51

Multivariate Approximation Theory: Selected Topics

E. W. CHENEY
University of Texas at Austin

CBMS-NSF
REGIONAL CONFERENCE SERIES
IN APPLIED MATHEMATICS

SPONSORED BY
CONFERENCE BOARD OF
THE MATHEMATICAL SCIENCES

SUPPORTED BY
NATIONAL SCIENCE
FOUNDATION

CBMS-NSF REGIONAL CONFERENCE SERIES IN APPLIED MATHEMATICS

A series of lectures on topics of current research interest in applied mathematics under the direction of the Conference Board of the Mathematical Sciences, supported by the National Science Foundation and published by SIAM.

GARRETT BIRKHOFF, *The Numerical Solution of Elliptic Equations*
D. V. LINDLEY, *Bayesian Statistics, A Review*
R. S. VARGA, *Functional Analysis and Approximation Theory in Numerical Analysis*
R. R. BAHADUR, *Some Limit Theorems in Statistics*
PATRICK BILLINGSLEY, *Weak Convergence of Measures: Applications in Probability*
J. L. LIONS, *Some Aspects of the Optimal Control of Distributed Parameter Systems*
ROGER PENROSE, *Techniques of Differential Topology in Relativity*
HERMAN CHERNOFF, *Sequential Analysis and Optimal Design*
J. DURBIN, *Distribution Theory for Tests Based on the Sample Distribution Function*
SOL I. RUBINOW, *Mathematical Problems in the Biological Sciences*
P. D. LAX, *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*
I. J. SCHOENBERG, *Cardinal Spline Interpolation*
IVAN SINGER, *The Theory of Best Approximation and Functional Analysis*
WERNER C. RHEINBOLDT, *Methods of Solving Systems of Nonlinear Equations*
HANS F. WEINBERGER, *Variational Methods for Eigenvalue Approximation*
R. TYRRELL ROCKAFELLAR, *Conjugate Duality and Optimization*
SIR JAMES LIGHTHILL, *Mathematical Biofluiddynamics*
GERARD SALTON, *Theory of Indexing*
CATHLEEN S. MORAWETZ, *Notes on Time Decay and Scattering for Some Hyperbolic Problems*
F. HOPPENSTEADT, *Mathematical Theories of Populations: Demographics, Genetics and Epidemics*
RICHARD ASKEY, *Orthogonal Polynomials and Special Functions*
L. E. PAYNE, *Improperly Posed Problems in Partial Differential Equations*
S. ROSEN, *Lectures on the Measurement and Evaluation of the Performance of Computing Systems*
HERBERT B. KELLER, *Numerical Solution of Two Point Boundary Value Problems*
J. P. LASALLE, *The Stability of Dynamical Systems – Z. ARTSTEIN, Appendix A: Limiting Equations and Stability of Nonautonomous Ordinary Differential Equations*
D. GOTTLIEB and S. A. ORSZAG, *Numerical Analysis of Spectral Methods: Theory and Applications*
PETER J. HUBER, *Robust Statistical Procedures*
HERBERT SOLOMON, *Geometric Probability*
FRED S. ROBERTS, *Graph Theory and Its Applications to Problems of Society*
JURIS HARTMANIS, *Feasible Computations and Provable Complexity Properties*
ZOHAR MANNA, *Lectures on the Logic of Computer Programming*
ELLIS L. JOHNSON, *Integer Programming: Facets, Subadditivity, and Duality for Group and Semi-Group Problems*
SHMUEL WINOGRAD, *Arithmetic Complexity of Computations*
J. F. C. KINGMAN, *Mathematics of Genetic Diversity*

(continued on inside back cover)

E. WARD CHENEY
University of Texas at Austin

Multivariate Approximation Theory: Selected Topics

SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS
PHILADELPHIA, PENNSYLVANIA 1986

Library of Congress Catalog Card Number 86-61533.

ISBN 0-89871-207-6.

Printed for the Society for Industrial and Applied Mathematics by Capital City Press, Montpelier, Vermont.

MULTIVARIATE APPROXIMATION THEORY – SELECTED TOPICS

Contents

Preface

These notes provide an account of lectures given at a Regional Conference on Approximation Theory and Numerical Analysis that was sponsored by the Conference Board of the Mathematical Sciences and supported financially by the National Science Foundation. The host institution was the University of Alaska in Fairbanks, which provided not only all facilities but also additional financial support. Professor Gary Gislason organized the conference with the help of his colleagues in the Department of Mathematics. It is a pleasure to thank them publicly for the comfortable arrangements they made, their warm hospitality, and the efficient running of the conference.

