since w_d belongs to the smallest eigenvalue μ_d. The coherence or anisotropy of an image structure essentially is captured in the eigenvalue distribution of the structure tensor S_ρ. In 1999 [25], the quantity

$$\kappa := \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} (\mu_i - \mu_j)^2$$

is proposed to measure coherence. Strongly differing eigenvalues result in a large value of κ, while similar values produce a small κ-value indicating a structure with isotropic character. The matrix D has the same eigenvectors as S_ρ, however, its eigenvectors λ_i are altered via the tensor transfer map H according to

$$\lambda_i := H(\mu_i) := \alpha \qquad \text{for } i = 1, \ldots, d-1$$

and

$$\lambda_d := H(\mu_d) := \begin{cases} \alpha & \text{if } \kappa = 0, \\ \alpha + (1-\alpha) \exp\left(-\frac{C}{\kappa}\right) & \text{else.} \end{cases}$$

with a threshold $C > 0$. With this choice of D, the CED-filter (1) smoothes mainly along the coherence direction w_d with a diffusivity λ_d that increases with κ. Note that $\min\{\lambda_i, i = 1, \ldots, d\} = \alpha > 0$, which causes D to be uniformly positive definite and enforces a diffusion no matter how isotropic ($\kappa \downarrow 0$) the image structure becomes.

It is our task in this chapter to extend the notion of structure and diffusion tensor from scalar images to matrix fields. Ultimately we aim at coherence enhancing diffusion filtering of matrix fields. To do so, we have to clarify what is meant by the partial derivative, the gradient and, most of all, the suitable structure tensor of a matrix field. This is done in the next section.

3 Basic Differential Calculus for Matrix Fields

In this section, we provide briefly the *basic definitions* for the formulation of a differential calculus for matrix fields. This material is instigated in [6] but for a more detailed exposition the reader is referred to [7].

1. **Functions of matrices.** The standard definition of a function h on $\mathrm{Sym}_n(\mathbb{R})$ is given in [16]:

$$h(U) = V^\top \mathrm{diag}(h(\lambda_1), \ldots, h(\lambda_n))V \in \mathrm{Sym}_n(\mathbb{R}),$$

   if $U = V^\top \mathrm{diag}(\lambda_1, \ldots, \lambda_n)V$ is the spectral-/eigendecomposition of the symmetric matrix U, and if $\lambda_1, \ldots, \lambda_n$ lie in the domain of definition of h. We encountered already an example of a function of a symmetric matrix; the diffusion tensor as a function of the structure tensor S_ρ with coherence κ under the tensor transfer map H, $D = H(S_\rho)$.
2. **Partial derivatives.** Let $\omega \in \{x_1, \ldots, x_d, t\}$ denote a spatial or temporal variable, and set $(x,t) = (x_1, \ldots, x_d, t)$. The partial derivative for matrix fields is naturally defined *component-wise* as the limit of a difference quotient:

$$\begin{aligned}\overline{\partial}_\omega U(x,t) &= \lim_{h\to 0} \frac{U((x,t) + h \cdot e_k) - U(x,t)}{h} \\ &= \left(\lim_{h\to 0} \frac{u_{ij}((x,t) + h \cdot e_k) - u_{ij}(x,t)}{h} \right)_{i,j} \\ &= \left(\partial_\omega u_{ij}(x,t)\right)_{i,j},\end{aligned}$$

   where $e_k := (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}^{d+1}$ denotes the kth unit vector of space-time $\mathbb{R}^{d+1}$. The generalization to directional derivatives is straightforward. In this case, ω would denote an appropriate unit vector. Higher order partial differential operators, such as the Laplacian or other more sophisticated operators, find their natural counterparts in the matrix-valued framework in this way. It is worth mentioning that for the operators $\overline{\partial}_\omega$ a *product rule* holds:

$$\overline{\partial}_\omega(A(x) \cdot B(x)) = (\overline{\partial}_\omega A(x)) \cdot B(x)) + A(x) \cdot (\overline{\partial}_\omega B(x)).$$

3. **Generalized gradient of a matrix field.** The gradient of a matrix field with sufficiently smooth component functions is defined via

$$\overline{\nabla} U(x) := (\overline{\partial}_{x_1} U(x), \ldots, \overline{\partial}_{x_d} U(x))^\top \in (\mathrm{Sym}_n(\mathbb{R}))^d.$$

Hence, the generalized gradient $\overline{\nabla} U(x)$ at a voxel x is regarded as an element of the module $(\mathrm{Sym}_n(\mathbb{R}))^d$ over $\mathrm{Sym}_n(\mathbb{R})$ in close analogy to the scalar setting, where $\nabla u(x) \in \mathbb{R}^d$. Hence, in the sequel we will call a mapping from R^d into the module $(\mathrm{Sym}_n(\mathbb{R}))^d$ a *module field* rather than a vector field, the later one being a mapping from R^d into a vector space. Note that this definition of a generalized gradient is different from one that might be expected when viewing a matrix as a tensor (of second order). According to differential geometry concepts, their derivatives are tensors of third order. However, we adopt an operator-algebraic point of view: The matrices are self-adjoint operators that can be added, multiplied with a scalar, and concatenated. Thus, they form an algebra, and we aim at consequently replacing the field $\mathbb{R}$ by the algebra $\mathrm{Sym}_n(\mathbb{R})$ in the scalar, that is, $\mathbb{R}$-based formulation of differential calculus.

4. For the sake of completeness, we include the formal definition of the *generalized structure tensor of a matrix field* here. We will discuss its derivation, properties and application in the next section. The novel structure tensor for a matrix field is given by

$$\begin{aligned} \overline{\mathcal{S}}_L\,(U(x)) &:= G_\rho * \left(\overline{\nabla} U(x) \cdot (\overline{\nabla} U(x))^\top\right) \\ &= \left(G_\rho * \left(\overline{\partial}_{x_i} U(x) \cdot \overline{\partial}_{x_j} U(x)\right)\right)_{i,j=1,\ldots,d} . \end{aligned} \tag{2}$$

5. **Symmetrized product of symmetric matrices.** The product of two symmetric matrices $A, B \in \mathrm{Sym}_n(\mathbb{R})$ is not symmetric unless the matrices commute. However, it is vital to our interests to have a symmetrized matrix product at our disposal. There are numerous options to define a symmetrized matrix product, however, we concentrate on a specific one known from algebra and called Jordan product:

$$A \bullet_J B = \frac{1}{2}(AB + BA) \quad \text{for } A, B \in \mathrm{Sym}_n(\mathbb{R})\,. \tag{3}$$

For commuting A and B, we have $A \bullet_J B = A \cdot B$. This product is commutative and distributive but not associative. It is one half of the anti-commutator of A and B, but due to its additive structure no determinant product rule holds. Most important, it does not preserve the positive semidefinitness of its arguments as the following simple example shows:

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \bullet_J \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{2}\left(\begin{pmatrix} 2 & 2 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix}\right) = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} \text{ with } \det \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} = -1\,.$$

Hence, simply multiplying each matrix in a DT-MRI-field by a positive definite matrix employing the Jordan product might produce a matrix field that cannot be interpreted anymore as DT-MRI data, which is not desirable.

Remarks:
(1) It should be mentioned that the logarithmic multiplication introduced in [1] and given by $A \bullet_L B := \exp(\log(A) + \log(B))$ is defined only for positive definite matrices. However, the matrix-valued diffusion proposed here requires the multiplication to be able to cope with the factor matrices being indefinite. Furthermore, matrix fields that are not necessarily positive semidefinite should also be within the reach of our PDE-based filtering. Hence, the logarithmic multiplication is not suitable for our purpose.

(2) The proposed notions for a calculus on symmetric matrix fields are extensions of the calculus of scalar multivariate functions. As such, it must be possible to regain the scalar calculus from the newly introduced matrix-valued framework by specification. There are two ways to view scalar calculus as a special case of the matrix calculus: Clearly, setting $n = 1$ turns the matrix field into a scalar function. However, one can also embed the set of real numbers $\mathbb{R}$ into the set of symmetric matrices $\mathrm{Sym}_n(\mathbb{R})$ by the identification $\mathbb{R} \ni r \longleftrightarrow r \cdot I_n$ with the $n \times n$ identity matrix I_n. Hence, aside from having a certain simplicity, it is mandatory that the proposed extensions collapse to the scalar calculus when making the transition from scalar functions to matrix fields in one way or the other.

We summarized the definitions from above and juxtapose them with their scalar counterparts in the subsequent Table 1. The matrix field $U(x)$ is assumed to be diagonizable with $U = (u_{ij})_{ij} = V^\top \mathrm{diag}(\lambda_1, \ldots, \lambda_n) V$, where $V \in O(n)$, the set of all orthogonal $n \times n$-matrices, and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$.

Table 1. Extensions of elements of scalar valued calculus to the matrix-valued setting

Setting	Scalar valued	Matrix-valued
function	$h : \begin{cases} \mathbb{R} \longrightarrow \mathbb{R} \\ x \mapsto h(x) \end{cases}$	$h : \begin{cases} \mathrm{Sym}_n(\mathbb{R}) \longrightarrow \mathrm{Sym}_n(\mathbb{R}) \\ U \mapsto V^\top \mathrm{diag}(h(\lambda_1), \ldots, h(\lambda_n)) V \end{cases}$
partial derivatives	$\partial_\omega u$, $\omega \in \{t, x_1, \ldots, x_d\}$	$\overline{\partial}_\omega U := (\partial_\omega u_{ij})_{ij}$, $\omega \in \{t, x_1, \ldots, x_d\}$
gradient	$\nabla u(x) := (\partial_{x_1} u(x), \ldots, \partial_{x_d} u(x))^\top$, $\nabla u(x) \in \mathbb{R}^d$	$\overline{\nabla} U(x) := (\overline{\partial}_{x_1} U(x), \ldots, \overline{\partial}_{x_d} U(x))^\top$, $\overline{\nabla} U(x) \in (\mathrm{Sym}_n(\mathbb{R}))^d$
structure tensor	$G_\rho * \left(\overline{\nabla} u(x) \cdot (\overline{\nabla} u(x))^\top \right)$	$\overline{\mathcal{S}}_L (U(x)) := G_\rho * \left(\overline{\nabla} U(x) \cdot (\overline{\nabla} U(x))^\top \right)$
product	$a \cdot b$	$A \bullet_J B := \frac{1}{2}(AB + BA)$

4 The Full Structure Tensor $\overline{\mathcal{S}}_L$ for Matrix Fields

4.1 Derivation of $\overline{\mathcal{S}}_L$

With the terminology introduced earlier, we infer for the directional derivative of $x \mapsto U(x)$ in direction $v \in S^{d-1}$:

$$\begin{aligned}\overline{\partial}_v U(x) &:= \frac{d}{dh} U(x + h \cdot v) \mid_{h=0} = \sum_{k=1}^{d} (\partial_{x_k} u_{ij}(x))\, v_k \\ &= (\overline{\nabla} U(x))^\top \begin{pmatrix} v_1\, I_n \\ \vdots \\ v_d\, I_n \end{pmatrix} \\ &= v_1\, \overline{\partial}_{x_1} U(x) + \cdots + v_d\, \overline{\partial}_{x_d} U(x) \in \mathrm{Sym}_n(\mathbb{R})\end{aligned}$$

This expression is a symmetric matrix but aside from that in complete analogy to the corresponding real-valued term in the scalar case. In the scalar setting, the direction of steepest descent/ascent would be given by the direction that optimizes the directional derivative. However, in the matrix-valued setting, the entries of the generalized gradient are matrices, and finding an optimal unit vector v that optimizes the matrix-valued directional derivative is hindered by practical as well as theoretical obstacles: An optimization relies on the presence of a total ordering, but on $\mathrm{Sym}_n(\mathbb{R})$ only partial ordering relations do exist and even after choosing a partial ordering, e.g. the Loewner ordering [9], it is not clear how to obtain the optimal v in a reasonable computational effort. A remedy for this difficulty is the projection of the set of symmetric matrices $\mathrm{Sym}_n(\mathbb{R})$ onto the real numbers by a so-called linear form. Each linear form on $\mathrm{Sym}_n(\mathbb{R})$ is of the form of a scalar product,

$$M \longmapsto \langle A, M \rangle := \mathrm{tr}_A(M) := \mathrm{tr}(A \cdot M)$$

with a matrix $A \in \mathrm{Sym}_n(\mathbb{R})$, see [16]. Then one simply has

$$\begin{aligned}&\mathrm{argmax}\{\mathrm{tr}_A(\overline{\partial}_v U(x)) \mid \|v\| = 1\} = \\ &\frac{1}{\sqrt{\sum_1^d (\mathrm{tr}_A(\overline{\partial}_{x_i} U(x)))^2}} \left(\mathrm{tr}_A(\overline{\partial}_{x_1} U(x)), \ldots, \mathrm{tr}_A(\overline{\partial}_{x_d} U(x))\right).\end{aligned}$$

We write $\mathrm{tr} = \mathrm{tr}_I$. Depending on the choice of A, we obtain the direction of strongest change at one point x in the matrix field, or which boils down to the same thing: the strongest change in the scalar image $x \longmapsto \mathrm{tr}_A(U(x))$. Apparently this approach suffers from the same weakness as the direct direction estimation in scalar images, the danger of cancellation through averaging. This reveals the need for a structure tensor for matrix fields and also its basic construction principle. It is close at hand to define a structure tensor for matrix fields as follows:

$$\overline{\mathcal{S}}_L\left(U(x)\right) := G_\rho * \left(\overline{\nabla} U(x) \cdot (\overline{\nabla} U(x))^\top\right) = \left(G_\rho * \left(\overline{\partial}_{x_i} U(x) \cdot \overline{\partial}_{x_j} U(x)\right)\right)_{i,j=1,\ldots,d}$$

Here $G_\rho *$ indicate a convolution with a Gaussian of standard deviation ρ or, more general, another appropriate averaging procedure. We list some immediate properties of this construct:

1. $\overline{\mathcal{S}}_L\left(U(x)\right)$ is a symmetric $nd \times nd$-block matrix with d^2 blocks of size $n \times n$, $\overline{\mathcal{S}}_L\left(U(x)\right) \in \mathrm{Sym}_d(\mathrm{Sym}_n(\mathbb{R})) = \mathrm{Sym}_{nd}(\mathbb{R})$. Typically for the 3D medical DT-MRI data one has $d = 3$ and $n = 3$, yielding a 9×9-matrix $\overline{\mathcal{S}}_L$. The symmetry follows from the fact that for all $i, j = 1, \ldots, d$

$$\begin{aligned}\left(\overline{\partial}_{x_i} U(x) \cdot \overline{\partial}_{x_j} U(x)\right)^\top &= \left(\overline{\partial}_{x_j} U(x)\right)^\top \cdot \left(\overline{\partial}_{x_i} U(x)\right)^\top \\ &= \overline{\partial}_{x_j} U(x) \cdot \overline{\partial}_{x_i} U(x)\,.\end{aligned}$$

2. The structure tensor $\overline{\mathcal{S}}_L$ can be diagonalized as

$$\overline{\mathcal{S}}_L\left(U\right) = \sum_{k=1}^{nd} \rho_k w_k w_k^\top$$

with real eigenvalues λ_k (w.l.o.g. arranged in decreasing order) and an orthonormal basis $\{w_k\}_{i=1,\ldots,nd}$ of $\mathbb{R}^{nd}$.

The usefulness of this construction will rise and fall with a positive answer to the following question:

How can we extract useful d-dimensional directional information from this full structure tensor $\overline{\mathcal{S}}_L$?

In the case of a structure tensor $S \in \mathrm{Sym}_d^+(\mathbb{R})$ for a scalar image, its eigenvectors provide all the important directional information of the image. However, in the matrix-valued setting the eigenvectors w_i of $\overline{\mathcal{S}}_L\left(U\right)$ are nd-dimensional vectors and they lack immediate physical interpretation. Hence, as such they do not provide really useful d-dimensional directional information. Instead, let us find an analog to the Rayleigh quotient of S

$$\mathrm{argmax}\,\left\{v^\top \left(G * \nabla u\ \nabla u^\top\right) v \;\middle|\; v \in \mathbb{R}^d,\ \|v\| = 1\right\}$$

in the matrix field framework. We consider

$$(v_1\, I_n \cdots v_d\, I_n)\overline{\mathcal{S}}_L\left(U(x)\right) \begin{pmatrix} v_1\, I_n \\ \vdots \\ v_d\, I_n \end{pmatrix}$$

$$= \begin{pmatrix} v_1^2\ G_\rho * (\overline{\partial}_{x_1} U)^2 & \cdots & v_1 v_d\ G_\rho * (\overline{\partial}_{x_1} U \cdot \overline{\partial}_{x_d} U) \\ \vdots & \ddots & \vdots \\ v_1 v_d\ G_\rho * (\overline{\partial}_{x_d} U \cdot \overline{\partial}_{x_1} U) & \cdots & v_d^2\ G_\rho * (\overline{\partial}_{x_d} U)^2 \end{pmatrix}.$$

Now we have to choose the real numbers $v_1, \ldots, v_d$ in an *optimal* way. Again we are facing difficulties caused by the absence of total ordering for matrices. In the case of a scalar image u (a matrix field with 1×1-matrices), the optimal v is the unit eigenvector corresponding to the largest eigenvalue of the structure tensor S. It will turn out to be convenient to utilize the notion of operator matrix.

The idea is to *reduce* $\overline{\mathcal{S}}_L(U) \in \mathrm{Sym}_{nd}(\mathbb{R})$ to a structure tensor $S(U) \in \mathrm{Sym}_d(\mathbb{R})$ in a generalized projection step employing the block operator matrix

$$\mathrm{Tr}_A := \begin{pmatrix} \mathrm{tr}_A & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathrm{tr}_A \end{pmatrix} \tag{4}$$

containing the trace operation. Again we set $\mathrm{Tr} := \mathrm{Tr}_I$. This operator matrix acts on elements of the space $(\mathrm{Sym}_n(\mathbb{R}))^d$ as well as on block matrices via formal block-wise matrix multiplication,

$$\begin{pmatrix} \mathrm{tr}_A & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathrm{tr}_A \end{pmatrix} \begin{pmatrix} M_{11} & \cdots & M_{1d} \\ \vdots & \ddots & \vdots \\ M_{d1} & \cdots & M_{dd} \end{pmatrix} = \begin{pmatrix} \mathrm{tr}_A(M_{11}) & \cdots & \mathrm{tr}_A(M_{1d}) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}_A(M_{d1}) & \cdots & \mathrm{tr}_A(M_{dd}) \end{pmatrix},$$

provided that the square blocks M_{ij} are compatible with tr_A, that means here, have the same size as A. The reason for choosing tr_A as reduction operators is their homogeneity:

$$\mathrm{tr}_A(t\ M) = t\ \mathrm{tr}_A(M) \quad \text{for all } t \in \mathbb{R}.$$

Note that $A = I$ provides the leading example since

$$\begin{pmatrix} \mathrm{tr}_A & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathrm{tr}_A \end{pmatrix} \begin{pmatrix} M_{11} & \cdots & M_{1d} \\ \vdots & \ddots & \vdots \\ M_{d1} & \cdots & M_{dd} \end{pmatrix} = \begin{pmatrix} \mathrm{tr} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathrm{tr} \end{pmatrix} \begin{pmatrix} A & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A \end{pmatrix} \begin{pmatrix} M_{11} & \cdots & M_{1d} \\ \vdots & \ddots & \vdots \\ M_{d1} & \cdots & M_{dd} \end{pmatrix}.$$

The subsequent result gives a first insight into the role of this reduction operation and its connection to other structure tensors:

Proposition: (The standard tensor as an elementary reduction of $\overline{\mathcal{S}}_L$)
Let $U(x) \in \mathrm{Sym}_n(\mathbb{R})$ be a d-dimensional matrix-field. Then the standard tensor J_ρ is a reduced version of the full tensor $\overline{\mathcal{S}}_L$,

$$\mathrm{Tr}\ \overline{S_L}(U) = J_\rho(U) \in \mathrm{Sym}_d(\mathbb{R}).$$

Proof: First we realize that we can disregard the convolution with a Gaussian G_ρ with integration scale ρ or any other linear averaging process since the trace operation commutes with such linear mappings. Therefore, we can deal with the partial derivatives $\partial_{x_i} u_{i,j}$ of the matrix components directly.

Fix $p, q \in \{1, \ldots, d\}$. Then the (p, q)-th component of the standard structure tensor is given by

$$J_\rho(U) = \sum_{i,j}^{d} \partial_{x_p} u_{i,j} \cdot \partial_{x_q} u_{i,j} \,.$$

However, we obtain for the (p,q)-th component of the *reduced version* of the full structure tensor $S_L(U)$

$$\begin{aligned} \mathrm{tr}(\overline{\partial}_{x_p} U \overline{\partial}_{x_q} U) = \sum_{i=1}^{d} \left(\overline{\partial}_{x_p} U \overline{\partial}_{x_q} U\right)_{i,i} &= \sum_{i=1}^{d}\sum_{j=1}^{d} (\partial_{x_p} u_{i,j} \cdot \partial_{x_q} u_{j,i}) \\ &= \sum_{i=1}^{d}\sum_{j=1}^{d} (\partial_{x_p} u_{i,j} \cdot \partial_{x_q} u_{i,j}), \end{aligned}$$

where the symmetry of the matrix U accounts for the last equality. This proves the assertion.

The reduction operation is accompanied by an extension operation defined via the *Kronecker product*:

Definition: (Extension via Kronecker product)
The I_n-extension operation is the mapping from $\mathrm{Sym}_d(\mathbb{R})$ to $\mathrm{Sym}_{nd}(\mathbb{R})$ given by the Kronecker product $\otimes$:

$$\begin{pmatrix} v_{11} & \cdots & v_{1d} \\ \vdots & \ddots & \vdots \\ v_{d1} & \cdot & v_{dd} \end{pmatrix} \longmapsto \begin{pmatrix} v_{11} & \cdots & v_{1d} \\ \vdots & \ddots & \vdots \\ v_{d1} & \cdots & v_{dd} \end{pmatrix} \otimes \begin{pmatrix} I_n & \cdots & I_n \\ \vdots & \ddots & \vdots \\ I_n & \cdots & I_n \end{pmatrix} := \begin{pmatrix} v_{11}I_n & \cdots & v_{1d}I_n \\ \vdots & \ddots & \vdots \\ v_{d1}I_n & \cdots & v_{dd}I_n \end{pmatrix} .$$

If the $d \times d$-matrix $(v_{ij})_{ij}$ is Kronecker-multiplied with

$$\begin{pmatrix} C & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C \end{pmatrix} \begin{pmatrix} I_n & \cdots & I_n \\ \vdots & \ddots & \vdots \\ I_n & \cdots & I_n \end{pmatrix} = \begin{pmatrix} C & \cdots & C \\ \vdots & \ddots & \vdots \\ C & \cdots & C \end{pmatrix}$$

we speak of a C-extension.