My principal objective in the lectures and in these notes was to describe the current status of several branches of multivariate approximation theory and, if possible, to entice more mathematicians into undertaking research on these matters. I especially had in mind the topics of best approximation, algorithms, and projection operators, since these topics are rife with challenging problems. As part of the survey, I tried to point out the many gaps in the current body of knowledge and to furnish copious references.

The central theme is the perennial problem of "best approximation," usually formulated in a normed linear space whose elements are functions of several real variables. First we ask, "What subclasses of functions are suitable for approximating other functions?" Here interest focuses naturally on functions that are simple combinations of univariate functions. The important tensor-product subspaces play the principal rôle here because of their simple linear structure. One chapter is devoted to introductory material on the tensor product of Banach spaces, as seen from the perspective of approximation theory.

I extend my thanks to all the participants in the conference, for they cheerfully suffered the lectures and offered interesting points of view in addition to stimulating questions.

The manuscript was put into camera-ready form by Ms. Jan Duffy and Ms. Margaret Combs of the Mathematics Department, University of Texas. (The typesetting system used was TEX.) I am very much indebted to Mss. Duffy and Combs for their painstaking work, and their unfailing good judgment in matters æsthetic.

Austin, April 1986 E.W. Cheney

CHAPTER 1
Introduction

One of the principal preoccupations in approximation theory is the study of the following generic problem. A Banach space X is given; its elements are entities that one may wish to approximate. In X a linear subspace U is prescribed; its elements are the objects which will be used for approximation. (Sometimes these are called the **approximants.**) If x is an element of X and u an element of U, then we interpret u as an approximation to x. The quality of this approximation is measured by $\|x - u\|$. The interesting problem of **optimal approximation** or **best approximation** now arises in a natural way: for a fixed x, find an element u in U for which $\|x - u\|$ is as small as possible. To make this more precise, define the **distance** from x to U by the equation

$$\operatorname{dist}(x, U) = \inf_{u \in U} \|x - u\|.$$

If this infimum is *attained* by one or more elements of U, then these elements are called **best approximations** of x in U. Thus, a member u of U is a best approximation of x if and only if

$$\|x - u\| = \operatorname{dist}(x, U).$$

Classical approximation theory has dealt largely with the approximation of functions of one (real or complex) variable. For specific choices of X and U, the theory answers questions concerning the existence of best approximations, their unicity, their characterization and computation. Typical spaces that play the rôle of X are $C(S), L_1(S), L_2(S)$. Typical subspaces that play the rôle of U are the polynomials of degree at most n, the trigonometric polynomials of order n, and spaces of spline functions with fixed knots.

The problems described above are classified as **linear** since the sets of approximants are linear subspaces. (However, the determination of best approximations is usually a *nonlinear* problem.) Some familiar examples of nonlinear sets of approximants are rational functions, splines with variable knots, and various constrained sets of polynomials.

Multivariate approximation theory concerns the approximation of functions of several real or complex variables. In some intuitive sense, multivariate functions should be more complicated than univariate functions, in general. One must expect, for instance, that a multivariate function will be more expensive to compute. This expectation is borne out by experience in solving partial differential equations numerically. If $f(s, t)$ is a solution of a partial differential equation, one might require approximate values of f for 100 different values of s and 100 different values of t. Thus 10,000 computations will be required to give $f(s, t)$ for all combinations of discrete values of s and t. If there are n independent variables instead of just 2, the corresponding number of points in $\mathbb{R}^n$ would be 100^n. So the cost of computing

a function would seem to be an exponential function of the number of variables. This is the first of a number of features that distinguish multivariate approximation theory from its univariate counterpart. In this introductory chapter, some of the other differences between univariate and multivariate approximation will be enumerated and elaborated. Each of these differences gives rise to a direction of investigation that has made the subject of multivariate approximation a fertile field for interesting research.

The subject of interpolation changes in its emphasis as we pass from univariate to multivariate approximation, for reasons to be discussed now. In univariate interpolation, say on an interval $[a, b]$, the subspaces ideal for interpolation are the so-called **Haar** subspaces. An n-dimensional vector space H of functions defined on a domain S is called a **Haar** space if each nonzero element of H has at most $n-1$ zeros in S. An equivalent property is this: for any n distinct points in S, say $s_1, \cdots, s_n$, and for any real numbers $\lambda_1, \cdots, \lambda_n$ there corresponds an element $h \in H$ such that