4.2 A Novel Diffusion Tensor $\overline{D}$ for Matrix Fields

It is now possible to give an analog $\overline{D}$ to the diffusion tensor D in the framework of matrix fields. We proceed in four steps:

1. The matrix field $\mathbb{R}^d \ni x \mapsto U(x)$ provides us with a module field of generalized gradients $\overline{\nabla} U(x)$ from which we construct the generalized structure tensor $\overline{\mathcal{S}}_L\, U(x)$ possibly with a certain integration scale ρ. This step corresponds exactly to the scalar case.
2. We infer d-dimensional directional information by reducing $\overline{\mathcal{S}}_L\, U(x)$ with tr_A with the help of the block operator matrix given in (4) leading to a symmetric $d \times d$-matrix S, for example $S = J_\rho$ if $A = I_n$,

$$S := \begin{pmatrix} \mathrm{tr}_A & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathrm{tr}_A \end{pmatrix} \overline{\mathcal{S}}_L \, U(x) \, .$$

3. The symmetric $d \times d$-matrix S is spectrally decomposed, and the tensor transfer map H is applied to S yielding the diffusion tensor D,

$$D := H(S) \, .$$

4. Finally, we enlarge the $d \times d$-matrix D to a $nd \times nd$-matrix $\overline{D}$ by the extension operation:

$$\overline{D} = D \otimes \left(\begin{pmatrix} C & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & C \end{pmatrix} \begin{pmatrix} I_n & \cdots & I_n \\ \vdots & \ddots & \vdots \\ I_n & \cdots & I_n \end{pmatrix} \right) .$$

This last step gives another possibility to steer the filter process by the choice of the matrix C. However, this is the subject of current research. For this work, we restricted ourselves to $C = I_n$.

5 Coherence-Enhancing Diffusion Filtering for Matrix Fields

Now we have gathered the necessary ingredients to formulate the matrix-valued equivalent to the scalar coherence enhancing diffusion as expressed in equation (1):

$$\begin{aligned} \overline{\partial}_t U - \sum_{i=1}^{d} \overline{\partial}_{x_i} \left(\overline{D} \bullet \overline{\nabla} U \right) = 0 \qquad & \text{in } \Omega \times I, \\ \overline{\partial}_n U = 0 \qquad & \text{in } \partial\Omega \times I, \\ U(x,0) = F(x) \qquad & \text{in } \Omega \, . \end{aligned} \tag{5}$$

Note that the Jordan-multiplication in $\overline{D} \bullet \overline{\nabla} U$ is understood in the block-wise sense of partitioned matrices. Moreover, we translated the divergence differential operator $\mathrm{div}\ u = \sum_{i=1}^{d} \partial_{x_i} u_i$ acting on a vector-valued function $u = (u_1, \ldots, u_d)$ into its matrix-valued counterpart acting on a module field $W \in \mathrm{Sym}_n(\mathbb{R})^d$ by

$$\overline{\mathrm{div}}\ W = \sum_{i=1}^{d} \overline{\partial}_{x_i} W_i \, .$$

The numerical algorithm is inspired by the explicit scheme in [24]. We used a matrix-valued version employing the calculus framework for matrix fields as presented before. With a typical time step size of $\tau = \frac{1}{10}$, we observed numerical stability throughout our experiments.

6 Experiments

We use two data sets in our numerical experiments: The artificial matrix fields of 3×3-symmetric matrices exhibit various coherent structures ranging from simple line-like to curved features, see Figs. 1–4. We will use these data to demonstrate the gap-closing, denoising, and enhancing properties of our technique. The other matrix fields stems from a 2D slice extracted from a 3D DT-MRI data set of size of a $128 \times 128 \times 30$ of a human head, see Fig. 5.

The data are represented as ellipsoids via the level sets of the quadratic form $\{x^\top A^{-2} x = const. : x \in \mathbb{R}^3\}$ associated with a matrix $A \in \mathrm{Sym}^+(3)$. By

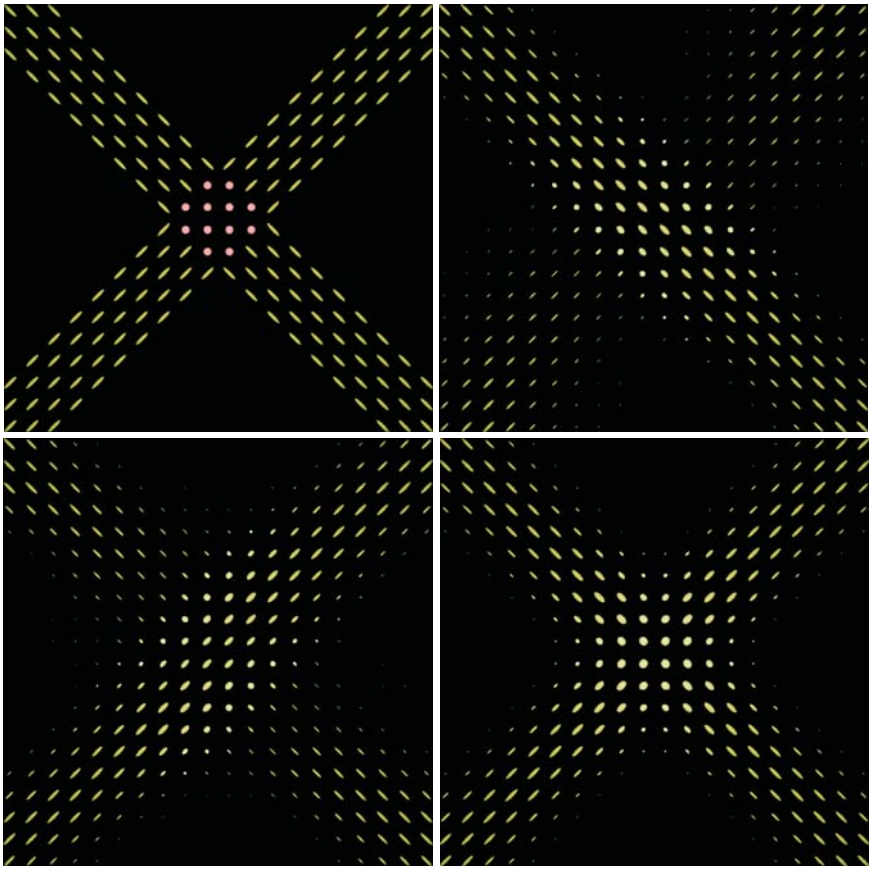


Fig. 1. (a) Top left: Artificial data set of ellipsoids indicating a crossing. **(b) Top right:** Effect of coherence-enhancing filtering if preference is given to the direction $(1, -1)$ **(c) Bottom left:** Effect of coherence-enhancing filtering if preference is given to the direction $(1, 1)$. **(d) Bottom right:** Effect of coherence-enhancing filtering if no directional priority is established.

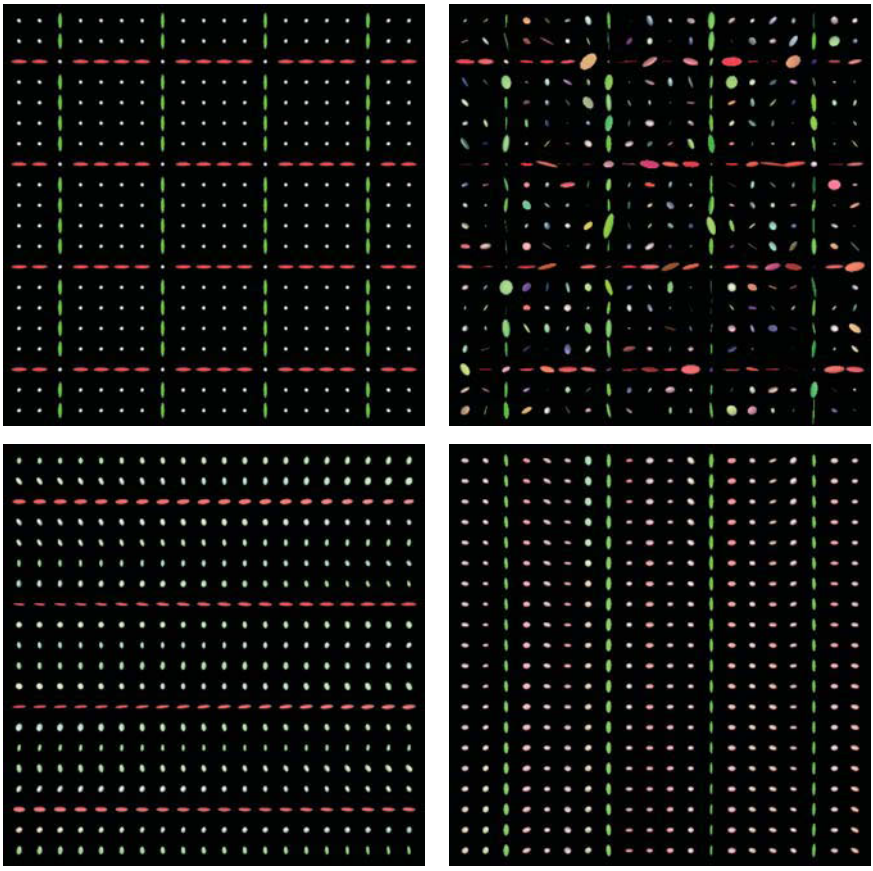


Fig. 2. (a) Top left: Original matrix field with grid-like structure. **(b) Top right:** (a) polluted with truncated Gaussian noise in the eigenvalues while the orthogonal matrices result from three Euler matrices with uniformly distributed angles. **(c) Bottom left:** Result of CED-filtering with preference on the horizontal x-direction. **(d) Bottom right:** The same but with preference on the vertical y-direction.

using A^{-2}, the length of the semi-axes of the ellipsoid correspond directly with the three eigenvalues of the matrix. We have added random positive definite matrices to the data to demonstrate the denoising capabilities of our CED-filtering concept. The eigenvectors of this noise were obtained by choosing Gaussian-distributed numbers with standard deviation $\sigma = 1,000$ and taking the absolute value for positive semidefiniteness. The high standard deviation can be explained by the fact that in real-world data the typical eigenvalues are in the order of magnitude of $1,000$. The eigenvectors of the artificial noise result in choosing three uniformly distributed angles and rotating the matrix by these angles around the coordinate axes. The resulting data is shown in Fig. 2.

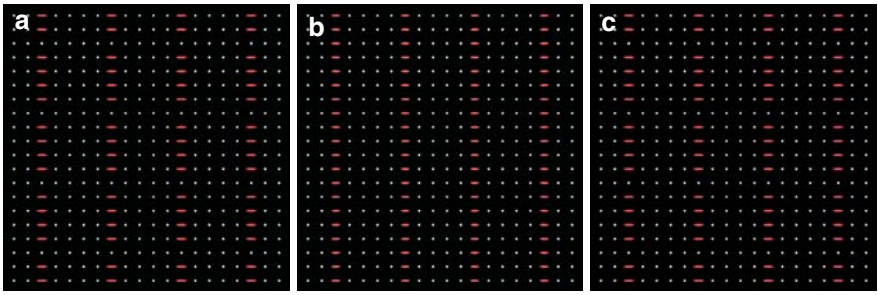


Fig. 3. (a) Left: Original image with coherent structure in y-direction and ellipsoids pointing in x-direction. **(b) Middle:** Result of CED-filtering with preference on the horizontal x-direction. **(c) Right:** Result of CED-filtering with preference on the horizontal y-direction.

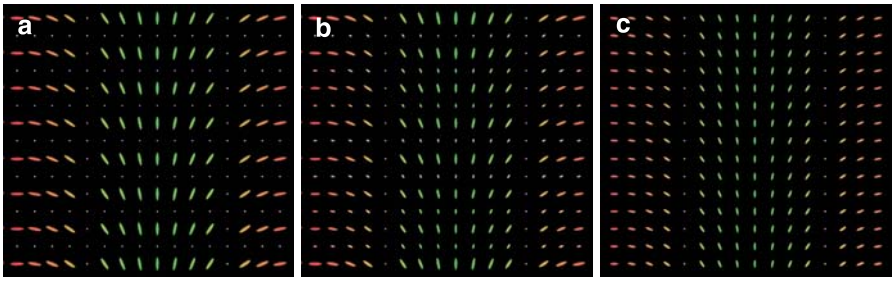


Fig. 4. (a) Left: Artificial incomplete coherent structure. **(b) Middle:** After CED-filtering with stopping time $t = 0.3$. **(c) Right:** After CED-filtering with stopping time $t = 3$

The artificial data set displayed in Fig. 1 imitates a crossing of nerve fibers. Depending on the choice of the reduction matrix A in tr_A either the diagonal directed downward, $A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$, or the one directed upward, $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, is given preference in the CED-filtering results. If no priority is set, $A = I_2$, a homogeneous structure is developing in the center, as it is expected due to the high symmetry of the image, see Fig. 1d.

An impression of the denoising and gap-closing capabilities can be obtained from the results in Fig. 2. The noisy version of an artificial data set with a grid-like structure is CED-filtered. As before, the directional preferences are conveyed by the reduction matrices $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ for the x-direction and $A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ for the y-direction. The noise is removed and the lines in the selected direction are getting completed.

Fig. 3 shows the results of directionally selective CED-filtering if the direction of the coherent structure (here the y-direction) does *not* coincide with the direction of the ellipsoids (here the x-direction). By selecting the x-direction

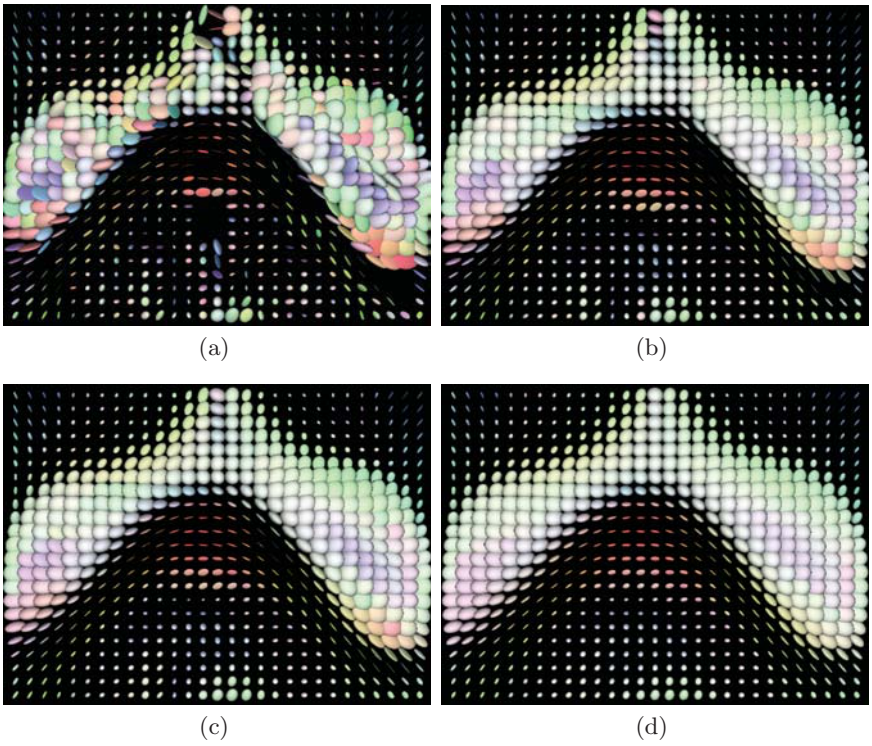


Fig. 5. (a) Top left: Original DT-MRI data: lower part of the Corpus Callosum fiber area. **(b) Top right:** After CED-filtering with stopping time $t = 1$. **(c) Bottom left:** CED-filtering with $t = 3$. **(d) Bottom right:** CED-filtering with $t = 5$.

with a proper choice of A, we allow for the enhancement of coherent structures, that is, the closing of lines if there is a change in this x-direction. In Fig. 3b we have such a change, hence the present lines in y-direction are completed. We do not have a change in y-direction (the balls and ellipsoids have the same y-extension), hence the selection of the y-direction via $A = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ triggers no enhancement at all. Therefore this directional CED-filtering has no effect, see Fig. 3c.

The experiment depicted in Fig. 4 demonstrates that even areas with no information may constitute a coherent structure that is preserved by CED-filtering without directional preferences. The (almost) empty lines in x-direction are getting filled, while the two lines in y-direction remain untouched by the filtering. The explanation is that in x-direction we have changes in the shape and orientation of the ellipsoids in the vicinity of the empty lines, hence, the gap-closing quality of CED-filtering comes into effect. However, proceeding

in y direction no changes in the surrounding ellipsoids is discernible, rendering the CED-filtering idle in this direction.

We applied CED-filtering without directional preference to real DT-MRI data, to investigate its usefulness as a pre-processing step, e.g. for fiber tracking. The results are displayed in Fig. 5 confirming the regularizing effect and the enhancement of coherent structures, such as the fiber bundles below the Corpus Callosum.

7 Summary and Future Work

In this chapter, we have developed a novel structure tensor concept for matrix fields. This approach is based on an operator-algebraic view on matrices and their rich algebraic properties. We have shown how to infer directional information from this high dimensional data by specifying directional preferences, and clarified the standard structure tensor for matrix fields as a special case of our extended concept. An application of the new tensor concept enabled us to develop a directionally selective coherence-enhancing diffusion filtering of matrix fields by employing a generic differential calculus framework for matrices. The matrix-valued CED-filtering exhibits similar behaviour as its scalar counterpart.

Current work encompasses the investigation of further opportunities to steer the filtering process, e.g. in the extension step, and its relation to other customizable tensor concepts for matrix fields. Future research will focus on further applications of the extended structure tensor concepts in image processing for matrix fields.

Acknowledgments

We are grateful to Anna Vilanova i Bartrolí (Eindhoven Institute of Technology) and Carola van Pul (Maxima Medical Center, Eindhoven) for providing us with the DT-MRI data set.

References

[1] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Fast and simple calculus on tensors in the log-Euclidean framework. In J. Duncan and G. Gerig, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2005, Part I*, volume 3749 of *LNCS*, Springer, Berlin, pp. 115–122, 2005.

[2] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, 2006.

[3] J. Bigun. *Vision with Direction*. Springer, Berlin, 2006.
[4] T. Brox and J. Weickert. A TV flow based local scale estimate and its application to texture discrimination. *Journal of Visual Communication and Image Representation*, 17(5):1053–1073, 2006.
[5] T. Brox, J. Weickert, B. Burgeth, and P. Mrázek. Nonlinear structure tensors. *Image and Vision Computing*, 24(1):41–55, 2006.
[6] B. Burgeth, A. Bruhn, S. Didas, J. Weickert, and M. Welk. Morphology for matrix-data: Ordering versus PDE-based approach. *Image and Vision Computing*, 25(4):496–511, 2007.
[7] B. Burgeth, S. Didas, L. Florack, and J. Weickert. A generic approach for singular PDEs for the processing of matrix fields. In F. Sgallari, F. Murli, and N. Paragios, editors, *Scale Space and Variational Methods in Computer Vision*, volume 4485 of *Lecture Notes in Computer Science*, Springer, Berlin, pp. 556–567, 2007.
[8] B. Burgeth, S. Didas, L. Florack, and J. Weickert. A generic approach to diffusion filtering of matrix-fields. *Computing*, 81:179–197, 2007.
[9] B. Burgeth, N. Papenberg, A. Bruhn, M. Welk, C. Feddern, and J. Weickert. Morphology for higher-dimensional tensor data via Loewner ordering. In C. Ronse, L. Najman, and E. Decencière, editors, *Mathematical Morphology: 40 Years On*, volume 30 of *Computational Imaging and Vision*, Springer, Dordrecht, pp. 407–418, 2005.
[10] C. Chefd'Hotel, D. Tschumperlé, R. Deriche, and O. Faugeras. Constrained flows of matrix-valued functions: Application to diffusion tensor regularization. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision – ECCV 2002*, volume 2350 of *Lecture Notes in Computer Science*, Springer, Berlin, pp. 251–265, 2002.
[11] O. Coulon, D. C. Alexander, and S. A. Arridge. A regularization scheme for diffusion tensor magnetic resonance images. In M. F. Insana and R. M. Leahy, editors, *Information Processing in Medical Imaging – IPMI 2001*, volume 2082 of *Lecture Notes in Computer Science*, Springer, Berlin, pp. 92–105, 2001.
[12] S. Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics and Image Processing*, 33:116–125, 1986.
[13] C. Feddern, J. Weickert, B. Burgeth, and M. Welk. Curvature-driven PDE methods for matrix-valued images. *International Journal of Computer Vision*, 69(1):91–103, 2006.
[14] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, Interlaken, Switzerland, pp. 281–305, 1987.
[15] Y. Gur and N. Sochen. Diffusion over tensor fields via lie group pde flows. In M. Farber, R. Ghrist, M. Burger, and D. Koditschek, editors, *Topology and Robotics*, Contemporary Mathematics, Springer, Berlin, pp. 59–74, 2007.

[16] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1990.

[17] H. Knutsson. Representing local structure using tensors. In *The 6th Scandinavian Conference on Image Analysis*, pp. 244–251, Oulu, Finland, June 1989. Report LiTH–ISY–I–1019, Computer Vision Laboratory, Linköping University, Sweden, 1989.

[18] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor mri processing. *Journal of Mathematical Imaging and Vision*, 25(3):423–444, 2006.

[19] T. McGraw, B. C. Vemuri, Y. Chen, M. Rao, and T. H. Mareci. DT-MRI denoising and neuronal fiber tracking. *Medical Image Analysis*, 8:95–111, 2004.

[20] X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

[21] T. Schultz, B. Burgeth, and J. Weickert. Flexible segmentation and smoothing of DT-MRI fields through a customizable structure tensor. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, and T. Malzbender, editors, *Advances in Visual Computing*, volume 4291 of *Lecture Notes in Computer Science*, Springer, Berlin, pp. 454–464, 2006.

[22] D. Tschumperlé and R. Deriche. Diffusion tensor regularization with constraints preservation. In *Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, Kauai, HI, pp. 948–953, December 2001. IEEE Computer Society Press.

[23] D. Tschumperlé and R. Deriche. Orthonormal vector sets regularization with PDE's and applications. *International Journal of Computer Vision*, 50(3):237–252, 2002.

[24] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998.

[25] J. Weickert. Coherence-enhancing diffusion filtering. *International Journal of Computer Vision*, 31(2/3):111–127, 1999.

[26] J. Weickert. Coherence-enhancing diffusion of colour images. *Image and Vision Computing*, 17(3/4):199–210, 1999.

[27] J. Weickert and T. Brox. Diffusion and regularization of vector- and matrix-valued images. In M. Z. Nashed and O. Scherzer, editors, *Inverse Problems, Image Analysis, and Medical Imaging*, volume 313 of *Contemporary Mathematics*, AMS, Providence, pp. 251–268, 2002.

[28] J. Weickert and H. Hagen, editors. *Visualization and Processing of Tensor Fields*. Springer, Berlin, 2006.

Coordinates-Based Diffusion Over the Space of Symmetric Positive-Definite Matrices

Yaniv Gur and Nir Sochen

Department of Applied Mathematics, Tel-Aviv university, Ramat-Aviv,
Tel-Aviv 69978, Israel,
{yanivg,sochen}@post.tau.ac.il

Summary. In some recent applications, the objects of interest are matrices. Diffusion Tensor MRI or in short DTI is an example of such an application. In this case, the image is divided into voxels where each voxel is described by a 3×3 symmetric positive-definite (SPD) matrix. In this chapter, we present an intrinsic approach for diffusion over the space of $n \times n$ symmetric positive-definite matrices, denoted by P_n. The basis of this framework is the description of P_n as a Riemannian manifold by means of the *local* coordinates and a natural Riemannian metric. One may choose various coordinate systems to cover the P_n manifold. We show that the analytical calculations, as well as the numerical implementation, become simple by choosing *Iwasawa* coordinates. Then, we define a $GL(n)$-invariant metric over P_n with respect to these coordinates. The metric is defined by means of the scalar product on Sym_n, the space of $n \times n$ symmetric matrices. The "image" is described here as a section of a fiber bundle. Then, the metric over P_n is combined with the metric over the two- or three-dimensional image domain in order to form the metric over the section. By means of the Beltrami framework, we define a functional over the space of sections. Variation of this functional leads to a set of $\frac{1}{2}n(n+1)$ coupled equations of motion with respect to the local coordinates on P_n. The solution of these equations defines a structure-preserving flow on this manifold. Finally, we demonstrate this framework on the case of P_3 by smoothing of real DTI datasets.

1 Introduction

Regularization of Diffusion Tensor Magnetic Resonance Imaging, DT-MRI (or in short DTI) data has been attracted much attention over the past few years (e.g., [1–3, 15, 25–27]). The need for regularization of DTI data motivated the development of new techniques to deal with tensors. However, complications arise when one wishes to regularize a tensor field. Tensors have properties that we would like to preserve during the regularization process. In DTI, for example, the tensors are symmetric and positive-definite. These properties have to be preserved to maintain the physical information encoded in these tensors. As a result, special analytical, as well as numerical, structure preserving schemes had to be developed.

In 2002 [24], it was proposed to regularize DTI datasets via the polar decomposition. In this method, the eigenvalues (diffusivities) and the eigenvectors (orientations) are regularized independently. However, the polar decomposition is not unique and therefore a realignment step for the eigenvectors in every iteration is needed. Also, treating the eigenvalues and the eigenvectors separately may result in eigenvector permutations that will lead to wrong results.

Some of the structure preserving flows are based on the coordinate-free methods for ODEs evolving on Lie-group manifolds [7, 8]. These methods rely strongly on the interplay between the Lie-group and the Lie-algebra via smooth mappings. In [2, 5], the exponential mapping was used to construct structure preserving flows for DTI orientations. Similarly, the exponential mapping may be used to map the linear space of $n \times n$ symmetric matrices of order n, Sym_n onto P_n, the space of SPD matrices. This fact, combined with the definition of a *global* Riemannian distance between two tensors, is the basis for various frameworks for DTI regularization [4, 13, 16]. However, the implementation of the global distance between two tensors involves operations on tensors and therefore it is time consuming. The Log-Euclidean distance, proposed in [4] is cost-effective but is not $GL(n)$-invariant.

In this chapter, we propose a novel framework for smoothing of DTI datasets. The basis of this framework is the description of P_n as a Riemannian manifold. The manifold is parameterized via the *local* coordinates, and the infinitesimal distance between two tensors is measured in terms of a natural $GL(n)$-invariant Riemannian metric.

This framework arises from a pure geometric point of view. By using the mathematical notion of fiber-bundles, we define a product space of two spaces: the image domain (the base manifold) and the feature space of $n \times n$ SPD matrices, P_n (the fiber). In this mathematical framework, a specific DT image is described as a section in the fiber bundle where the feature space is P_3. These issues are discussed in Sect. 2.

In Sect. 2.1, we discuss the basics of the Riemannian geometry of P_n. The space P_n is identified with the coset space $GL(n)/O(n)$ and the underlying algebraic structure is employed to define a $GL(n)$-invariant Riemannian metric. Then, the metric over the sections in the fiber bundle is defined in terms of this metric.

In Sect. 3, we introduce a functional over sections by means of the Beltrami framework. Using the calculus of variations, we derive the equations of motion with respect to the coordinates on the section. The coordinates evolve with respect to the geometry of the section via the induced metric.

The proper choice of coordinate system to parameterize P_n is important both analytically and numerically. In Sect. 4, we discuss the partial and the full *Iwasawa* coordinates. We show that the *Iwasawa* decomposition that naturally respects the coset algebraic structure, yields a natural coordinate system. This choice of coordinates to parameterize P_n simplifies the analytical calculations as well as the numerical implementation. Then, the flow on P_3 is

defined by a set of six coupled Beltrami equations with respect to the Iwasawa coordinates. The numerical implementation of these equations is almost straightforward and no time consuming operations are needed. Therefore, compared to previous works (e.g., [16]), running times are improved, while the $GL(n)$-invariance property of the metric is maintained. We demonstrate this framework on a real DTI dataset at the end of the section.

2 Images and Fiber Bundles

Generally, an image may be described locally as a product of two spaces: the image domain (two- and three-dimensional usually) and the feature space, which is composed of objects such as intensity, RGB values, orientation fields, etc. A product of two-spaces may be described mathematically via the concept of a fiber bundle. A fiber bundle is characterized by the total space E, a structural group G (a group of homeomorphism of the fiber F onto itself) together with a continuous surjective map $\pi : E \to B$. Locally, the total space E is described as product of the base manifold and the fiber $B \times F$. Globally, this may not be the case, since global features of the total space are not observed locally (e.g., the Mobius band where the twist cannot be observed locally). When the identification of the total space with the direct product of the base space and the fiber is possible, this is known as a trivial bundle. In image processing language, the total space is the spatial-feature manifold. For example, using this concept, a gray-value image is described locally by the product $\mathbb{R}^2 \times \mathbb{R}^+$.

A section of the bundle is a mapping $f : B \to E$ such that $\pi \circ f = I$. In our language, a given image is a section in a fiber bundle. For each pixel in the image domain, we have a unique assignment of a feature space object (see Fig. 1).

In DTI we have a tensor-valued image, where for each pixel there is unique assignment of a three-dimensional SPD matrix. Thus, a DT image is a section in the fiber bundle with the map $f : \mathbb{R}^m \to \mathbb{R}^m \times P_3$ where $\mathbb{R}^m$ is in practice two- or three-dimensional Euclidean image domain. Thus, the spatial-feature manifold is described by the product $\mathbb{R}^m \times P_3$. The structural group in this case is simply $GL(3)$, which is the generalized linear group of 3×3 real and nonsingular matrices. The action of $GL(3)$ on P_3 is given by $Y[g] = g^T Y g$ for $Y \in P_3$ and $g \in GL(3)$. This description is extendable to any n with the fiber bundle $\mathbb{R}^m \times P_n$. In this case, the structural group is $GL(n)$.

2.1 The Riemannian Geometry of P_n

One of the key points in the proposed framework is the definition of a meaningful metric and a coordinate system on the spatial-feature manifold. In this section, we will discuss the metric issue while the coordinate choice is discussed in Sect. 4.

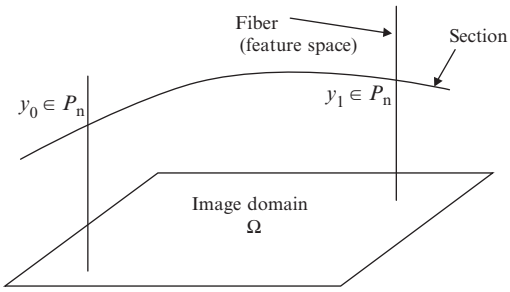


Fig. 1. Description of an image as a section of a fiber bundle. The base manifold B is the image domain Ω, the fiber is the feature space F (e.g., intensity values, color DT, etc.). Here the image domain is Euclidean $\Omega \subset \mathbb{R}^n$. Locally the bundle is described as a direct product $E = B \times F$ where E is the total space. In DTI images the fiber is taken to be the space of SPD matrices P_n. A DT image is composed of a particular selection of the SPD matrix to each point of the base manifold. Therefore, a DT image is a section in the fiber bundle. This section is referred to here as the image manifold.

The metric over the Euclidean image domain with Cartesian coordinate system is simply $ds^2_{\mathbb{R}^2} = dx^2 + dy^2$. In order to write the metric over the feature manifold, we have to discuss the Riemannian structure of P_n, thus, to equip P_n with a Riemannian metric. Let us briefly review the important facts about P_n. Detailed discussions on the Riemannian geometry of P_n can be found in [9, 12].

The space of SPD matrices may be identified with an open cone in R^m where $m = n(n+1)/2$, i.e., for any $V, W \in P_n$ and for any positive scalar $c > 0$, we have $V + W \in P_n$ and $cV, cW \in P_n$.

Without going into detail, the tangent space at every point $Y \in P_n$ may be identified with the vector space of $n \times n$ symmetric matrices, Sym_n. The inner product at any point $Y \in P_n$ is defined by

$$\langle v, w \rangle_Y = tr(Y^{-1} v Y^{-1} w), \quad \forall\, v, w \in T_Y P_n, \tag{1}$$

thus, $\|v\|^2_Y = tr((Y^{-1}v)^2)$.

Let $Y(t)$ be a curve in P_n. Since $Y'(t) \in T_{Y(t)} P_n$, the length of the curve is given by

$$s = L(t) = \int_a^t \left\{ tr\left[(Y(\tau)^{-1} Y'(\tau))^2 \right] \right\}^{1/2} d\tau. \tag{2}$$

Then, the square of the differential of s is given by

$$\left(\frac{ds}{dt} \right)^2 = tr(Y(t)^{-1} Y'(t))^2. \tag{3}$$

Finally, the Riemannian metric over P_n is given by the abbreviation

$$ds^2_{P_n} = tr((Y^{-1}dY)^2), \tag{4}$$

where $Y^{-1} = (y_{ij})^{-1}_{1\leq i,j\leq n}$ and $dY = (dy_{ij})_{1\leq i,j\leq n}$.

The Riemannian metric is, by definition, positive-definite [12, 23]. Also, it can be easily shown that this metric is invariant under the action of $GL(n)$. Let $W = Y[g]$, where the differential is given by $dW = dY[g]$. Then, upon plugging everything in $ds^2_{P_n}$, it follows that

$$\begin{aligned} ds^2_{P_n} &= tr((Y^{-1}dY)^2) \\ &= tr((gW^{-1}g^Tg^{-T}dWg^{-1})^2) = tr((W^{-1}dW)^2). \end{aligned} \tag{5}$$

With respect to the Riemannian metric, the notion of geodesics is defined. Since P_n is a geodesically complete Riemannian space, according to the Hopf–Rinow theorem, any two points in this space may be joined by a minimizing geodesic [12]. Its length is the geodesic distance between the two points. Moreover, it has been shown in [10] that there is precisely one geodesic connecting any two points on P_n. The geodesic distance between any two points $A, B \in P_n$ is given by [12]

$$d(A,B) = \left(\sum_{i=1}^{n} log^2\lambda_i\right)^{1/2}, \tag{6}$$

where λ_i denotes the i'th eigenvalue of the matrix $A^{-1}B$. This geodesic distance is the so-called Riemannian distance, which was used in [16] as a distance measure between SPD matrices. It can be easily shown that the geodesic distance is invariant under the inversion $A \to A^{-1}$. This inversion is an involutive isometry on P_n. Hence, P_n is a Riemannian symmetric space.

After the metric over P_n was introduced, the metric over the spatial-feature manifold $M = \mathbb{R}^2 \times P_n$, for example, is simply given by $ds^2_M = dx^2 + dy^2 + tr((Y^{-1}dY)^2)$. Finally, we would like to mention that in [6] we used a similar geometric construction in order to define flows on compact Lie-groups, which are another example of Riemannian symmetric spaces.

3 Calculus of Variations in Fiber Bundles

Using functional analysis on sections in the fiber bundle, we are able to derive a flow on the section. This will be done via the Beltrami framework [20], where the functional is the Polyakov action [18]. This approach is intrinsic, where the explicit parameterization on the section is used. The Beltrami framework has been intensively discussed in the literature (e.g., [11, 19–21]) and therefore we only briefly review its main ideas.

3.1 The Polyakov Action

Denote by (Σ, γ) the image manifold and its metric and by (M, h) the spatial-feature manifold (the fiber bundle) and its metric, then, the section of interest is expressed locally by the map $X : \Sigma \to M$. A functional over the space of sections is given in local coordinates by the following expression:

$$S = \int_{\Sigma} d^n x \sqrt{|\gamma|} \gamma^{\mu\nu} \partial_\mu X^i \partial_\nu X^j h_{ij}(X), \tag{7}$$

which is known as the Polyakov action [18]. The integration is taken over two- or three-dimensional image manifold ($n = 2$ or $n = 3$), $|\gamma|$ is the determinant of the image metric, $\gamma^{\mu\nu}$ denotes the components of the inverse of the metric tensor, and h_{ij} are the components of the embedding space metric (the spatial-feature fiber bundle manifold). The coordinates in the spatial-feature space are denoted by X^i. The values of μ and ν range from 1 to the dimension of Σ (the dimension of the image manifold) and the values of the i and j indices range from 1 to the dimension of the embedding space.

Using calculus of variations with respect to the embedding coordinates X^i, we obtain the Euler–Lagrange equations for this action:

$$-\frac{1}{2\sqrt{|\gamma|}} h^{il} \frac{\delta S}{\delta X^l} = \frac{1}{\sqrt{|\gamma|}} \partial_\mu \left(\sqrt{|\gamma|} \gamma^{\mu\nu} \partial_\nu X^i \right) + \Gamma^i_{jk} \gamma^{\mu\nu} \partial_\mu X^j \partial_\nu X^k. \tag{8}$$

By the gradient descent method, we obtain a set of PDEs (i.e., Beltrami equations) with respect to the embedding coordinates. However, since the coordinates of the image domain are fixed, the interesting equations are for the coordinates of the fiber (P_n in our case). Hence,

$$X^i_t = \frac{1}{\sqrt{|\gamma|}} \partial_\mu \left(\sqrt{|\gamma|} \gamma^{\mu\nu} \partial_\nu X^i \right) + \Gamma^i_{jk} \gamma^{\mu\nu} \partial_\mu X^j \partial_\nu X^k, \tag{9}$$

where $i = 1, ...dim(M) - dim(\Sigma)$ and where Γ^i_{jk} are the Christoffel symbols. When the embedding space is Euclidean with Cartesian coordinates, all the Christoffel symbols vanish. The Christoffel symbols are calculated using the metric of the embedding space

$$\Gamma^i_{jk} = \frac{1}{2} h^{il} \left(\partial_j h_{lk} + \partial_k h_{jl} - \partial_l h_{jk} \right). \tag{10}$$

3.1.1 The Two-Dimensional Induced Metric

The induced metric minimizes the Polyakov functional. Therefore, the image manifold (or section) metric is induced from the fiber bundle's metric. In this way, the flow depends not only on the geometry of the image domain but also on the geometry of the data. Moreover, the induced metric is a dynamical variable, which changes along the flow.

Although the object of interest is the spatial-feature manifold, distances are measured on the image manifold. We assume an isometric embedding, i.e., infinitesimal distances on the spatial-feature manifold are equal to infinitesimal distances on the image manifold. As a consequence, we may use the chain rule such that $dY = Y_x dx + Y_y dy$. Then, we plug this expression into the metric over the section to get the components of the induced metric

$$\gamma_{\mu\nu} = \delta_{\mu\nu} + tr(Y^{-1}\partial_\mu Y Y^{-1}\partial_\nu Y), \tag{11}$$

where $\mu, \nu = 1, 2$ are the indices of the local coordinates on the image manifold. In matrix form and in terms of the spatial coordinates, it may be written as

$$(\gamma_{\mu\nu}) = \begin{pmatrix} 1 + tr((Y^{-1}Y_x)^2) & tr((Y^{-1}Y_x)(Y^{-1}Y_y)) \\ tr((Y^{-1}Y_y)(Y^{-1}Y_x)) & 1 + tr((Y^{-1}Y_y)^2) \end{pmatrix}. \tag{12}$$

4 Diffusion on the Space of SPD Matrices

In this section, we obtain the main result of this chapter. The framework and the mathematical tools that we have introduced in the previous sections are used here to define flows on P_n. The results are general, however, we put an emphasis on P_3 for the DTI application. At the end of this section, we present a full numerical example of DTI smoothing using this framework.

In Sect. 3.1.1, we defined the induced metric from the spatial-feature manifold $\mathbb{R}^2 \times P_n$ to the image manifold. For an Euclidean image domain, this metric has the form $ds^2 = \left(\delta_{\mu\nu} + tr(Y^{-1}\partial_\mu Y Y^{-1}\partial_\nu Y)\right) dx^\mu dx^\nu$. We would like to write the metric in terms of the coordinates on the P_n manifold explicitly.

The P_n space is identified with an open cone in $\mathbb{R}^m$ with $m = \frac{1}{2}n(n+1)$. The m different entries of the symmetric matrix may be identified with the Cartesian coordinates of $\mathbb{R}^m$. Therefore, it is straightforward to parameterize P_n using these coordinates. Regularization of DTI datasets using this coordinate system has been proposed in [14]. Let

$$Y = \begin{pmatrix} y_1 & y_2 & y_3 \\ y_2 & y_4 & y_5 \\ y_3 & y_5 & y_6 \end{pmatrix}, \tag{13}$$

where the differential matrix is given by

$$dY = \begin{pmatrix} dy_1 & dy_2 & dy_3 \\ dy_2 & dy_4 & dy_5 \\ dy_3 & dy_5 & dy_6 \end{pmatrix}. \tag{14}$$

Taking the trace of $(Y^{-1}dY)^2$ one gets a long and cumbersome expression with complicated expressions involving the y_i's in the denominator. The metric tensor is even more cumbersome, and there are 78 Christoffel symbols

associated with it. Moreover, the induced metric that has to be calculated in every iteration is also complicated. As a result, the numerical implementation is problematic.

An important issue in analysis on manifolds is the right choice of coordinate system for the problem. Therefore, we may try to choose a different coordinate system to simplify the analytical calculations as well as the numerics.

The polar coordinates that correspond to the polar decomposition of SPD matrices are often used to parameterize P_n (e.g., pages 23–24 in [23]). Although there is a nice and elegant metric associated with this coordinate system, it suffers from two main problems:

1. The polar decomposition is not unique. There are two possible different decompositions for P_n. The eigenvectors of $Y \in P_n$ in these two decompositions are in opposite directions. Therefore, artificial discontinuities may be created. As a result, one has to apply a heuristic realignment step in every iteration of the numerical solution (e.g., [24]).
2. Using the polar decomposition, one has to deal with diffusion on orientations field (the directions of the eigenvectors). The manifold of the orientation space is compact, i.e., the spheres S^n. Generally, diffusion on orientation fields leads to numerical problems (e.g., [11, 17, 22]) due to the impossibility of covering a compact manifold with only one coordinate chart.

Luckily, there is a third set of coordinates called Iwasawa coordinates, which correspond to the Iwasawa decomposition of symmetric matrices. We discuss the partial and the full Iwasawa decomposition in next section. Surprisingly, they turn out to be the natural parameterization on P_n for analytical as well as for numerical considerations.

Let us first review the partial Iwasawa decomposition. Then, we will show that the full Iwasawa decomposition is obtained by repeatedly applying the partial Iwasawa decomposition to the matrices $Y \in P_n$.

4.1 Partial Iwasawa Coordinates

It has been shown (see [23] and references therein) that in the partial Iwasawa coordinates a $n \times n$ SPD matrix may be decomposed as follows:

$$Y = \begin{pmatrix} V & 0 \\ 0 & W \end{pmatrix} \begin{bmatrix} I_p & X \\ 0 & I_q \end{bmatrix}, \tag{15}$$

where $V \in P_p$, $W \in P_q$, $X \in \mathbb{R}^{q \times p}$, $n = p + q$. The notation $A[B]$ is for $B^T AB$. This decomposition is unique for a given p, q and $Y \in P_n$.

A direct calculation of the metric $ds^2 = tr((Y^{-1}dY)^2)$ shows that the metric on P_n with respect to the partial Iwasawa coordinates is given by

$$ds^2_Y = tr(V^{-1}dV)^2 + tr(W^{-1}dW)^2 + tr(2dXW^{-1}(dX)^T V), \tag{16}$$

where dX is the differential of X.

In the special case where $V \in P_{n-1}$, we have

$$Y = \begin{pmatrix} V & 0 \\ 0 & w \end{pmatrix} \begin{bmatrix} I_p & x \\ 0 & 1 \end{bmatrix}, \tag{17}$$

where $w > 0$ and $x \in \mathbb{R}^{n-1}$. In this case, the metric takes the form

$$ds_Y^2 = tr(V^{-1}dV)^2 + (w^{-1}dw)^2 + 2w^{-1}V[dx]. \tag{18}$$

To perform smoothing of DTI datasets, we would like to calculate the metric tensor and the induced metric on P_3.

According to (17), the partial Iwasawa decomposition for an arbitrary SPD matrix is given by

$$Y = \begin{pmatrix} c & d & 0 \\ d & e & 0 \\ 0 & 0 & f \end{pmatrix} \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix}. \tag{19}$$

This decomposition is unique and may be calculated easily for a given matrix.

From (18), we get the metric tensor for this coordinate system:

$$(h_{ij}) = \begin{pmatrix} \frac{2c}{f} & \frac{2d}{f} & 0 & 0 & 0 & 0 \\ \frac{2d}{f} & \frac{2e}{f} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{e^2}{s} & \frac{-2de}{s} & \frac{d^2}{s} & 0 \\ 0 & 0 & \frac{-2de}{s} & \frac{2(d^2+ce)}{s} & \frac{-2cd}{s} & 0 \\ 0 & 0 & \frac{d^2}{s} & \frac{-2cd}{s} & \frac{c^2}{s} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{f^2} \end{pmatrix}, \tag{20}$$

where $s = (ce - d^2)^2$. Note that since we are dealing with SPD matrices, V is also symmetric and positive-definite, i.e., $ce - d^2 > 0$. Also $f > 0$ and therefore, the components of the metric tensor cannot be singular at any point.

The next task is to calculate the Christoffel symbols, which are given by the formula

$$\Gamma^i_{jk} = \frac{1}{2}h^{il}\left(\partial_j h_{lk} + \partial_k h_{jl} - \partial_l h_{jk}\right), \tag{21}$$

where h^{il} are the entries of the inverse of h. From straightforward, but cumbersome calculation, which we performed with symbolic manipulation software, one obtains 33 nonvanishing Christoffel symbols. We do not list the Christoffel components here. However, we note that they have a compact form. As for the components of the metric tensor, the positive-definiteness of Y implies that the Christoffel components cannot be singular at any point on the P_n manifold.

Finally, we would like to calculate the induced metric

$$ds^2 = \left(\delta_{\mu\nu} + \beta tr(Y^{-1}\partial_\mu Y Y^{-1}\partial_\nu Y)\right) dx^\mu dx^\nu, \tag{22}$$

where $\beta > 0$ is a scaling parameter that takes into account the differences in dimension between the spatial coordinates and the coordinates of the feature space.

After plugging everything into the Beltrami equation, we get six coupled equations for the coordinates of the partial Iwasawa decomposition:

$$X_t^i = \frac{1}{\sqrt{|\gamma|}}\partial_\mu\left(\sqrt{|\gamma|}\gamma^{\mu\nu}\partial_\nu X^i\right) + \Gamma^i_{jk}\gamma^{\mu\nu}\partial_\mu X^j\partial_\nu X^k, \tag{23}$$

where $i = 1, \ldots, 6$.

These equations are solved together in each iteration using standard finite differences schemes. The components of the induced metric and the Christoffel symbols are evaluated in each iteration. An important result of this framework is that no constraint on the positive-definiteness of the matrices is needed. Indeed the ellipticity of the Laplace–Beltrami operator that generates the flow implies the validity of the extremum principle. This, in turn, means that *the positive-definiteness of the initial condition is a necessary and sufficient condition for the flow to stay on* P_n.

4.2 Full Iwasawa Decomposition

This coordinate system is obtained by applying the partial Iwasawa decomposition to the matrix $V \in P_{n-1}$ in the previous subsection. This procedure can continue until complete diagonalization of V. In the case of interest here, $V \in P_2$. Thus, applying the partial Iwasawa decomposition to V, we have

$$V = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}. \tag{24}$$

The result of this operation for $Y \in P_3$ is

$$Y = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}\begin{bmatrix} 1 & d & e \\ 0 & 1 & f \\ 0 & 0 & 1 \end{bmatrix} = A[N], \tag{25}$$

where A is a positive diagonal matrix and N is an element of the nilpotent group. Note that the Iwasawa decomposition for Lie-groups is given by $G = KAN$, where K is a compact subgroup, A is an Abelian subgroup, and N is a nilpotent subgroup. Taking G as $GL(n)$, we choose K as $O(n)$. Since $P_n \sim GL(n)/O(n)$, it may be parameterized with AN. However, the product AN is not symmetric. Therefore the symmetric product is defined by $A[N] = N^T AN$ where, since $Y \in P_n$, the matrix A has to be positive-definite. This explains intuitively the form of the Iwasawa coordinate system for P_n.

Calculating ds^2_Y for the full Iwasawa decomposition in terms of A and N, we get

$$ds_Y^2 = tr((A^{-1}dA)^2) + 2tr(A^{-1}[N^{-T}]A[dN]). \tag{26}$$

We would expect a term of the form $tr((N^{-1}dN)^2)$ in the metric, but for a matrix N, which is an upper diagonal matrix with ones on its diagonal, this term vanishes.

Finally, the metric tensor for the full Iwasawa decomposition is given by

$$(h_{ij}) = \begin{pmatrix} 1\frac{1}{a^2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{b^2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{c^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2a(c+bf^2)}{bc} & -\frac{2af}{c} & 0 \\ 0 & 0 & 0 & -\frac{2af}{c} & \frac{2a}{c} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2b}{c} \end{pmatrix}. \tag{27}$$

Note that the Iwasawa decomposition of a SPD matrix implies that $a, b, c > 0$. Therefore, the components of the metric tensor are not singular.

There are only 26 Christoffel symbols associated with this metric tensor. Using the new metric tensor, the induced metric is calculated easily and the set of the Beltrami equations is constructed. The results in the case of the full Iwasawa coordinates are similar to those obtained with the partial Iwasawa coordinates.

In Fig. 2, we demonstrate smoothing of a real noisy DTI dataset using the full Iwasawa coordinates. In this DTI dataset, the brain ventricles are presented as well as the corpus callosum and the cingulum fibers.

Using the Iwasawa decomposition, the corresponding coordinates are extracted easily from the original data. Then, we diffused over the noisy field using the Beltrami equations with respect to the Iwasawa coordinates. The size of the DTI slice was 128 $\times$128, where running time is approximately 5 s for 50 iterations on MatLab platform (on a laptop with Pentium 1.73 GHz and 1 Gb of RAM). It can be seen that smoothing occurs in homogenous regions like the ventricles, while important features are preserved. For example, in the detail in the second row, it can be seen that the directions of the fibers in the corpus callosum are preserved. Moreover, the directions of the fibers in the regularized image are more continuous than in the noisy image.

The results are presented using three different DTI data representations. The first (Fig. 2) is a representation of a SPD matrix as an ellipsoid, where the length of its axes is determined by the diffusivities (the eigenvalues of the SPD matrix). The orientations of the ellipsoids in the 3D space are determined by the eigenvectors of the SPD matrix. Each ellipsoid is color-coded according to the components of the principal eigenvector (the eigenvector of the largest eigenvalue).

In Fig. 3 (top row), the same results are presented using the fractional anisotropy (FA) representation, a scalar image that shows the level

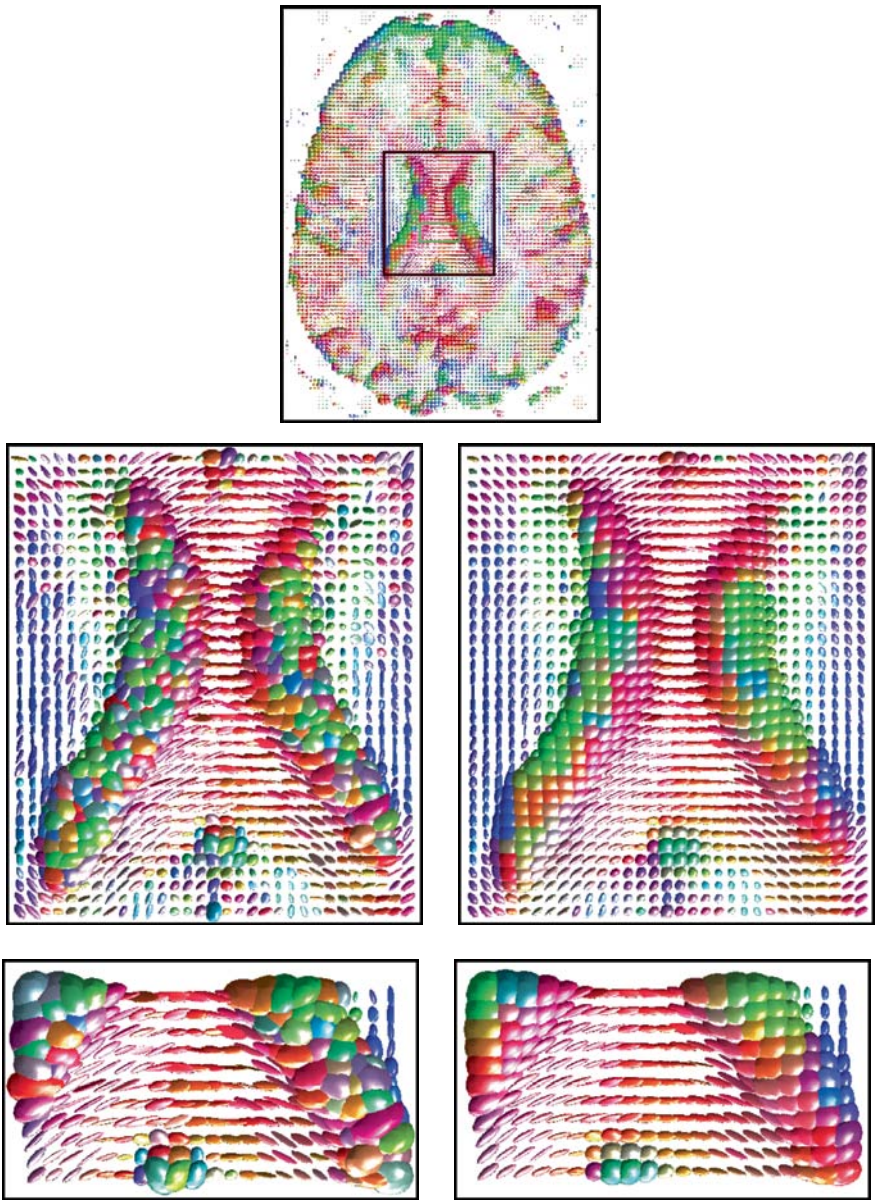


Fig. 2. *Top*: The regularized DTI slice. The marked regions of interest are zoomed below. *Middle*: Zoom on the brain ventricles area (the *black* ROI). On the left side is the original DTI data. On the right side is the regularized data using the full Iwasawa decomposition. *Bottom*: Enlargement of the corpus callosum fibers area (the *green* ROI) of the original and the regularized image. The ellipsoids are color coded according to the directions of the principal eigenvectors. These results are obtained after 50 iterations with $\beta = 1$.

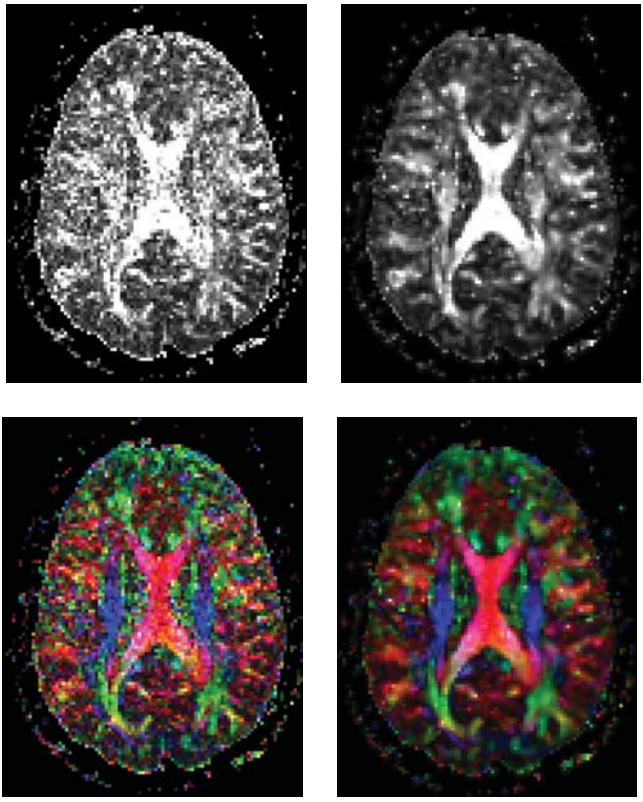


Fig. 3. *Top row*: The FA images of original data (left) and the regularized data (*right*). *Bottom*: FA weighted color coded image before the regularization (*left*) and after the regularization (*right*). These results are obtained after 50 iterations with $\beta = 1$.

of anisotropy as intensity[1]. The FA image of the smoothed tensor field shows that the noise is removed and important features (e.g., fibers) are extracted.

Finally, the third representation (Fig. 3, bottom row) is a FA weighted color-coded image of the principal eigenvectors. Thus, the FA image and the color-coded FA image illustrate separately that diffusivities as well as orientations are smoothed, while important features are preserved.

[1] The values of the FA image are calculated as follows: $FA = \sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - <D>)^2 + (\lambda_2 - \langle D \rangle)^2 + (\lambda_3 - <D>)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}$ where $< D >= \frac{1}{3}(\lambda_1 + \lambda_2 + \lambda_3)$ is the total diffusivity i.e. the mean of the three diffusivities.

5 Summary

In this chapter, we presented a geometric framework for regularization of data that is given in P_n, the space of SPD matrices. In particular, we discussed regularization of DTI data that belong to P_3. We described the DT image as a section of a fiber bundle where the metric over the section is given by means of the natural $GL(n)$-invariant metric on P_n. The manifold P_n is described thus as a Riemannian manifold. Using calculus of variations on P_n via the Polyakov action, a set of coupled Beltrami equations for the local coordinates on P_n was derived. For P_3, we obtained a set of six coupled Beltrami equations for the six Iwasawa coordinates. We have shown that there are several parameterizations on P_n. However, parameterization of the P_n manifold via Iwasawa coordinates simplifies the analytical as well as the numerical calculations. As a result, the proposed algorithm converges very fast and the results are satisfying. This was demonstrated via regularization of real noisy DTI dataset.

References

[1] Batchelor, P., G., Moakher, M., Atkinson, D., Calamante, F. and Connelly, A.: A Rigorous Framework for Diffusion Tensor Calculus. Magnetic Resonance in Medicine **53** (2005) 221–225.
[2] Chefd'hotel, C., Tschumperlé, D., Deriche, R. and Faugeras, O.: Regularizing Flows for Constrained Matrix-Valued Images. Journal of Mathematical Imaging and Vision **20** (2004) 147–162.
[3] Feddern, C., Weickert, J., Burgeth, B. and Welk, M.: Curvature–Driven PDE Methods for Matrix-Valued Images. IJCV **69** (2006) 91–103.
[4] Fillard, P., Arsigny V., Pennec, X. and Ayache, N.: Log-Euclidean Metrics for Fast and Simple Calculus on Diffusion Tensors. Magnetic Resonance in Medicine **56** (2006) 411–421.
[5] Gur, Y. and Sochen, N.: Diffusion over Tensor Fields via Lie Group PDE Flows: Lagrangian Action Approach. Proceedings of Topology and Robotics, FIM ETH, Zurich, July 2006. AMS series Contemporary Mathematics **438** (2007) 59–74.
[6] Gur, Y. and Sochen, N.: Coordinates Free Diffusion over Compact Liegroups. In proceedings of the 1st International Conference on Scale-Space and Variational Methods, Ischia Island, Italy, May 2007. Lecture Notes in Computer Science **4485** (2007) 580–591.
[7] Hairer, E., Lubich, C. and Wanner, G.: Geometric Numerical Integration. Springer, Berlin(2002).
[8] Iserles, A., Munthe-Kaas, H., Z., Nørsett, S., P. and Zanna, A.: Lie-Group Methods. Acta Numerica **9** (2000) 215–365.
[9] Jorgenson, J. and Lang, S.: $Pos_n(R)$ and Eisenstein Series. Lecture Notes in Mathematics **1868**, Springer, Berlin (2005).

[10] Jost, J.: Riemannian Geometry and Geometric Analysis, Springer, New York (2001).
[11] Kimmel, R. and Sochen, N.: Orientation Diffusion or How to Comb a Porcupine. Journal of Visual Communication and Image Representation **13** (2001) 238–248.
[12] Lang, S.: Fundamentals of Differential Geometry, Springer, New York (1999).
[13] Lenglet, C., Rousson, M., Deriche, R. and Faugeras, O.: Statistics on Multivariate Normal Distributions: A Geometric Approach and its Application to Diffusion Tensor MRI. Technical Report 5242, INRIA, June (2004).
[14] Moakher, M. and Zerai, M.: Riemannian Curvature-Driven Flows for Tensor-Valued Data. In proceedings of the 1st International Conference on Scale-Space and Variational Methods, Ischia Island, Italy, May 2007. Lecture Notes in Computer Science **4485** (2007) 592–602.
[15] Pasternak, O., Sochen, N. and Asaf, Y.: Separation of White Matter Fascicles From Diffusion MRI Using Φ-Functional Regularization. The International Conference on Magnetic Resonance in Medicine, Kyoto, Japan (2004).
[16] Pennec, X., Fillard, P. and Ayache, N.: A Riemannian Framework for Tensor Computing. International Journal of Computer Vision **66** (2006) 41–66.
[17] Perona, P.: Orientation Diffusion. IEEE Transactions on Image Processing **7** (1998) 457–467.
[18] Polyakov, A., M.: Quantum Geometry of Bosonic Strings. Physics Letters B **103** (1981) 207–210.
[19] Shafrir, D. Sochen, N., Deriche, R.: Regularization of Mappings between Implicit Manifolds of Arbitrary Dimension and Codimension, VLSM '05: Proceedings of the 3rd IEEE Workshop on Variational, Geometric and Level-set methods in Computer Vision, Beijing, China (2005). Lecture Notes in Computer Science **3752** (2005) 344–355.
[20] Sochen, N., Kimmel, R. and Malladi, R.: From High Energy Physics to Low Level Vision. Scale-Space Theories in Computer Vision (1997) 236–247.
[21] Sochen, N., Deriche, R. and Lopez-Perez, L.: The Beltrami Flow over Implicit Manifolds, ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision (2003) 832–839.
[22] Tang, B., Sapiro, G. and Caselles, V.: Direction Diffusion. IEEE International Conference on Computer Vision (1999) 1245–1252.
[23] Terras, A.: Harmonic Analysis on Symmetric Spaces and Applications, vol. 2. Springer, New York (1988).
[24] Tschumperlé, D. and Deriche, R.: Orthonormal Vector Sets Regularization with PDE's and Applications. International Journal of Computer Vision **50** (2002) 237–252.

[25] Weickert, J. and Hagen, H. (Eds.): Visualization and Processing of Tensor Fields. Springer, Berlin (2005).
[26] Westin, C., -F., Maier, S., E., Khidhir, B., Everett, P., Jolesz, F., A. and Kikinis, R.: Image Processing for Diffusion Tensor Magnetic Resonance Imaging. Proceedings of the Second International Conference on Medical Image Computing and Computer-Assisted Intervention. Lecture Notes In Computer Science **1679** (1999) 441–452.
[27] Zhizhou Wang, Vemuri, B., C., Chen, Y., Mareci, T., H., A.: Constrained Variational Principle for Direct Estimation and Smoothing of the Diffusion Tensor Field from Complex DWI. IEEE Transactions on Medical Imaging **23** (2004) 930–939.

Variational Methods for Denoising Matrix Fields

S. Setzer[1], G. Steidl[1], B. Popilka[1], and B. Burgeth[2]

[1] University of Mannheim, Department of Mathematics and Computer Science, Mannheim, Germany
{ssetzer,steidl}@math.uni-mannheim.de, bpopilka@rumms.uni-mannheim.de
[2] Saarland University, Mathematical Image Analysis Group, Faculty of Mathematics and Computer Science
burgeth@mia.uni-saarland.de

Summary. The restoration of scalar-valued images via minimization of an energy functional is a well-established technique in image processing. Recently, higher-order methods have proved their advantages in edge preserving image denoising. In this chapter, we transfer successful techniques like the minimization of the Rudin–Osher–Fatemi functional and the infimal convolution to matrix fields, where our functionals couple with different matrix channels. For the numerical computation, we use second-order cone programming. Moreover, taking the operator structure of matrices into account, we consider a new operator-based regularization term. This is the first variational approach for denoising tensor-valued data that takes the operator structure of matrices, in particular the operation of matrix multiplication into account. Using matrix differential calculus, we deduce the corresponding Euler–Lagrange equation and apply it for the numerical solution by a steepest descent method.

1 Introduction

Matrix-valued data have gained significant importance in recent years, e.g. in diffusion tensor magnetic resonance imaging (DT-MRI) and technical sciences (inertia, diffusion, stress, and permittivity tensors). Like most measured data, these matrix-valued data are also polluted by noise and require restoration. Regularization methods have been applied very successfully for denoising of scalar-valued images, where recently higher-order methods, e.g. in connection with the infimal convolution [9] have provided impressive results. In this chapter, we want to transfer these techniques to matrix fields. However, unlike vectors, matrices can be multiplied providing matrix-valued polynomials and also functions of matrices. These useful notions rely decisively on the strong interplay between the different matrix entries. Thus the corresponding regularization terms should take the relation between the different matrix channels into account.

Filtering methods for matrix fields based on matrix-valued nonlinear partial differential equations (PDEs) have been proposed in [6] for singular and in [7] for Perona–Malik-type diffusivity functions. These approaches rely on an operator-algebraic point of view on symmetric matrices as instances of self-adjoint Hilbert space operators, and are based on a basic differential calculus for matrix fields. This chapter continues our work in [22] and contains the first variational approach for denoising tensor fields that takes the operator structure of matrices into account. Since the proposed techniques exploit the greater algebraic potential of matrices, when compared to vectors, they ensure appropriate matrix channel coupling, and more important, are also applicable to indefinite matrix fields.

Approaches to positive definite matrix field filtering with a differential geometric background have been suggested previously [9, 14, 19, 23, 29]. In their setting, the set of positive definite matrices is endowed with a structure of a manifold, and the methodology is geared towards application to DT-MRI data. For other smoothing techniques for DT-MRI data, we refer to [24, 28]. Comprehensive survey articles on the analysis of matrix fields utilizing a wide range of different techniques can be found in [27] and the literature cited therein.

This chapter is organized as follows: In Sect. 2, we start by considering various variational methods for denoising scalar-valued images, in particular we adapt the infimal convolution technique to our discrete setting and introduce a corresponding simplified version. In Sect. 3, we turn to the matrix-valued setting. After briefing the necessary preliminaries in Sect. 3.1, we consider component-based regularization terms related to the Rudin–Osher–Fatemi approach and to infimal convolution in Sect. 3.2. These functionals couple the different matrix channels as originally proposed by Tschumperlé and Deriche in 2001 [23]. In Sect. 3.3, we introduce a new operator-based functional and derive the corresponding Euler–Lagrange equation, which contains the Jordan product of matrices. In contrast to the ordinary matrix product, the Jordan product of two symmetric matrices is again a symmetric matrix. Finally, in Sect. 4, we present numerical examples comparing the component-based and the operator-based approach as well as the first-order and infimal convolution methods.

2 Variational Methods for Scalar-Valued Images

2.1 First-Order Methods

A well-established method for restoring a scalar-valued image u from a given degraded image f consists in calculating

$$\arg\min_u \int_\Omega (f-u)^2 \;+\; \alpha\, \Phi(|\nabla u|^2)\, dxdy \tag{1}$$

with a regularization parameter $\alpha > 0$ and an increasing function $\Phi : [0, \infty] \to \mathbb{R}$ in the penalizing term. The first summand encourages similarity between the restored image and the original one, while the second term rewards smoothness. The appropriate choice of the function Φ ensures that important image structures such as edges are preserved while areas with small gradients are smoothed.

The Euler–Lagrange equation of (1) is given by

$$0 = f - u + \alpha \operatorname{div}(\Phi'(|\nabla u|^2)\nabla u). \tag{2}$$

Thus, the minimizer u can be considered as the steady state ($t \to \infty$) of the reaction–diffusion equation

$$\partial_t u = f - u + \alpha \operatorname{div}(\Phi'(|\nabla u|^2)\nabla u) \tag{3}$$

with initial image $u(\cdot, 0) = f$ and homogeneous Neumann boundary conditions. For an interpretation of (2) as a fully implicit time discretization of a diffusion equation see [21, 25].

In this chapter, we mainly focus on the frequently applied ROF-model introduced by Rudin, Osher, and Fatemi [20], which uses the function

$$\Phi(s^2) := \sqrt{s^2} = |s|. \tag{4}$$

For this function, the functional (1) is strictly convex and the penalizing functional $J(u) = \int_\Omega \sqrt{u_x^2 + u_y^2}\, dxdy$ is positively homogeneous, i.e. $J(\alpha u) = \alpha J(u)$ for $\alpha > 0$. Since Φ in (4) is not differentiable at zero, we have to use its modified version

$$\Phi(s^2) = \sqrt{s^2 + \varepsilon^2}, \tag{5}$$

with a small additional parameter ε, if we want to apply (3).

For digital image processing, we consider a discrete version of (1). Let us introduce this discrete version in matrix–vector notation. For the sake of simplicity, we restrict our attention to quadratic images $\mathbf{f} \in \mathbb{R}^{n,n}$. We transform $\mathbf{f}$ into a vector $f \in \mathbb{R}^N$ with $N = n^2$ in the following way

$$\operatorname{vec} \mathbf{f} := \begin{pmatrix} f_0 \\ \vdots \\ f_{n-1} \end{pmatrix},$$

where f_j denotes the j-th column $\mathbf{f}$. The partial derivatives in (1) are discretized by forward differences. More precisely, we introduce the difference matrix $D_1 := \begin{pmatrix} D_x \\ D_y \end{pmatrix}$ with $D_x := I_n \otimes D$, $D_y := D \otimes I_n$ and

$$D := \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ & & \ddots & \ddots & \ddots & & \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -1 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}. \tag{6}$$

Here $A \otimes B$ is the *Kronecker product* (tensor product) of A and B. Now our discrete version of (1) reads

$$\arg\min_{u\in\mathbb{R}^N} \frac{1}{2}\,||f-u||_{\ell_2}^2 \;+\; \alpha||\,|D_1 u|\,\|_{\ell_1}, \tag{7}$$

where $|D_1 u| \in \mathbb{R}^N$ is defined componentwise by $|D_1 u| = \left((D_x u)^2 + (D_y u)^2\right)^{\frac{1}{2}}$. For computations, it is useful to consider the dual formulation of (7). Since we will need the dual form of various similar functionals later, let us consider more generally

$$\arg\min_{u\in\mathbb{R}^N} \frac{1}{2}\,||f-u||_{\ell_2}^2 \;+\; \alpha\,||\,|Lu|\,\|_{\ell_1}, \tag{8}$$

with $L \in \mathbb{R}^{pN,N}$ and $|w|(j) = \left(\sum_{k=0}^{p-1} w(j+kN)^2\right)^{1/2}$, $j = 1,\ldots,N$. For $L := D_1$, the functional (8) coincides with (7). Since the penalizing functional $J(u) = ||\,|Lu|\,\|_{\ell_1}$ is positively homogeneous, its Legendre–Fenchel dual J^* is the indicator function of the convex set

$$\mathcal{C} := \{v : \langle v, w\rangle \le J(w), \quad \forall w \in \mathbb{R}^N\}$$

and the minimizer $\hat{u}$ of (8) is given by $\hat{u} = f - \Pi_{\alpha\mathcal{C}} f$, where $\Pi_{\mathcal{C}}$ denotes the orthogonal projection of f onto $\mathcal{C}$. It can be proved, see, e.g. [12], that

$$\mathcal{C} := \{L^{\mathrm{T}} V :\; \|\,|V|\,\|_\infty \le 1\} \tag{9}$$

and consequently $\hat{u} = f - L^{\mathrm{T}} V$, where V is a solution of

$$\|f - L^{\mathrm{T}} V\|_{\ell_2}^2 \quad \to \min \quad \text{s.t.} \quad \|\,|V|\,\|_{\ell_\infty} \le \alpha. \tag{10}$$

This minimization problem can be solved by second-order cone programming (SOCP) [13], or by Chambolle's descent algorithm [8, 12].

2.2 Higher-Order Methods

For various denoising problems, higher-order methods with functionals including higher-order derivatives have been proved useful. In particular, the drawback of so-called staircasing effect known from the ROF-model can be avoided in this way. An example is shown in Fig. 1.

Here we focus only on the infimal convolution method introduced by Chambolle and Lions [9]. We consider

$$\arg\min_{u\in\mathbb{R}^N} \frac{1}{2}\,||f-u||_{\ell_2}^2 \;+\; (J_1 \square J_2)(u), \tag{11}$$

where,

$$(J_1 \square J_2)(u) := \inf_{u_1,u_2\in\mathbb{R}^N} \{J_1(u_1) + J_2(u_2) \;:\; u_1 + u_2 = u\}$$

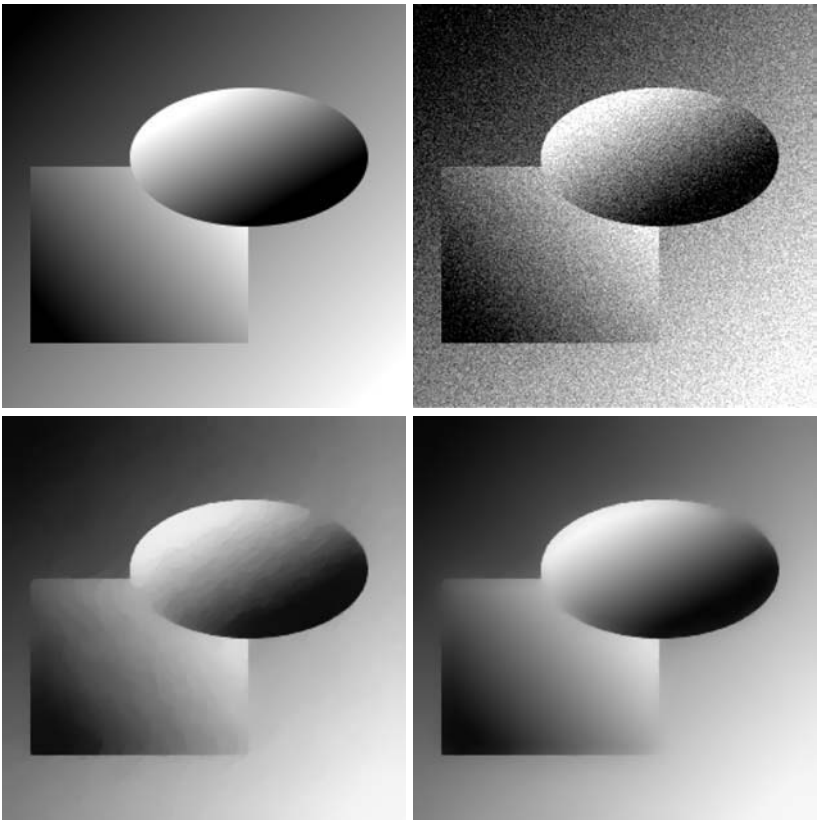


Fig. 1. *T*op: Original image (*left*), noisy image (*right*). *B*ottom: Denoised image by the ROF method (7) with $\alpha = 50$ and by the modified infimal convolution approach (15) with $\alpha_1 = 50$, $\alpha_2 = 180$ (right). The ROF based image shows the typical staircasing effect. The performance of the inf-conv method is remarkable

denotes the so-called *infimal convolution* of J_1 and J_2. Note that the infimal convolution is closely related to the dilation operation in morphological image processing. In this chapter, we will only deal with

$$J_1(u) := \alpha_1 \| \, |D_1 u| \, \|_{\ell_1}, \quad \text{and} \quad J_2(u) := \alpha_2 \| \, |D_2 u| \, \|_{\ell_1}, \tag{12}$$

where $D_2 := \begin{pmatrix} D_{xx} \\ D_{yy} \end{pmatrix} = \begin{pmatrix} I_n \otimes D^{\mathrm{T}} D \\ D^{\mathrm{T}} D \otimes I_n \end{pmatrix}$ is a discrete second-order derivative operator. Alternatively, we can also use $|(D_{xx}^{\mathrm{T}}, D_{yy}^{\mathrm{T}}, D_{xy}^{\mathrm{T}}, D_{yx}^{\mathrm{T}})^{\mathrm{T}} u|$, which is related to the Frobenius norm of the Hessian of u. It is easy to check that for (12), the infimum in (11) is attained, so that (11) can be rewritten as

$$\arg \min_{u_1, u_2 \in \mathbb{R}^N} \frac{1}{2} \, ||f - u_1 - u_2||_{\ell_2}^2 \; + \; \alpha_1 \| \, |D_1 u_1| \, \|_{\ell_1} + \alpha_2 \| \, |D_2 u_2| \, \|_{\ell_1}. \tag{13}$$

Using that $(J_1 \square J_2)^*(u) = J_1^*(u) + J_2^*(u)$ for proper convex functionals, the minimizer $\hat{u}$ of (11) is given by $\hat{u} = f - \Pi_{\alpha_1 \mathcal{C}_1 \cap \alpha_2 \mathcal{C}_2} f$, where J_i^* are the indicator functions of the convex sets $\mathcal{C}_i$ associated with J_i, $i = 1, 2$. Consequently, by (9), we obtain that $\hat{u} = f - v$, where v is the solution of

$$\begin{aligned} \|f - v\|_2^2 \quad \to \min \quad \text{s.t.} \quad & v = D_1^{\mathrm{T}} V_1 = D_2^{\mathrm{T}} V_2, \\ & \| \, |V_1| \, \|_\infty \le \alpha_1, \; \| \, |V_2| \, \|_\infty \le \alpha_2. \end{aligned} \tag{14}$$

Now we have that

$$D_2^{\mathrm{T}} = D_1^{\mathrm{T}} \begin{pmatrix} D_x & 0 \\ 0 & D_y \end{pmatrix}.$$

Consequently, assuming that $V_1 = \begin{pmatrix} D_x & 0 \\ 0 & D_y \end{pmatrix} V_2$, we may rewrite (14) as

$$\begin{aligned} \|f - D_2^{\mathrm{T}} V\|_2^2 \quad \to \min \quad \text{s.t.} \quad & \| \, | \begin{pmatrix} D_x V^1 \\ D_y V^2 \end{pmatrix} | \, \|_\infty \le \alpha_1, \\ & \| \, |V| \, \|_\infty \le \alpha_2. \end{aligned} \tag{15}$$

Note that this minimization problem is similar but not equivalent to (14). The solution of (14), (15) or of the primal problem (13) can be computed by SOCP.

3 Variational Methods for Matrix-Valued Images

In this section, we want to transfer the regularization methods for scalar-valued images reviewed in the previous section to matrix-valued images. While it seems to be straightforward to replace the square of scalar values in the data fitting term by the squared Frobenius norm, the penalizing term may be established in different ways as we will explain in Sects. 3.2 and 3.3. First we need is some notation.

3.1 Preliminaries

Let $\mathrm{Sym}_m(\mathbb{R})$ be the vector space of symmetric $m \times m$ matrices. This space can be treated as a Euclidean vector space with respect to the trace inner product

$$\langle A, B \rangle := \operatorname{tr} AB = (\operatorname{vec} A, \operatorname{vec} B),$$

where $(\cdot, \cdot)$ on the right-hand side denotes the Euclidean inner vector product. Then

$$\langle A, A \rangle = \operatorname{tr} A^2 = \|A\|_F^2 = \|\operatorname{vec} A\|_{\ell_2}^2$$

is the squared *Frobenius norm* of A. In addition to this, vector structure matrices are (realizations of) linear operators and carry the corresponding features. In particular they can be applied successively. Unfortunately, the

original matrix multiplication does not preserve the symmetry of the matrices. The *Jordan-product* of matrices $A, B \in \mathrm{Sym}_m(\mathbb{R})$ defined by

$$A \bullet B := \frac{1}{2}(AB + BA)$$

preserves the symmetry of the matrices but not the positive semi-definiteness.

In $\mathrm{Sym}_m(\mathbb{R})$, the positive semi-definite matrices $\mathrm{Sym}_m^+(\mathbb{R})$ form a closed convex set whose interior consists of the positive definite matrices. More precisely, $\mathrm{Sym}_m^+(\mathbb{R})$ is a cone with base [1, 4, 5]. In our numerical examples, we will consider only the cases $m = 2$ and $m = 3$, where the positive definite matrices can be visualized as ellipses, resp. ellipsoids, i.e. $A \in \mathrm{Sym}_3^+(\mathbb{R})$ can be visualized as the ellipsoid

$$\{x \in \mathbb{R}^3 : x^{\mathrm{T}} A^{-2} x = 1\}$$

whose axis lengths are given by the eigenvalues of A.

3.2 Component-Based Regularization

In the following, let $F : \mathbb{R}^2 \supset \Omega \to \mathrm{Sym}_m(\mathbb{R})$ be a matrix field. In this subsection, we transfer (1) to matrix-valued images in a way that emphasizes the individual matrix components. We will see that for this approach, the denoising methods from the previous section can be translated in a direct way. However, the specific question arises whether these methods preserve positive definiteness.

Instead of (1), we are dealing with

$$\arg\min_U \int_\Omega \|F - U\|_F^2 + \alpha \Phi\left(\mathrm{tr}\,(U_x^2 + U_y^2)\right)\, dxdy, \tag{16}$$

where the partial derivatives are taken componentwise. The penalizing term $J(U)$ in (16) was first mentioned by Deriche and Tschumperlé [23]. Rewriting this term as

$$J(U) = \int_\Omega \Phi\big(\|U_x\|_F^2 + \|U_y\|_F^2\big)\, dxdy = \int_\Omega \Phi\big(\sum_{j,k=1}^{n} \nabla u_{jk}^{\mathrm{T}} \nabla u_{jk}\big)\, dxdy \tag{17}$$

we see its component-based structure implied by the Frobenius norm. However, due to the sum on the right-hand side, Φ is applied to coupled matrix coefficients. By [3], the Euler–Lagrange equation of (17) is given by

$$0 = F - U + \alpha\left(\partial_x(\Phi'(\mathrm{tr}(U_x^2 + U_y^2))U_x + \partial_y(\Phi'(\mathrm{tr}(U_x^2 + U_y^2))U_y\right). \tag{18}$$

Again, we are only interested in the function Φ given by (4).

For computations, we consider the discrete counterpart of (16), where we again replace the derivative operators by simple forward difference operators

$$\arg\min_U \sum_{i,j=0}^{N-1} \frac{1}{2} \|F(i,j) - U(i,j)\|_F^2 \; + \; \alpha\, J(U), \tag{19}$$

$$J(U) := \sum_{i,j=0}^{N-1} \left(\|U(i,j) - U(i-1,j)\|_F^2 + \|U(i,j) - U(i,j-1)\|_F^2 \right)^{1/2}$$

with $U(-1,j) = U(i,-1) = 0$. The functional in (19) is strictly convex and thus has a unique minimizer.

We say that the discrete matrix field $F : \mathbb{Z}_n^2 \to \mathrm{Sym}_m^+(\mathbb{R})$ has all eigenvalues in an interval $\mathcal{I}$ if all the eigenvalues of every matrix $F(i,j)$ of the field lie in $\mathcal{I}$. By the following proposition, the minimizer of (19) preserves positive definiteness.

Proposition 1. *Let all eigenvalues of* $F : \mathbb{Z}_n^2 \to \mathrm{Sym}_m^+(\mathbb{R})$ *be contained in the interval* $[\lambda_{\min}, \lambda_{\max}]$. *Then the minimizer* U *of* (19) *has all eigenvalues in* $[\lambda_{\min}, \lambda_{\max}]$.

The proof in the appendix is based on Courant's Min–Max principle and the projection theorem for convex sets.

To see how the methods from Sect. 2 carry over to matrix fields, we rewrite (19) in matrix–vector form. To this end, let $N = n^2$ and $M := m(m+1)/2$. We change $F : \mathbb{Z}_n^2 \to \mathrm{Sym}_m(\mathbb{R})$ into the vector

$$f := \begin{pmatrix} \varepsilon_{1,1} & \mathrm{vec}\,(F_{1,1}) \\ \vdots & \\ \varepsilon_{1,m} & \mathrm{vec}\,(F_{1,m}) \\ \varepsilon_{2,2} & \mathrm{vec}\,(F_{2,2}) \\ \vdots & \\ \varepsilon_{2,m} & \mathrm{vec}\,(F_{2,m}) \\ \vdots & \\ \varepsilon_{m,m} & \mathrm{vec}\,(F_{m,m}) \end{pmatrix} \in \mathbb{R}^{MN},$$

where $F_{k,l} := (F_{k,l}(i,j))_{i,j=0}^{n-1}$ and $\varepsilon_{k,l} := \begin{cases} \sqrt{2} \text{ for } k \neq l \\ 1 \quad \text{otherwise} \end{cases}$.

Then (19) becomes

$$\arg\min_{u \in \mathbb{R}^{MN}} \frac{1}{2} \|f - u\|_{\ell_2}^2 \; + \; \alpha \| \, | \, (I_M \otimes D_1)\, u | \, \|_{\ell_1}. \tag{20}$$

This problem has just the structure of (8) with $L := I_M \otimes D_1 \in \mathbb{R}^{2MN,MN}$ and $p = 2M$. Thus it can be solved by applying SOCP to its dual given by (10).

Similarly, we can transfer the infimal convolution approach to the matrix-valued setting. Obviously, we have to find

$$\arg \min_{u_1,u_2 \in \mathbb{R}^{MN}} \frac{1}{2}\|f-u_1-u_2\|_{\ell_2}^2+\alpha_1\| \, |(I_M \otimes D_1)u_1| \, \|_{\ell_1}+\alpha_2\| \, |(I_M \otimes D_2)u_1| \, \|_{\ell_1}.$$

In our numerical examples, we solve the corresponding modified dual problem

$$\begin{aligned} \|f - (I_M \otimes D_2^{\mathrm{T}})\, V\|_2^2 \;\to \min \quad \text{s.t.} \;\; & \| \, | \left(I_M \otimes \begin{pmatrix} D_x & 0 \\ 0 & D_y \end{pmatrix} \right) V | \, \|_\infty \le \alpha_1, \\ & \| \, |V| \, \|_\infty \le \alpha_2, \end{aligned} \tag{21}$$

by SOCP.

3.3 Operator-Based Regularization

In this subsection, we introduce a regularization term that emphasizes the operator structure of matrices. For $A \in \mathrm{Sym}_m(\mathbb{R})$ with eigenvalue decomposition $A = Q\Lambda Q^{\mathrm{T}}$, let $\Phi(A) = Q\Phi(\Lambda)Q^{\mathrm{T}}$, where $\Lambda := \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$ and $\Phi(\Lambda) := \mathrm{diag}\,(\Phi(\lambda_1), \ldots, \Phi(\lambda_n))$. We consider the following minimization problem

$$\arg\min_U \int_\Omega \|F - U\|_F^2 + \alpha \,\mathrm{tr}\left(\Phi(U_x^2 + U_y^2)\right) dxdy. \tag{22}$$

In contrast to (16), the trace is taken after applying Φ to the matrix $U_x^2 + U_y^2$. By the next proposition, we have that the functional in (22) with Φ defined by (4) is strictly convex.

Proposition 2. *For given $F : \mathbb{R}^2 \supset \Omega \to \mathrm{Sym}_m(\mathbb{R})$ and $\Phi(s^2) = \sqrt{s^2}$, the functional in* (22) *is strictly convex.*

The proof is given in the appendix.

Remark. The solution of (22) in general does not preserve positive definiteness. For an example see [22].

The next proposition shows that the functional (22) has an interesting Gâteaux derivative.

Proposition 3. *Let Φ be a differentiable function. Then the Euler–Lagrange equations for minimizing the functional in* (22) *are given by*

$$0 = F - U + \alpha \left(\partial_x \left(\Phi'(U_x^2 + U_y^2) \bullet U_x\right) + \partial_y \left(\Phi'(U_x^2 + U_y^2) \bullet U_y\right)\right). \tag{23}$$

The proof of the proposition is provided in the appendix and makes use of matrix differential calculus. In contrast to (18), the Jordan product of matrices appears in (23) and the function Φ' is applied to matrices.

We apply Proposition 3 to compute a minimizer of (22) by using a difference method to solve the corresponding reaction–diffusion equation for $t \to \infty$

$$U_t = F - U + \alpha \left(\partial_x \left(\Phi'(U_x^2 + U_y^2) \bullet U_x \right) + \partial_y \left(\Phi'(U_x^2 + U_y^2) \bullet U_y \right) \right) \tag{24}$$

with Φ as in (5), homogeneous Neumann boundary conditions and initial value F. More precisely, we use the iterative scheme

$$U^{(k+1)} = (1-\tau)U^{(k)} + \tau F + \tau\alpha \left(\partial_x \left(G^{(k)} \bullet U_x^{(k)} \right) + \partial_y \left(G^{(k)} \bullet U_y^{(k)} \right) \right) \tag{25}$$

with sufficiently small time step size τ and $G^{(k)} := \Phi'((U_x^{(k)})^2 + (U_y^{(k)})^2)$. The inner derivatives including those in G are approximated by forward differences and the outer derivatives by backward differences, so that the penalizing term becomes

$$\begin{aligned} &\frac{1}{h_1} \left(G(i,j) \bullet \frac{U(i+1,j) - U(i,j)}{h_1} - G(i-1,j) \bullet \frac{U(i,j) - U(i-1,j)}{h_1} \right) \\ &+ \frac{1}{h_2} \left(G(i,j) \bullet \frac{U(i,j+1) - U(i,j)}{h_2} - G(i,j-1) \bullet \frac{U(i,j) - U(i,j-1)}{h_2} \right), \end{aligned}$$

where h_i, $i = 1, 2$ denote the pixel distances in x- and y-direction. Alternatively, we have also worked with symmetric differences for the derivatives. In this case, we have to replace, e.g. $G(i,j)$ in the first summand by $\tilde{G}(i+1,j) + \tilde{G}(i,j))/2$ and $\tilde{G}$ is now computed with symmetric differences.

Finally, we mention that a diffusion equation related to (24) was examined in [6]. Moreover, in [30] an anisotropic diffusion concept for matrix fields was presented where the function Φ was also applied to a matrix.

4 Numerical Results

Finally, we present numerical results demonstrating the performance of the different methods. All algorithms were implemented in MATLAB. Moreover, we have used the software package MOSEK for SOCP.

SOCP [16] amounts to minimize a linear objective function subject to the constraints that several affine functions of the variables have to lie in a *second-order cone* $\mathcal{C}^{n+1} \subset \mathbb{R}^{n+1}$ defined by the convex set

$$\mathcal{C}^{n+1} = \left\{ \begin{pmatrix} x \\ \bar{x}_{n+1} \end{pmatrix} = (x_1, \ldots, x_n, \bar{x}_{n+1})^\mathrm{T} : \|x\|_2 \leq \bar{x}_{n+1} \right\}.$$

With this notation, the general form of a SOCP is given by

$$\inf_{x \in \mathbb{R}^n} f^\mathrm{T} x \quad \text{s.t.} \begin{pmatrix} A_i x + b_i \\ c_i^\mathrm{T} x + d_i \end{pmatrix} \in \mathcal{C}^{n+1}, \; i = 1, \ldots, r. \tag{26}$$

Alternatively, one can also use the rotated version of the standard cone:

$$\mathcal{K}^{n+2} := \Big\{ \big(x, \bar{x}_{n+1}, \bar{x}_{n+2}\big)^{\mathrm{T}} \in \mathbb{R}^{n+2} : \|x\|_2^2 \leq 2\,\bar{x}_{n+1}\bar{x}_{n+2} \Big\}.$$

This allows us to incorporate quadratic constraints. Problem (26) is a convex program for which efficient, large scale solvers are available [18]. For rewriting our minimization problems as a SOCP, see [22].

We start by comparing the component-based regularization with the operator-based regularization. First we focus on the 1D matrix-valued function $F : \mathbb{Z}_{16} \to \mathrm{Sym}_2^+(\mathbb{R})$ in Fig. 2. We added white Gaussian noise with standard deviation 0.1 to all components of the original data in $[0, 1]$. Then we computed the minimizer of the component-based functional (19) (left) and of the operator-based functional (22) (right) both by SOCP. The latter was computed using the fact that the operator-based functional can be rewritten for tensor-valued 1D signals as

$$\arg\min_U \int_\Omega \|F - U\|_F^2 + \alpha \,\mathrm{tr}|U_x|\, dx$$

with

$$\mathrm{tr}\,|U| = \max\{\big(4u_{12}^2 + (u_{11} - u_{22})^2\big)^{1/2}, |u_{11} + u_{22}|\}$$

for $U \in \mathrm{Sym}_2^+(\mathbb{R})$, cf. [22]. The middle of the figure shows the Frobenius norm of the difference between the original and the denoised signal $(\sum_{i=1}^N \|F(i) - \hat{U}(i)\|_F^2)^{1/2}$ in dependence on the regularization parameter α. We remark that the shape of the curve and its minimal point do not change if we use the error measure $\sum_{i=1}^N \|F(i) - \hat{U}(i)\|_F$ instead. The actual minima with respect to the Frobenius norm are given by min = 0.2665 at $\alpha = 0.8$ for (19) and by min = 0.2276 at $\alpha = 0.8$ for (22). The denoised signals corresponding to the smallest error in the Frobenius-norm are shown at the bottom of Fig. 2. It appears that the operator-based method performs slightly better with respect to these error norms. The visual results confirm this impression. The larger ellipses obtained by the first method (19) slightly overlap while there are gaps between the smaller ones. We do not have this effect for the minimizer of (22) on the right-hand side.

Next we consider the 2D matrix–valued function $F : \mathbb{Z}_{32}^2 \to \mathrm{Sym}_2^+(\mathbb{R})$ in Fig. 3. To all components of the original data in [0,2] we added white Gaussian noise with standard deviation 0.6. As in the previous example, we compare the minimizer of the component-based approach (16) resp. (19) with those of the operator-based approach (22). For computing the minimizer of the first functional we applied SOCP, while the second one was computed via the reaction–diffusion equation (25) with time step size $\tau = 0.00025$. The iterations were stopped when the relative error in the ℓ_2-norm between two consecutive iterations became smaller than 10^{-8} (approximately 20,000 iterations) although the result becomes visually static much earlier. The middle

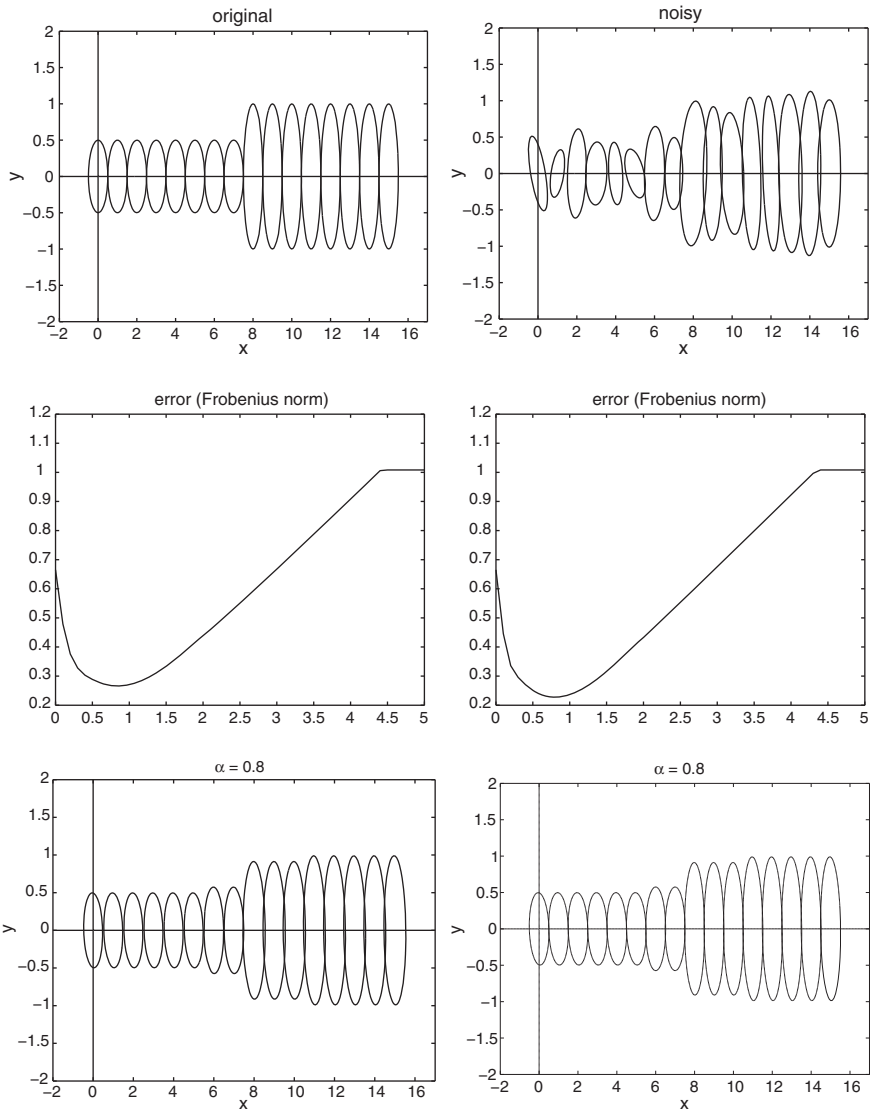


Fig. 2. Denoising of a matrix-valued signal. *T*op: Original signal (*left*), noisy signal (*right*). *M*iddle: Error of the Frobenius norm in dependence on the regularization parameter α for the minimizers of the component-based functional (19) (*left*) and the operator-based functional (22) (*right*). *B*ottom: Denoised image for α corresponding to the smallest error in the Frobenius norm for the component-based functional (*left*) and the operator-based functional (*right*)

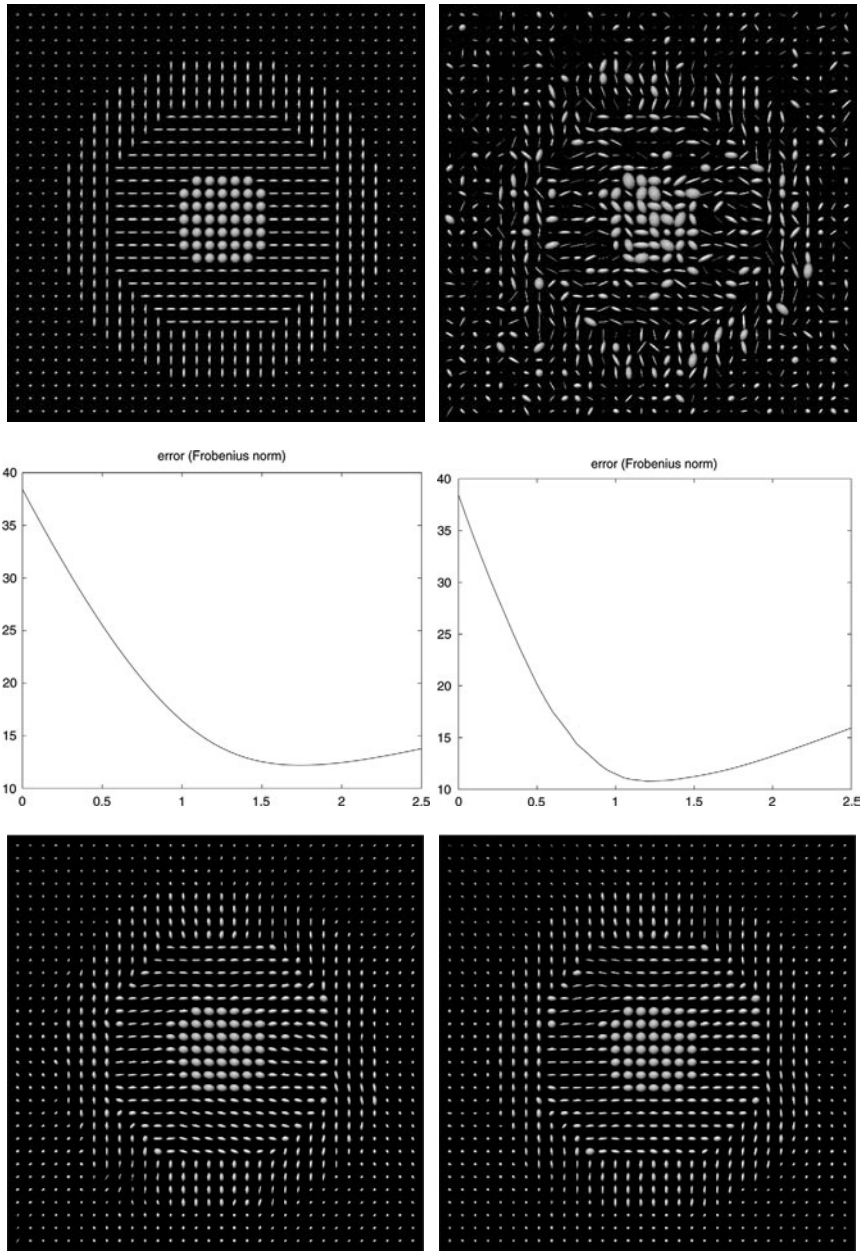


Fig. 3. Denoising of a $\mathrm{Sym}_2(\mathbb{R})$-valued image. *T*op: Original image (*left*), noisy image (*right*). Middle: Error of the Frobenius norm in dependence on the regularization parameter α for the minimizers of the component-based functional (19) (*left*) and the operator-based functional (22) (*right*). *B*ottom: Denoised image for α corresponding to the smallest error in the Frobenius norm for the component-based functional (*left*) and the operator-based functional (*right*)

row of the figure contains the error plots for both methods. The actual minima with respect to. the Frobenius norm are given by min = 12.19 at $\alpha = 1.75$ for (19) and by min = 10.79 at $\alpha = 1.2$ for (22). Hence, with respect to the computed errors the operator-based method outperforms the component-based one. The corresponding denoised images are shown in the bottom row of the figure.

In the following two examples, we consider bivariate matrix-valued functions which map to $\mathrm{Sym}_3(\mathbb{R})$. We use ellipsoids to visualize this kind of data as described in Sect. 3.1. Furthermore, the colour of the ellipsoid associated with a matrix A is chosen with respect to the normalized eigenvector corresponding to the largest eigenvalue of A. Figure 4 shows a function $F : \mathbb{Z}_{12}^2 \to \mathrm{Sym}_3(\mathbb{R})$. As in the previous case, we added white Gaussian noise to all components. The matrix components of the original data lie in the interval $[-0.5, 0.5]$ and the standard deviation of the Gaussian noise is 0.06. The denoising results are displayed in the last two rows of Fig. 4. We computed the minimizers of the component-based method (19) (top) by SOCP. The smallest error, measured in the Frobenius-norm, is 1.102 and was obtained for the regularization parameter $\alpha = 0.11$. In addition, we considered the minimizer of the infimal convolution approach (21) (bottom). Again, we applied SOCP and found the optimal regularization parameters to be $\alpha_1 = 0.14$ and $\alpha_2 = 0.08$ for this method. The corresponding Frobenius-norm error is 0.918. We see that the infimal convolution approach is also suited for matrix-valued data.

In our final experiment, we applied the two component-based methods (19) and (21) to a larger data set. Figure 5 shows the original data and the minimizers of (19) and (21). The components of the original data lie in $[-4000, 7000]$ and we used the regularization parameters $\alpha = 600$ for (19) and $\alpha_1 = 500$, $\alpha_2 = 600$ for (21), respectively.

A Proofs

*P*roof of Proposition 1. Using that the minimal and maximal eigenvalues $\lambda_{\min}(A), \lambda_{\max}(A)$ of a symmetric matrix A fulfill

$$\lambda_{\min}(A) = \min_{\|v\|=1} v^{\mathrm{T}} A\, v, \quad \lambda_{\max}(A) = \max_{\|v\|=1} v^{\mathrm{T}} A\, v,$$

it is easy to check that the set $\mathcal{C}$ of matrices having all eigenvalues in $[\lambda_{\min}, \lambda_{\max}]$ is convex and closed. Let $\mathcal{J}$ be the functional in (19). Assume that some matrices $\hat{U}(i,j)$ are not contained in $\mathcal{C}$. Let $P\hat{U}(i,j)$ denote the orthogonal projection (with respect to the Frobenius norm) of $\hat{U}(i,j)$ onto $\mathcal{C}$. Then we obtain by the projection theorem [11, p. 269] that

$$\begin{aligned} \|F(i,j) - P\hat{U}(i,j)\|_F &\leq \|F(i,j) - \hat{U}(i,j)\|_F, \\ \|P\hat{U}(i,j) - P\hat{U}(k,l)\|_F &\leq \|\hat{U}(i,j) - \hat{U}(k,l)\|_F. \end{aligned}$$

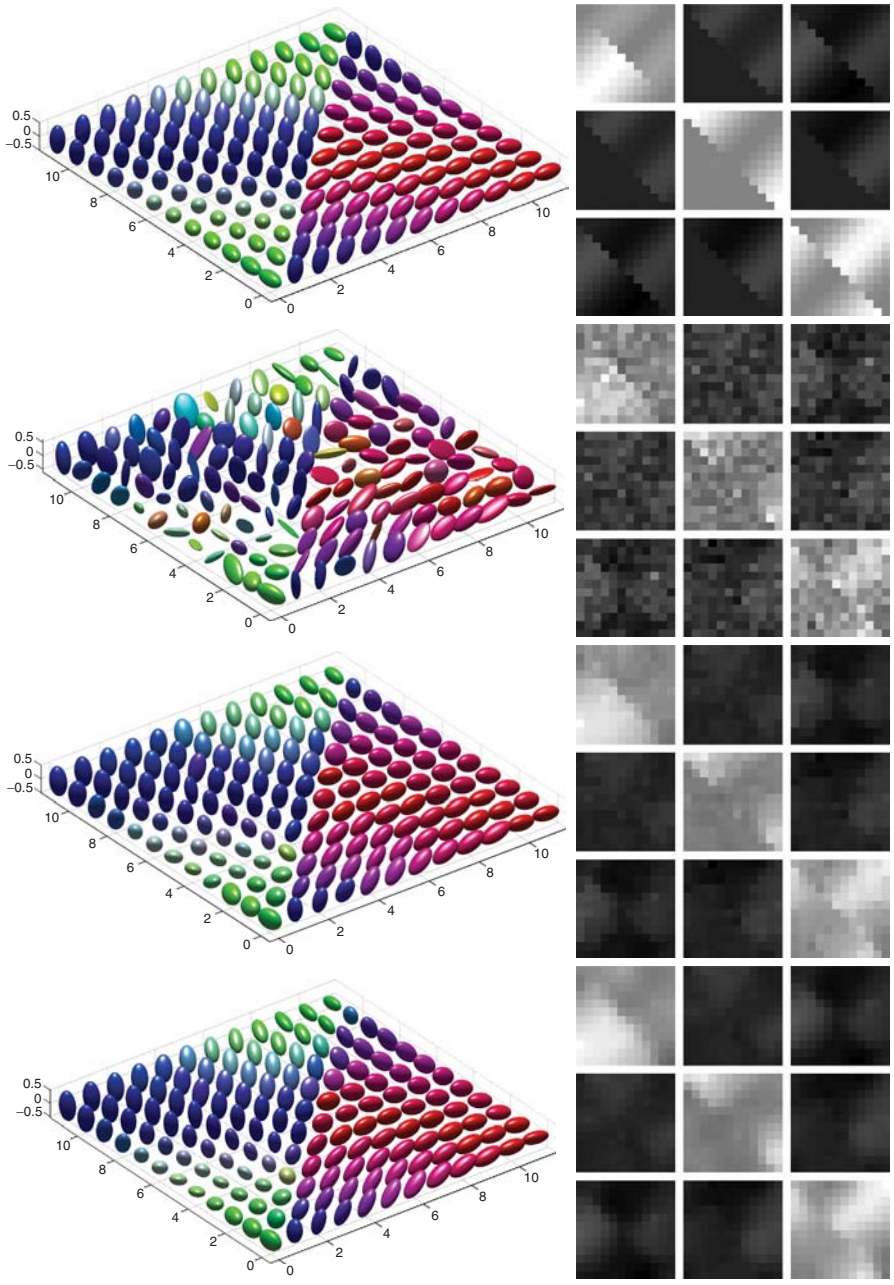


Fig. 4. Denoising of a $\mathrm{Sym}_3(\mathbb{R})$-valued image. *T*op to Bottom: Original image, noisy image, minimizer of the component-based method (19) for $\alpha = 0.11$, minimizer of the component-based infimal convolution approach (21) with parameters $\alpha_1 = 0.14$, $\alpha_2 = 0.08$. Visualization: ellipsoids (*left*), components of the matrix-valued data (*right*)

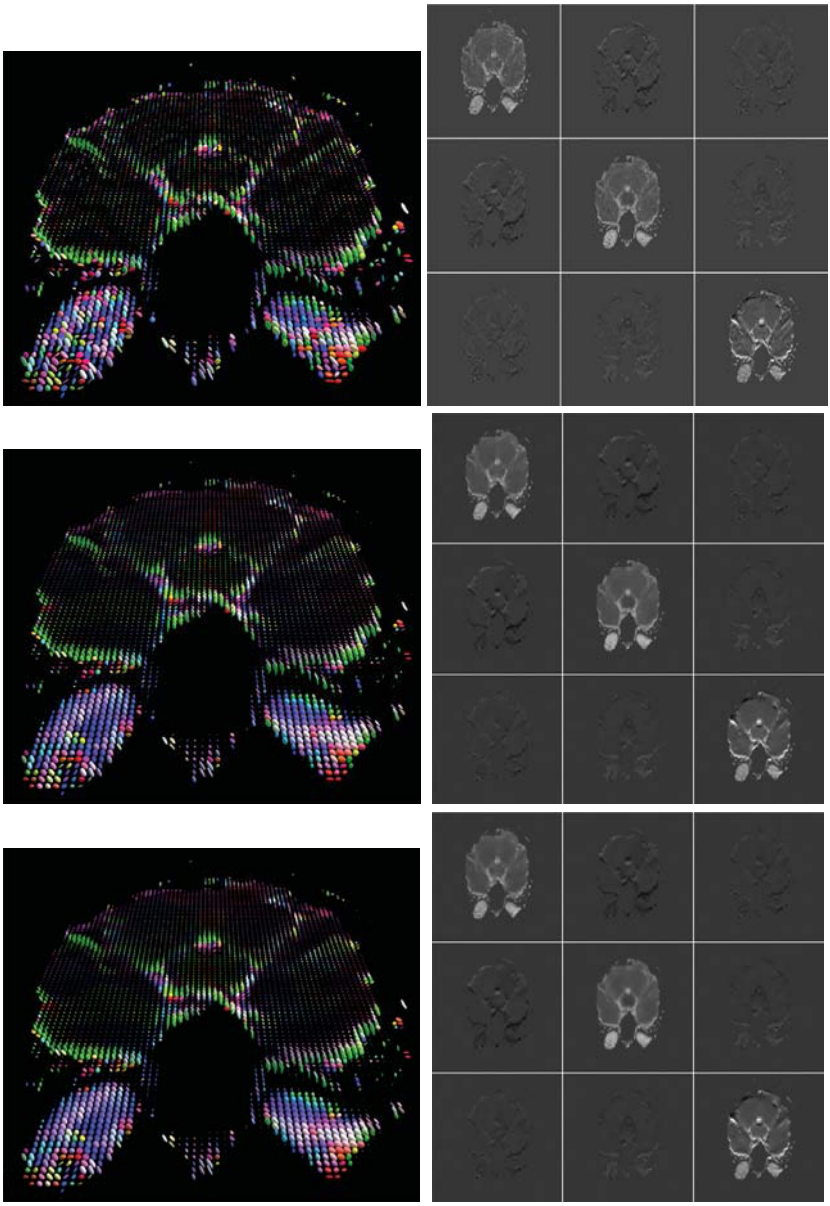


Fig. 5. Denoising of a real-world DT-MRI matrix field with values in $\mathrm{Sym}_3(\mathbb{R})$. *T*op: Original image. *M*iddle: Minimizer of the component-based method (19) for $\alpha = 600$. *B*ottom: Minimizer of the infimal convolution approach (21) for $\alpha_1 = 500$, $\alpha_2 = 600$

Consequently, $\mathcal{J}(P\hat{U}) \leq \mathcal{J}(\hat{U})$, which contradicts our assumption since the minimizer is unique. This completes the proof. □

Proof of Proposition 2. Since $\|F-U\|_F^2$ is strictly convex, it remains to show that the functional

$$J(U) := \operatorname{tr}\left(\sqrt{U_x^2+U_y^2}\right)$$

is convex. Moreover, since J is positively homogeneous, we only have to prove that J is subadditive, cf. [2, p. 34], i.e.

$$J(\tilde{U}+U) \leq J(\tilde{U}) + J(U).$$

This can be rewritten as

$$\operatorname{tr}\left(\sqrt{(\tilde{U}_x+U_x)^2+(\tilde{U}_y+U_y)^2}\right) \leq \operatorname{tr}\left(\sqrt{\tilde{U}_x^2+\tilde{U}_y^2}\right) + \operatorname{tr}\left(\sqrt{U_x^2+U_y^2}\right).$$

To prove this relation, we recall the definition of the *trace norm*, cf. [16, p. 197], which is defined as the sum of the singular values of a matrix $A \in \mathbb{R}^{s,t}$:

$$\|A\|_{\mathrm{tr}} = \operatorname{tr}(\sqrt{A^*A}).$$

Then we have for the symmetric matrices $\tilde{U}_x, \tilde{U}_y, U_x, U_y$ that

$$\| \begin{pmatrix} \tilde{U}_x+U_x \\ \tilde{U}_y+U_y \end{pmatrix} \|_{\mathrm{tr}} = \operatorname{tr}\left(\sqrt{(\tilde{U}_x+U_x)^2+(\tilde{U}_y+U_y)^2}\right).$$

Since $\|\cdot\|_{\mathrm{tr}}$ is a norm it follows that

$$\begin{aligned} \| \begin{pmatrix} \tilde{U}_x+U_x \\ \tilde{U}_y+U_y \end{pmatrix} \|_{\mathrm{tr}} &\leq \| \begin{pmatrix} \tilde{U}_x \\ \tilde{U}_y \end{pmatrix} \|_{\mathrm{tr}} + \| \begin{pmatrix} U_x \\ U_y \end{pmatrix} \|_{\mathrm{tr}} \\ &= \operatorname{tr}(\sqrt{\tilde{U}_x^2+\tilde{U}_y^2}) + \operatorname{tr}(\sqrt{U_x^2+U_y^2}) \end{aligned}$$

and we are done. □

Proof of Proposition 3. Let $\varphi(U_x,U_y) := \operatorname{tr}\left(\Phi(U_x^2+U_y^2)\right)$. The Euler–Lagrange equations of (22) are given, for $i,j=1,...,n$ and $i \geq j$, by

$$0 = \frac{\partial}{\partial u_{ij}}\|F-U\|_F^2 - \alpha\left(\frac{\partial}{\partial x}\left(\frac{\partial\varphi}{\partial u_{ijx}}\right) + \frac{\partial}{\partial y}\left(\frac{\partial\varphi}{\partial u_{ijy}}\right)\right).$$

For a scalar-valued function f and an $n \times n$ matrix X, we set $\frac{\partial f(X)}{\partial X} := \left(\frac{\partial f(X)}{\partial x_{ij}}\right)_{i,j=1}^{n}$. Then, by symmetry of F and U, the Euler–Lagrange equations can be rewritten in matrix–vector form as

$$W_n \circ \frac{U-F}{\alpha} = \frac{1}{2}\left(\frac{\partial}{\partial x}\left(\frac{\partial \varphi}{\partial U_x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial \varphi}{\partial U_y}\right)\right), \tag{27}$$

where W_n denotes the $n \times n$ matrix with diagonal entries 1 and other coefficients 2, and $A \circ B$ stands for the *Hadamard product* (componentwise product) of A and B.
We consider $f(X) := \operatorname{tr} \Phi(X^2)$. Then we obtain by [17, p. 178] and $\operatorname{tr}(A^{\mathrm{T}}B) = (\operatorname{vec} A)^{\mathrm{T}} \operatorname{vec} B$ that

$$\begin{aligned}\operatorname{vec} \frac{\partial f(X)}{\partial X} &= \operatorname{vec}\left(\operatorname{tr}\left(\Phi'(X^2)\,\frac{\partial (X^2)}{\partial x_{ij}}\right)\right)_{i,j=1}^{n} \\ &= \operatorname{vec}\left((\operatorname{vec} \Psi)^{\mathrm{T}} \operatorname{vec} \frac{\partial (X^2)}{\partial x_{ij}}\right)_{i,j=1}^{n}\end{aligned}$$

where $\Psi := \Phi'(X^2)$. By [17, p. 182] and since Ψ is symmetric this can be rewritten as

$$\operatorname{vec} \frac{\partial f(X)}{\partial X} = \operatorname{vec}\, W_n \circ ((I_n \otimes X) + (X \otimes I_n)) \operatorname{vec} \Psi.$$

Using that $\operatorname{vec}(ABC) = (C^{\mathrm{T}} \otimes A) \operatorname{vec} B$ we infer that

$$\operatorname{vec} \frac{\partial f(X)}{\partial X} = \operatorname{vec} W_n \circ \operatorname{vec}(X\Psi + \Psi X).$$

This implies that

$$\frac{\partial f(X)}{\partial X} = 2\, W_n \circ (\Psi \bullet X). \tag{28}$$

Applying (28) with $f(U_x) := \varphi(U_x, U_y)$ and $f(U_y) := \varphi(U_x, U_y)$, respectively, in (27) we obtain the assertion. □

Acknowledgements

The authors thank J. Weickert and S. Didas for their fruitful discussions.

References

[1] A. Barvinok. *A Course in Convexity, Graduate Studies in Mathematics*. AMS, Providence, RI, 2002.
[2] J. M. Borwein and A. S. Lewis. *Convex Analysis and Nonlinear Optimization*. Springer, New York, 2000.
[3] T. Brox, J. Weickert, B. Burgeth, and P. Mrázek. Nonlinear structure tensors. *Image and Vision Computing*, 24(1):41–55, 2006.

[4] B. Burgeth, A. Bruhn, S. Didas, J. Weickert, and M. Welk. Morphology for matrix-data: Ordering versus PDE-based approach. *Image and Vision Computing*, 25(4):496–511, 2007.

[5] B. Burgeth, A. Bruhn, N. Papenberg, M. Welk, and J. Weickert. Mathematical morphology for matrix fields induced by the Loewner ordering in higher dimensions. *Signal Processing*, 87(2):277–290, 2007.

[6] B. Burgeth, S. Didas, L. Florack, and J. Weickert. Singular PDEs for the processing of matrix-valued data. In F. Sgallari, A. Murli, and N. Paragios, editors, *Scale Space and Variational Methods in Computer Vision*, volume 4485 of *Lecture Notes in Computer Science*, pages 556–567. Springer, Berlin, 2007.

[7] B. Burgeth, S. Didas, and J. Weickert. A generic approach to diffusion filtering of matrix-fields. *Computing*, 81:179–197, 2007.

[8] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004.

[9] A. Chambolle and P. -L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997.

[9] C. Chefd'Hotel, D. Tschumperlé, R. Deriche, and O. Faugeras. Constrained flows of matrix-valued functions: Application to diffusion tensor regularization. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision – ECCV 2002*, volume 2350 of *Lecture Notes in Computer Science*, pages 251–265. Springer, Berlin, 2002.

[11] P. G. Ciarlet. *Introduction to Numerical Linear Algebra and Optimisation*. Cambridge University Press, Cambridge, 1989.

[12] S. Didas, S. Setzer, and G. Steidl. Combined ℓ_2 data and gradient fitting in conjunction with ℓ_1 regularization. *Advances in Computational Mathematics*, 30(1), 2008, (available online at www.springerlink.com).

[13] D. Goldfarb and W. Yin. Second-order cone programming methods for total variation-based image restoration. *SIAM J. Scientific Computing*, 2(27):622–645, 2005.

[14] Y. Gur and N. Sochen. Fast invariant Riemannian DT-MRI regularization. In *IEEE 11th International Conference on Computer Vision*, Rio de Janeiro, Brasil, 2007.

[16] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1990.

[16] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284(1):193–228, 1998.

[17] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. J. Wiley and Sons, Chichester, 1988.

[18] H. Mittelmann. An independent bechmarking of SDP and SOCP solvers. *Mathematical Programming Series B*, 95(2):407–430, 2003.

[19] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

[20] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[21] O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering. *Journal of Mathematical Imaging and Vision*, 12(1):43–63, 2000.

[22] G. Steidl, S. Setzer, B. Popilka, and B. Burgeth. Restoration of matrix fields by second-order cone programming. *Computing*, 81:161 – 178, 2007.

[23] D. Tschumperlé and R. Deriche. Diffusion tensor regularization with contraints preservation. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 948–953, Kauai, HI, 2001.

[24] Z. Wang, B. C. Vemuri, W. Chen, and T. H. Mareci. A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI. *IEEE Transactions on Medical Imaging*, 23(8):930–939, 2004.

[25] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998.

[30] J. Weickert and T. Brox. Diffusion and regularization of vector- and matrix-valued images. In M. Z. Nashed and O. Scherzer, editors, *Inverse Problems, Image Analysis, and Medical Imaging*, volume 313 of *Contemporary Mathematics*, pages 251–268. AMS, Providence, 2002.

[27] J. Weickert and H. Hagen, editors. *Visualization and Processing of Tensor Fields*. Springer, Berlin, 2006.

[28] Q. Xu, A. W. Anderson, J. C. Gore, and Z. Ding. Diffusion tensor image smoothing using efficient and effective anisotropic filtering. In *IEEE 11th International Conference on Computer Vision*, Rio de Janeiro, Brasil, 2007.

[29] M. Zéraïand M. Moakher. Riemannian curvature-driven flows for tensor-valued data. In F. Sgallari, A. Murli, and N. Paragios, editors, *Scale Space and Variational Methods in Computer Vision*, volume 4485 of *Lecture Notes in Computer Science*, pages 592–602, Springer, Berlin, 2007.

An Operator Algebraic Inverse Scale Space Method for Symmetric Matrix Valued Images

Johan Lie[1], Bernhard Burgeth[2], and Oddvar Christiansen[3]

[1] Centre for Integrated Petroleum Research Unifob/University of Bergen, Realfagsbyggat, Allegaten 41, 5007 Bergen, Norway
Johan.lie@cipr.uib.no
[2] Faculty of Mathematics and Computer Science, Saarland University, Bld. E1 1 66041 Saarbrücken, Germany
burgeth@mia.uni-saarland.de
[3] Department of Mathematics, University of Bergen, Johannes Brunsgate 12, 5008 Bergen, Oddvar
Christiansen@math.uib.no

Summary. During the last few years, a considerable amount of research has been conducted to study multiscale properties of images via partial differential equations. In this context, we can roughly divide the methodology into three different formulations, namely the scale space formulation, the regularization formulation, and the inverse scale space formulation. In this chapter, we propose an inverse scale space formulation for matrix valued images using the operator-algebraic approach recently introduced by Burgeth et al. in 2007 (B. Burgeth, S. Didas, L. Florack, and J. Weickert. A generic approach to diffusion filtering of matrix-fields. *Computing*, 81(2–3):179–197, 2007; B. Burgeth, S. Didas, L. Florack, and J. Weickert. A Generic Approach to the Filtering of Matrix Fields with Singular PDEs. In *Scale Space and Variational Methods in Computer Vision*. Volume 4485 of *Lecture Notes in Computer Science*, pages 556–567. Springer, Heidelberg, 2007). We perform numerical experiments on synthetic tensor fields and on real diffusion tensor data from DT-MRI of a human brain.

1 Introduction

In many applications there is a need for a multiscale representation of images. Partial differential equations (PDEs) provide a flexible framework for this purpose [8, 20, 31]. Traditionally, PDEs have been used via two intrinsically related approaches, namely parabolic equations of diffusive nature and elliptic equations corresponding to the minimum of energy functionals. In between these two approaches we find the inverse scale space approach, which we will address in this chapter. Inverse scale space methods have earlier been studied to some details in the context of scalar valued images [4, 17, 21, 24]. In this

chapter we extend the so-called *relaxed inverse scale space flow* to matrix valued images via the operator algebraic framework recently introduced [5, 6]. Inverse scale space methods as a concept is related to regularization methods for ill-posed inverse problems, i.e. Tikhonov regularization [14, 15, 27].

In this chapter, we consider a matrix field as a mapping $F : \Omega \subset \mathbb{R}^d \longrightarrow M_n(\mathbb{R})$, from a d-dimensional image domain into the set of $n \times n$-matrices with real entries, $F(x) = (f_{p,q}(x))_{p,q=1,\ldots,n}$. Essential for us is the subset of symmetric matrices $\mathrm{Sym}_n(\mathbb{R})$. The set of positive (semi-) definite matrices, denoted by $\mathrm{Sym}_n^{++}(\mathbb{R})$ (resp., $\mathrm{Sym}_n^{+}(\mathbb{R})$), consists of all symmetric matrices A with $\langle v, Av\rangle := v^\top Av > 0 \quad (\text{resp.}, \geq 0) \quad \text{for} \quad v \in \mathbb{R}^n \setminus \{0\}$.

Matrix fields play a vital role in many applications: In image processing as a form of the structure–tensor [12], in continuum mechanics where tensors are widely used to describe anisotropic behaviour, such as strain–stress or permittivity tensors. But most importantly in this setting medical imaging. Diffusion tensor magnetic resonance imaging (DT-MRI) [1] constitutes a modern and widely used image acquisition technique that measures a 3×3 matrix field with positive definite matrices. To each voxel, a so-called diffusion tensor is assigned describing diffusive properties of water molecules; thus it is intimately related to the geometry and organization of the tissue under examination. Hence the matrix field obtained is a valuable source of in vivo information about the underlying tissue structure, e.g. in the human brain [2, 22]. This information can, for example, be used for the construction of maps of the tissue-connectivity [19, 33, 34], or for construction of anisotropy measures like the fractional anisotropy (FA), the relative anisotropy (RA) and the barycentric index [32]. The fractional anisotropy is a measure which is routinely used by the medical practitioners.

The indicated variety of applications require the development of appropriate tools for the processing and analysis of matrix-valued data. Comprehensive survey articles on the analysis of DT-MRI matrix fields using various techniques can be found in [31]. The work here concentrates on the multistage representation of matrix fields via the inverse scale space methodology.

In the context of scalar images, we can write a standard parabolic flow as

$$\begin{aligned} u_t &= -p(u), \\ u(x,0) &= f, \end{aligned} \tag{1}$$

and the corresponding elliptic flow as

$$\begin{aligned} u_t &= -p(u) + \lambda(f-u), \\ u(x,0) &= f, \end{aligned} \tag{2}$$

with appropriate boundary conditions and f is the initial data. The symbol $p(u)$ denotes a differential operator involving u, usually stemming from a regularization functional. In the literature

$$p(u) = -\Delta u$$

and

$$p(u) = -\text{div}\left(\frac{\nabla u}{|\nabla u|}\right)$$

are used as canonical linear and nonlinear operators respectively. In applications, often the nonlinear operators are preferable, since linear operators tend to over-smooth the edges in the images [8, 20, 23].

As indicated by the names of the two flows, the parabolic PDE (1) develops the initial condition towards the mean value of u as time increases, while the elliptic PDE (2) governs an evolution from the initial condition towards a non-trivial steady-state. In order to make the two flows meaningful, a parameter estimation must be done. In the context of the parabolic flow, a stopping condition must be imposed and in the context of the elliptic flow, the weighting factor λ must be chosen. Finding the optimal parameter λ^* or t^* is in general difficult. However, approximations are usually easier to obtain.

By a modification of the flow (2), we get the relaxed inverse scale space flow

$$\begin{aligned} u_t &= -p(u) + \lambda(f - u + v), \\ v_t &= \alpha(f - u), \\ u(x,0) &= \text{mean}(f), \\ v(x,0) &= 0, \end{aligned} \tag{3}$$

with $\alpha \leq \frac{\lambda}{5}$ as a relaxation parameter. This relaxed inverse scale space flow was first introduced by Gilboa et al., and has been studied by several authors [4, 17]. In contrast to the standard scale space formulation, here interpreted as the flow (1) with the righthand side $p(u) = -\Delta u$, the relaxed inverse flow produces a group of images evolving from a very smooth initial condition mean(f) towards the data f. By the introduction of the inverse scale space, it is possible to traverse the scale space in both directions. It is even possible to iterate back and forth in the scale space. This could possibly be used to make more robust and effective scale spaces. For a more detailed explanation of the inverse scale space methods, see eg. [4, 17].

Attempts to extend the PDE methodology used for scalar images to the setting of matrix fields with positive (semi-) definite matrices have been made [7, 10, 11, 25, 28–30]. Matrix field regularization as suggested in [9, 13] is based on differential geometric considerations. Recently, Christiansen et al. generalized the vector TV model of Blomgren and Chan in a straightforward manner to yield a regularization method for matrix valued images [3, 10]. This approach is interesting in the sense that during the flow, the diffusion tensor is by construction positive definite, which is usually required in DT-MRI applications.

A different and more general approach was presented by Burgeth et al. in 2007 [5, 6]. There they introduced an operator-algebraic approach for the construction of matrix-valued PDEs for matrix fields. They show that it is possible to transfer the characteristic behaviour of scalar operators to matrix operators as well. Important scalar models like motion by mean curvature, self-snakes,

the Perona–Malik model and the ROF model are generalized to matrix valued models. By different choices of matrix products, they preserve intrinsic matrix properties like, for example, positive definiteness. In [5, 6] various instances of the parabolic equation (1) are studied within this context. The setup provides a coupling of the different matrix channels, treating diagonal and off-diagonal elements in a proper manner. This appropriate channel interaction is a key issue in the processing of matrix fields. As a proof-of-concept, we demonstrate in this chapter that the same operator-algebraic framework can be successfully used to obtain a relaxed inverse scale space flow (3) for matrix fields.

This chapter is organized as follows: The next section is devoted to the basic operations of the operator algebraic framework and we show how it can be used for the purpose of generalizing the inverse scale space methods to a matrix-valued setting. We discuss numerical implementation issues in Sect. 4. Section 4 also contains a validation of our matrix-valued inverse scale space methodology by a few numerical computations on synthetic diffusion tensor fields and real DT-MRI fields. Summarizing remarks in Sect. 5 concludes the chapter.

2 Inverse Methods for Matrix Valued Images

In this section, we give a sufficiently detailed description of how we generalize the relaxed inverse scale space method from a scalar setting to a matrix setting. Although the operator-algebraic framework has been described by Burgeth et al. [5, 6] we repeat its necessary parts here for the sake of completeness.

The definition of a function of a symmetric matrix is well established in linear algebra [16]:

$$g(U) := V^\top \mathrm{diag}(g(\lambda_1), \ldots, g(\lambda_n))V,$$

where $U := V^\top \mathrm{diag}(\lambda_1, \ldots, \lambda_n)V$ is the spectral decomposition of U and g is a real function applicable to the set of real eigenvalues of U. The usual product of two symmetric matrices is in general no longer symmetric unless they commute. Hence a suitable 'symmetric' multiplication has to be found. Among the numerous possibilities, we mention two choices. The so-called Jordan product is given by

$$A \bullet_J B := \frac{1}{2}(AB + BA).$$

The preconditioner product, named after its usage in numerical linear algebra, as a way to precondition symmetric linear systems, is defined as

$$A \bullet_P B := \sqrt{A}B\sqrt{A}\,.$$

with a positive semidefinite A.

In the numerical experiments in this chapter we use the Jordan product, from now on simply denoted by "•".

The rather obvious matrix-valued version of a difference quotient for a matrix field U gives rise to a component-wise definition of **spatial or temporal partial derivatives** [5, 6]

$$\overline{\partial}_\omega U(\omega_0) = \lim_{h\to 0} \frac{U(\omega_0+h)-U(\omega_0)}{h} = \left(\partial_\omega u_{ij}(\omega_0)\right)_{ij}. \tag{4}$$

It is close at hand that other linear operators like the arithmetic mean or convolution with a kernel are also interpreted component-wise in the matrix setting.

The set of spatial partial derivatives forms the spatial **gradient** for matrix fields $\overline{\nabla} U \in (\mathrm{Sym}_n(\mathbb{R}))^d$

$$\overline{\nabla} U(x) = \left[\overline{\partial}_{x_1} U(x), \overline{\partial}_{x_2} U(x), \cdots, \overline{\partial}_{x_d} U(x)\right]^T. \tag{5}$$

Note that this gradient is an element of the module $(\mathrm{Sym}_n(\mathbb{R}))^d$ rather than a higher order tensor as one might expect from a differential geometric point of view. We refer to a function with values in $(\mathrm{Sym}_n(\mathbb{R}))^d$ as a module field. Similarly, the generalized **divergence** defined by

$$\overline{\mathrm{div}}\ (A(x))^\top := \sum_{i=1}^{d} \overline{\partial}_{x_i} A_i(x)$$

for a module field $A(x) := \big(A_1(x), \ldots, A_d(x)\big)$ is again not a higher order tensor but simply a symmetric matrix. The definition of a **Laplacian** for a matrix field U is straightforward:

$$\overline{\Delta} U := \sum_{i=1}^{d} \overline{\partial}_{x_i}^2 U$$

Finally, there is a notion of p-**length** in the module $(\mathrm{Sym}_n(\mathbb{R}))^d$ realized for a 'fat vector' $W = (W_1, \ldots, W_d) \in (\mathrm{Sym}_n(\mathbb{R}))^d$ by

$$|W|_p := \sqrt[p]{|W_1|^p + \cdots + |W_d|^p} \in \mathrm{Sym}_n^+(\mathbb{R}).$$

The most essential term in equations (1), (2), and (3) is $p(u)$. In its most basic (linear) form, we have

$$p(u) = -\sum_{i=1}^{d} \partial_{x_i}\partial_{x_i} u = -\Delta u\,. \tag{6}$$

Using the operator-algebraic framework, $p(u)$ can be readily generalized to an operator acting on a matrix by

$$P_{\Delta}(U) = -\overline{\Delta} U. \tag{7}$$

However, the linear model is not very interesting for most practical purposes. The canonical nonlinear model where the Laplacian is replaced with the nonlinear term

$$p(u) = -\mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right) \tag{8}$$

can similarly be generalized to a matrix valued setting by

$$P_{TV}(U) = -\overline{\mathrm{div}}\left(\frac{1}{|\overline{\nabla} U|} \bullet \overline{\nabla} U\right), \tag{9}$$

where $|\overline{\nabla} U|$ stands for the 2-length of $\overline{\nabla} U$, that is, $|\overline{\nabla} U| := |\overline{\nabla} U|_2 = \sqrt{\sum_{i=1}^{d} |U_i|^2}$ and "$\bullet$" denotes either the Jordan or the preconditioning product. The diffusivity can be interpreted as the inverse of $|\overline{\nabla} U|$:

$$G = \frac{1}{|\overline{\nabla} U|} = |\overline{\nabla} U|^{-1} \tag{10}$$

since we have

$$|\overline{\nabla} U|^{-1} \bullet_J |\overline{\nabla} U| = |\overline{\nabla} U|^{-1} \bullet_P |\overline{\nabla} U| = |\overline{\nabla} U|^{-1} \cdot |\overline{\nabla} U| = I\,. \tag{11}$$

Having defined the regularization operator $P(U)$ (where P equals, e.g. P_{Δ} or P_{TV}) for a matrix valued image U, we rewrite the relaxed inverse scale space flow to matrix valued images by

$$\begin{aligned} U_t &= -P(U) + \lambda(F - U + V), \\ V_t &= \alpha(F - U), \\ U(x,0) &= \mathrm{simp}(F) \\ V(x,0) &= 0, \end{aligned} \tag{12}$$

with $\alpha \leq \lambda/5$ and U, F and V being matrix fields mapping $\mathbf{R}$ to $\mathrm{Sym}_n(\mathbb{R})$. Temporal derivatives are denoted for brevity as U_t and V_t. The expression $\mathrm{simp}(F)$ stands for a simplified version of F. In the simplest case, $\mathrm{simp}(F)$ might be the arithmetic mean of the matrix field F,

$$\mathrm{simp}(F) = \int_{\Omega} F(x)\; d\,x$$

or alternatively the solution $(x,t) \mapsto W(x,t)$ of another evolution process such as linear, Perona–Malik- or TV- diffusion. The initialization flow can, for example, start with initial value F and stop at a certain time τ:

$$\mathrm{simp}(F) = W(\cdot, \tau)\,.$$

From the scalar system (3), we know that with increasing time u flows from $\mathrm{simp}(f)$ towards f. Analogously, in the matrix valued setting, the flow (12) evolves from the simplified field $\mathrm{simp}(F)$ towards the matrix field F.

3 Numerical Implementation

In this section we include details on the numerical implementation of the method. We discretize the equation system (12) by standard finite differences. We employ the same numerical scheme as in [5]. Super- and sub-indexes denote temporal and spatial discretization respectively. For the temporal derivative we use an Euler explicit scheme, i.e.

$$\frac{\partial U}{\partial t} \approx \frac{U^{n+1}_{i,j,k} - U^{n}_{i,j,k}}{\Delta t}. \tag{13}$$

To simplify the notation, we sometimes omit one of the indexes. Thus we denote $U^n_{i,j,k}$ as $U_{i,j,k}$, or $U^n_{i,j,k}$ as U^n. The arithmetic mean of the diffusivity in the direction determined by the index is defined as

$$G_{i\pm\frac{1}{2},j,k} = \frac{G_{i\pm1,j,k} + G_{i,j,k}}{2}. \tag{14}$$

As a numerical approximation to P, we employ $\tilde{P}$ given by

$$\begin{aligned}\tilde{P}(U) = &\frac{1}{\Delta x}\Big(G_{i+\frac{1}{2},j,k} \bullet \frac{U_{i+1,j,k} - U_{i,j,k}}{\Delta x} - G_{i-\frac{1}{2},j,k} \bullet \frac{U_{i,j,k} - U_{i-1,j,k}}{\Delta x}\Big),\\ &+ \frac{1}{\Delta y}\Big(G_{i,j+\frac{1}{2},k} \bullet \frac{U_{i,j+1,k} - U_{i,j,k}}{\Delta y} - G_{i,j-\frac{1}{2},k} \bullet \frac{U_{i,j,k} - U_{i,j-1,k}}{\Delta y}\Big),\\ &+ \frac{1}{\Delta z}\Big(G_{i,j,k+\frac{1}{2}} \bullet \frac{U_{i,j,k+1} - U_{i,j,k}}{\Delta z} - G_{i,j,k-\frac{1}{2}} \bullet \frac{U_{i,j,k} - U_{i,j,k-1}}{\Delta z}\Big).\end{aligned}$$

This leads to the following numerical schemes for the equation system (12)

$$\begin{aligned}U^{n+1} &= U^n - \Delta t\Big(\tilde{P}(U^n) + \lambda(F - U^n + V^n)\Big),\\ V^{n+1} &= V^n - \alpha\Delta t(F - U^n),\\ U^0 &= \mathrm{simp}(F),\\ V^0 &= 0.\end{aligned} \tag{15}$$

We will use the numerical schemes introduced above to perform numerical experiments on matrix fields in the next section.

4 Numerical Experiments

Numerical experiments on artificial and real DT-MRI data will confirm that the inverse scale space concept can be transferred to the matrix fields via the operator-algebraic framework. For all the numerical experiments we initialize the algorithm with an over-smoothed version of the noisy matrix field F, $U(x,0) = \mathrm{simp}(F)$, as described in Sect. 2. From the simplified matrix

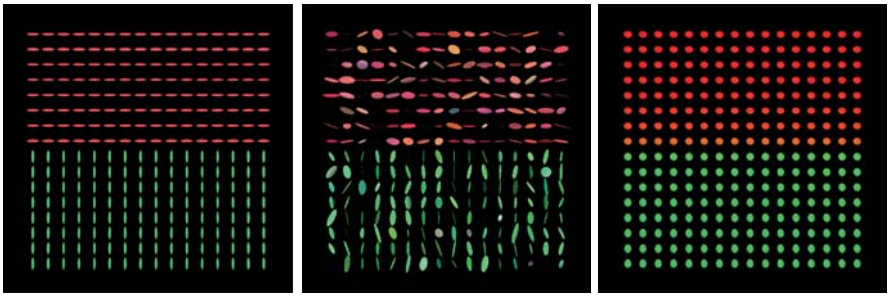
field $U(x,0)$, the matrix field $U(x,t)$ evolves towards the noisy matrix field F. All the processed matrix fields are 3D matrix fields. We display only one 2D-slice of the data set. All experiments are carried out with the parameters $\lambda = 2$ and $\alpha = \lambda/5$. This means that we only have one active parameter in the methodology, namely the stopping time t. This is similar to the scale space approach where the stopping time is a free parameter, and the regularization approach where the regularization parameter is a free parameter. In this chapter, we study the behaviour of the relaxed inverse scale space flow, but we do not search for a good stopping criteria. This is common practice for this kind of methods.

All computations are done on a computer with a 2 Opteron 270 Dual-core processor and 8 GB of memory. The implementation is done in MatLab® with some parts implemented in C using mex files [26].

4.1 A Simple Synthetic Field

In the first numerical experiment, we apply the proposed methodology on a simple three-dimensional tensor field with spatial dimensions $16 \times 16 \times 5$. All five layers in this data set are equal, and Fig. 1(a) shows one slice of the tensor field. The tensor field consists of two regions, one with principal diffusion direction along the east–west and one along the north–south direction. Each matrix in the regions have one single large eigenvalue of 1, and two small eigenvalues of 0.25. Thus the regions are anisotropic.

The clean tensor field has matrix elements in the range from 0 to 1, and we add normal distributed noise with zero mean and variance of 0.2, see Fig. 1(b). This tensor field is then evolved, using a simple forward model, into an over-smoothed tensor field, see Fig. 1(c). The noisy tensor field and the over-smoothed tensor field is then respectively used as the input parameters F and $\mathrm{simp}(F)$ in (12).



(a) (b) (c)

Fig. 1. (**a**) A simple synthetic field with two regions. (**b**) Normal distributed noise with zero mean and variance of 0.2 is added to the clean tensor field. (**c**). The noisy tensor field is evolved, using a simple forward model, into an over-smoothed tensor field

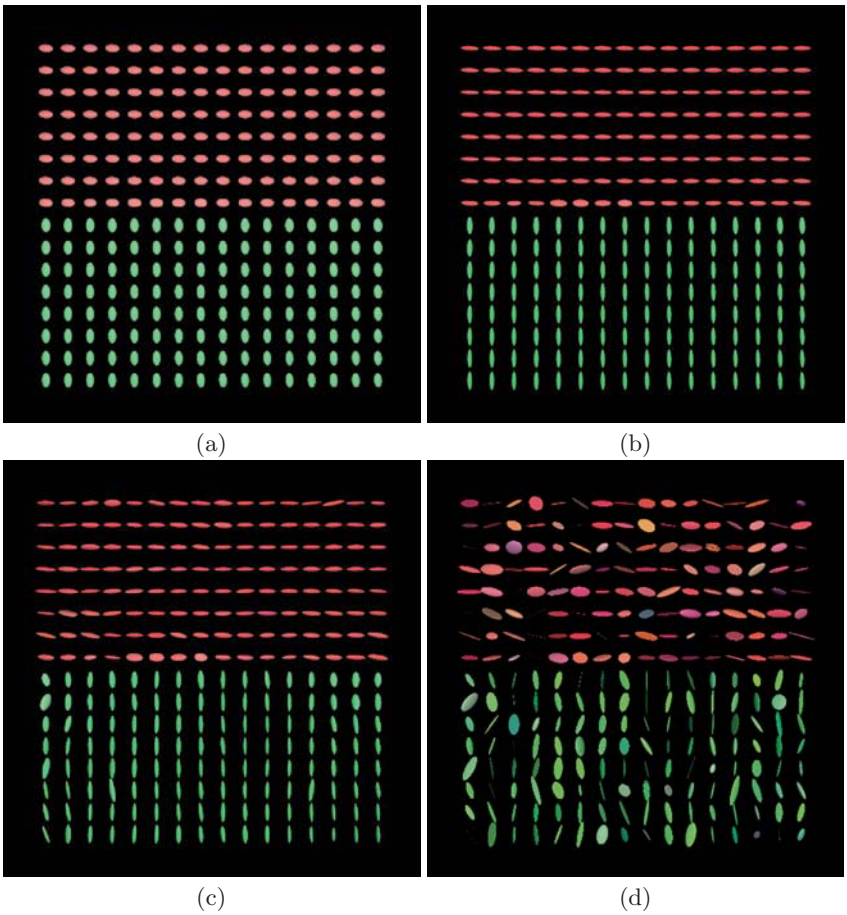
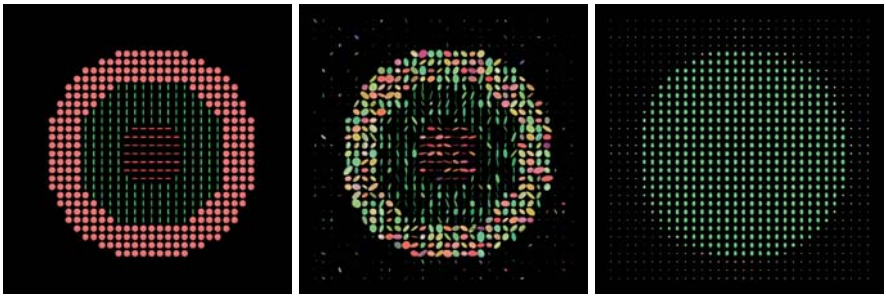


Fig. 2. The relaxed inverse scale space flow in a matrix setting. (**a**) After 300 iterations the over-smoothed tensor field has started to align along the directions in the true solution. (**b**) After 1,950 iterations we are close to the true solution. (**c**) After 5,000 iterations noise is beginning to appear in the matrix field. (**d**) After 20,000 iterations U is close to the noisy data F

In Fig. 2, we observe that more and more information is added to U as time progresses. After a while, the field is almost identical to the noise-free field and as anticipated, a while after this, the noise reappears in the field U. Eventually, $U \to F$ as $t \to \infty$, which also is to be expected.

4.2 A More Complex Synthetic Example

In the next example, we have a more complex synthetic three-dimensional tensor field with spatial dimensions $32 \times 32 \times 5$. As in the previous example, all five layers are equal, and Fig. 3(a) shows one slice of the tensor field. The



(a) (b) (c)

Fig. 3. (**a**) A more complex synthetic field with four regions. (**b**) Normal distributed noise with zero mean and variance of 0.2 is added to the clean tensor field. (**c**). The noisy tensor field is evolved, using a simple forward model, into an over-smoothed tensor field

tensor field consists of four regions, the outermost region is empty with each matrix equal to the null matrix. The second outermost region is circle shaped. Each matrix in this region has two large eigenvalues of 1 and a single small one of 0.25. The two innermost regions are more anisotropic. Each matrix in these regions have one single large eigenvalue of 1 and two smaller ones of 0.25.

The clean tensor field has matrix elements in the range from 0 to 1, and we add normal distributed noise with zero mean and variance of 0.2, see Fig. 3(b). This tensor field is then evolved, using a simple forward model (ROF model with $\lambda = 0$), into an over-smoothed tensor field, see Fig. 3(c). The noisy tensor field and the over-smoothed tensor field are then used as the input parameters F and $\mathrm{simp}(F)$ respectively in (12).

As in the previous example, we observe that more and more information is added to U as time increases, see Fig. 4. As anticipated the evolution from a degraded image over a somewhat denoised version to a variant close to the polluted version is observed.

By these experiments, we have demonstrated that, at least qualitatively, the relaxed inverse scale space flow in the matrix setting indeed behaves as expected from the scalar setting. The diffusion tensors field evolves from an essentially isotropic tensor field at $t = 0$, through a state of a structured anisotropic tensor field, and finally towards a noisy unstructured anisotropic tensor field.

4.3 Real Brain DTI

In the last numerical experiment, we apply the proposed methodology on a real diffusion tensor MRI tensor field. To be able to display a slice of the complex human brain tensor field as a whole, we use a colour-coded FA plot instead of the ellipsoid representation used in the previous experiments, see Fig. 5(a). However, the calculations are performed, as before, on the underlying matrix field.

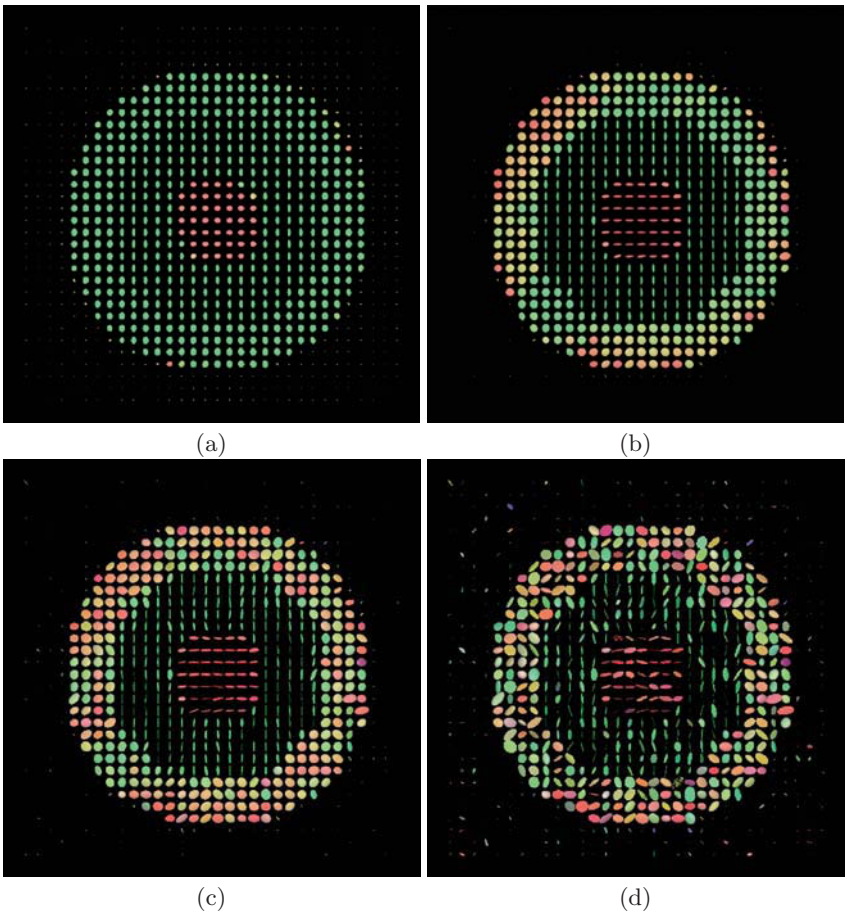
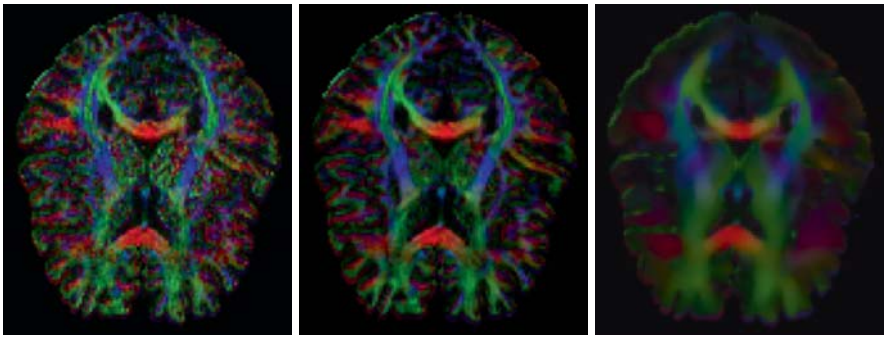


Fig. 4. The relaxed inverse scale space flow in a matrix setting. (**a**) After 500 iterations the over-smoothed tensor field has started to align along the directions in the true solution. (**b**) After 2,430 iterations we are close to the true solution. (**c**) After 5,000 iterations noise is beginning to appear in the matrix field. (**d**) After 20,000 iterations U is close to the noisy data F

The data used is a full 3D matrix valued data set with spatial dimensions $110 \times 126 \times 65$. The human subject data were acquired using a 3.0 T scanner (Magnetom Trio, Siemens Medical Solutions, Erlangen, Germany) with a 8-element head coil array and a gradient subsystem with the maximum gradient strength of $40\,\mathrm{mT\,m^{-1}}$ and maximum slew rate of 200 $\mathrm{mT\,m^{-1}\,ms^{-1}}$. The DTI data were based on spin-echo single shot EPI acquired utilizing generalized auto calibrating partially parallel acquisitions (GRAPPA) technique with acceleration factor of 2, and 64 reference lines. The DTI acquisition consisted of one baseline EPI, S_0, and six diffusion weighted images $S_1, \ldots, S_6$



(a) (b) (c)

Fig. 5. Colour coded FA plots of one slice of real DTI (**a**) The noisy input image. (**b**) High quality reference solution where 18 measurements are registered and averaged. (**c**) The noisy tensor field evolved into an over-smoothed tensor field

(b-factor of 1,000 $\mathrm{s\,mm^{-2}}$) along the same gradient directions as in the previous example. Each acquisition had the following parameters: TE / TR / averages was 91 ms/10,000 ms/2, FOV was 256×256 $\mathrm{mm^2}$, slice thickness/gap was 2 mm/0 mm, acquisition matrix was 192×192 pixels and partial Fourier encoding was 75%.

Since we are working with real data, we do not have access to an exact "true" solution. Instead we used a high quality reference dataset for comparison, see Fig. 5(b). This dataset was obtained by registering and averaging 18 acquisitions. The noisy dataset used as F in (12) is a four averaged acquisitions consuming about 20% of the acquisition time, compared to the higher-quality one. This dataset is then evolved into an over-smoothed dataset and used as the initial parameter $\mathrm{simp}(F)$, see Fig. 5(c).

From the experiment, see Fig. 6, we clearly infer that the inverse scale space methodology is well suited for the construction of multiscale representations of DTMRI fields of the human brain. The visualizations are made by the software DTIStudio, which is developed by Mori et al. [18].

5 Conclusions

In this chapter, we have employed the operator-algebraic framework of Burgeth et al. to make a straightforward generalization of the so-called relaxed inverse scale space flow to matrix images. We have performed numerical experiments, which indicate that the matrix flow qualitatively resembles the scalar flow. Further research must be performed to justify the usefullness of the methodology. Also, further work must be done to construct good stopping criteria for the inverse scale space flow.

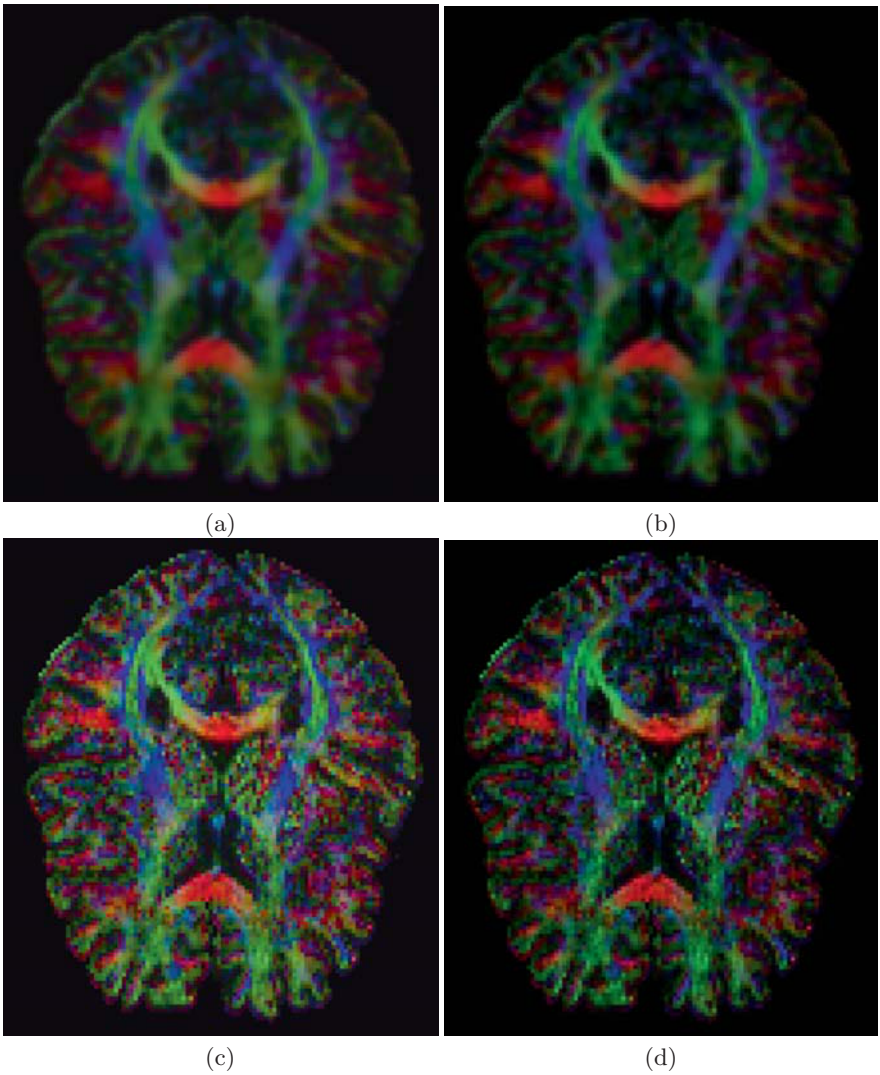


Fig. 6. One slice of a real diffusion tensor field of a human brain, colour-coded FA plot. (**a**) After 50 iterations we are approaching a good result. (**b**) After 110 iterations we are close to the optimal matrix field. (**c**) After 250 iterations noise is beginning to appear in the matrix field. (**d**) After 1,000 iterations U is close to the noisy data F

Acknowledgements

We want to thank Stephan Didas from the Mathematical Imaging and Analysis group at Saarland University for providing the synthetic data for this project, and for kindly helping us with the visualization of the tensor fields.

The visualization of the synthetic data was done by an in-house visualization software from the Mathematical Imaging and Analysis group at Saarland University. We also thank Dr. Siamak Ardekani at UCLA for kindly providing the real DTMRI data used in this project.

References

[1] P. J. Basser, J. Mattiello, and D. LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267, 1994.

[2] P.J Basser, J. Mattiello, and D. LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66(1):259–267, 1994.

[3] P. Blomgren and T. F Chan. Color TV: Total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3):304–309, 1998.

[4] M. Burger, G. Gilboa, S. Osher, and J. Xu. Nonlinear inverse scale space methods. *Communications in Mathematical Sciences*, 4(1):179–212, 2006.

[5] B. Burgeth, S. Didas, L. Florack, and J. Weickert. A generic approach to diffusion filtering of matrix-fields. *Computing*, 81(2–3):179–197, 2007.

[6] B. Burgeth, S. Didas, L. Florack, and J. Weickert. A Generic Approach to the Filtering of Matrix Fields with Singular PDEs. In *Scale Space and Variational Methods in Computer Vision*. volume 4485 of *Lecture Notes in Computer Science*, pages 556–567. Springer, Heidelberg, 2007.

[7] T. Brox, J. Weickert, B. Burgeth, and P. Mrázek. Nonlinear structure tensors. *Image and Vision Computing*, 24(1):41–55, 2006.

[8] T. F. Chan and J. Shen. *Image Processing And Analysis: Variational, PDE, Wavelet, And Stochastic Methods*. SIAM, Philadelphia, 2005.

[9] C. Chefd'Hotel, D. Tschumperlé, R. Deriche, and O. Faugeras. Constrained flows of matrix-valued functions: Application to diffusion tensor regularization. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision – ECCV 2002*, volume 2350 of *Lecture Notes in Computer Science*, pages 251–265. Springer, Berlin, 2002.

[10] O. Christiansen, T.-M. Lee, J. Lie, U. Sinha, and T. F. Chan. Total variation regularization of matrix valued images. *International Journal of Biomedical Imaging, special issue on Mathematics in Biomedical Imaging*, 2007. DOI: 10.1155/2007/27342.

[11] C. Feddern, J. Weickert, B. Burgeth, and M. Welk. Curvature-driven PDE methods for matrix-valued images. *International Journal of Computer Vision*, 69(1):91–103, 2006.

[12] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken, Switzerland, June 1987.

[13] Y. Gur and N. Sochen. Diffusion over tensor fields via Lie group PDE flows: Lagrangian action approach. In *Topology and robotics*, volume 438

of *Contemporary Mathematics*, pages 59–74. American Mathematical Society, Providence, RI, 2007.

[14] P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Number 4 in SIAM Monographs on Mathematical Modeling and Computation. SIAM, Philadelphia, 1997.

[15] P. C. Hansen. Deconvolution and regularization with toepliz matrices. *Numerical Algorithms*, 29:323–378, 2002.

[16] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1990.

[17] J. Lie and J. M. Nordbotten. Inverse scale spaces for nonlinear regularization. *Journal of Mathematical Imaging and Vision*, 27(1):41–50, 2007.

[18] S. Mori et al. Dti-studio. Available from: https://www.mristudio.org/.

[19] S. Mori and P. C. M. van Zijl. Fiber tracking: Principles and Strategies-a Techical Review. *NMR Biomed*, 15:468–480, 2002.

[20] S. Osher and R. Fedkiw. *Level set methods and dynamic implicit surfaces*, volume 153 of *Applied Mathematical Sciences*. Springer, New York, 2003.

[21] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling and Simulation*, 4(2):460–489, 2005.

[22] C. Pierpaoli, P. Jezzard, P. J. Basser, A. Barnett, and G. Di Chiro. Diffusion tensor MR imaging of the human brain. *Radiology*, 201(3):637–648, 1996.

[23] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithm. *Physica D*, 60:259–268, 1992.

[24] O. Scherzer and C. Groetsch. Inverse scale space theory for inverse problems. In M. Kerckhove, editor, *Scale-Space and Morphology in Computer Vision : Third International Conference, Scale-Space 2001*, volume 2106 of *Lecture Notes in Computer Science*. Springer, Berlin, 2001.

[25] T. Schultz, B. Burgeth, and J. Weickert. Flexible segmentation and smoothing of DT-MRI fields through a customizable structure tensor. In *Proceedings of the International Symposium on Visual Computing*, Lecture Notes in Computer Science. Springer, Berlin, 2007.

[26] The Mathworks. MatLab®, The Language of Technical Computing.

[27] A. Tikhonov and V. A. Arsenin. *Solutions of ill-posed problems*. VH Winston, New York, 1977.

[28] D. Tschumperlé and R. Deriche. Diffusion tensor regularization with constraints preservation. In *Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 948–953, Kauai, HI, December 2001.

[29] Z. Wang, B. C. Vemuri, Y. Chen, and T. Mareci. A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex dwi. *IEEE Transactions on Medical Imaging*, 23(8):930–939, 2004.

[30] J. Weickert and T. Brox. Diffusion and regularization of vector- and matrix-valued images. In M. Z. Nashed and O. Scherzer, editors, *Inverse*

Problems, Image Analysis, and Medical Imaging, volume 313 of *Contemporary Mathematics*, pages 251–268. AMS, Providence, 2002.

[31] J. Weickert and H. Hagen (Eds.). *Visualization and Processing of Tensor Fields*. Mathematics and Visualization. Springer, Berlin, 2005.

[32] C.-F Westin, S. E. Maier, H. Mamata, A. Nabavi, F. A. Jolesz, and R. Kikinis. Processing and visualization for diffusion tensor mri. *Medical Image Analysis*, 6:93–108, 2002.

[33] L. Zhukov and A. H. Barr. Oriented tensor reconstruction: Tracing neural pathways from diffusion tensor mri. *IEEE Visualization*, 387–394, 2002.

[34] L. Zhukov and A. H. Barr. Heart-muscle fiber reconstruction from diffusion tensor MRI. *IEEE Visualization 2003, Proceedings of Vis 2003*, 597–602, 2003.