$$h(s_i) = \lambda_i \qquad (1 \leq i \leq n). \tag{1}$$

Since H is of dimension n, the element h is necessarily unique. Equation (1), of course, expresses the interpolation property. The important theorem of Mairhuber [1956] states that the space $C(S)$, of continuous real-valued functions on a compact Hausdorff space S, contains a Haar subspace of dimension 2 or greater only if S is homeomorphic to a subset of the circumference of a circle. This means that interpolation, as we have described it above, is possible only on one-dimensional topological spaces. For domains in higher dimensions, $\mathbb{R}^2, \mathbb{R}^3, \cdots$, the Haar property is too severe a restriction. The issue here is a matter of the *order* in which we list the quantifiers in our definitions. For interpolation by a Haar subspace of dimension n, *any* set of n points is a satisfactory set of "nodes" for interpolation. If, however, we specify the set of nodes first and *then* seek a suitable subspace for interpolation, many possibilities exist. For example, the method of Shepard interpolation is particularly attractive. In order to describe this, let us begin by recalling the classical Lagrange interpolation formula. If $s_1, \cdots, s_n$ are distinct points in $\mathbb{R}$ then we define

$$Lx = \sum_{i=1}^{n} x(s_i)\ell_i, \qquad \ell_i(s) = \prod_{\substack{j=1 \\ j \neq i}}^{n} (s - s_j)/(s_i - s_j). \tag{2}$$

The functions ℓ_i are polynomials known as the **cardinal** functions in the process. They have the characteristic property $\ell_i(s_j) = \delta_{ij}$. The linear operator L produces, from a given function x, a polynomial Lx, which has the interpolation property: $(Lx)(s_i) = x(s_i)$. What makes this process work is the fact that the function ϕ defined by

$$\phi(s, t) = s - t$$

has the property

$$\phi(s, t) = 0 \quad \text{if and only if} \quad s = t. \tag{3}$$

With *any* function having property (3), an interpolation process can be defined by

$$Ax = \sum_{i=1}^{n} x(s_i)h_i\,, \qquad h_i(s) = \prod_{\substack{j=1 \\ j \neq i}}^{n} \phi(s, s_j)/\phi(s_i, s_j). \tag{4}$$

This is the basic idea of Shepard interpolation (Shepard [1968]). Many choices are possible for the function ϕ, and there are no constraints engendered by the topological nature of S. If S is a metric space, with metric d, an elegant choice for ϕ is d.

Shepard's method also includes the following refinement. Suppose that ϕ is a nonnegative function; i.e., we replace assumption (3) by

$$\phi(s,t) > 0 \quad \text{if} \quad s \neq t \quad \text{and} \quad \phi(s,s) = 0 \quad \text{for all} \quad s. \tag{5}$$

Now define functions g_i by the equation

$$g_i(s) = \prod_{\substack{j=1 \\ j\neq i}}^{n} \phi(s, s_j).$$

Next, let $\overline{g}_i = g_i / \sum_{j=1}^{n} g_j$, and define an operator B by

$$Bx = \sum_{i=1}^{n} x(s_i)\overline{g}_i.$$

One verifies readily these properties of B:

$$\overline{g}_i(s_j) = \delta_{ij}, \tag{6}$$

$$\overline{g}_i(s) > 0 \quad \text{if} \quad s \quad \text{is not a node,} \tag{7}$$

$$\sum_{i=1}^{n} \overline{g}_i = 1, \tag{8}$$

$$B \geq 0; \;\; i.e., \;\; x \geq 0 \Longrightarrow Bx \geq 0. \tag{9}$$

$$\text{If} \;\; x \;\; \text{is a constant function then} \;\; Bx = x. \tag{10}$$

The special case of Shepard interpolation in $\mathbb{R}^k$ in which ϕ is the square of the Euclidean metric is especially favorable, since the basis elements g_i are then analytic functions. Properties (9) and (10) above are "shape-preserving" attributes of the operator B.

The reader who wishes to learn more about Shepard interpolation should consult these additional sources: Barnhill [1977], Barnhill, Dube and Little [1983], Farwig [1984], Gordon and Wixom [1978], and Schumaker [1976].

A third distinctive feature of multivariate approximation theory is the common occurrence of *infinite-dimensional* subspaces for approximants. One example of this arises in approximating a function of two variables by the sum of two functions of one variable:

$$f(s,t) \approx x(s) + y(t).$$

We might assume that f is a continuous function on a Cartesian product $S \times T$ and that x and y are continuous functions on S and T respectively. Since x ranges over $C(S)$ and y ranges over $C(T)$, the approximating subspace is an immense linear subspace in $C(S \times T)$. It is obviously infinite dimensional, if $S \cup T$ is infinite.

Another type of infinite dimensional subspace, which turns out to be quite useful in computer-aided design and in solving integral equations, is involved in this approximation problem:

$$f(s,t) \approx x_0(s) + tx_1(s) + t^2x_2(s) + \cdots + t^n x_n(s). \tag{11}$$

The function on the right in this equation is sometimes called a **pseudo-polynomial.** For each fixed s it is a polynomial in t, but the coefficients x_i can be allowed to be arbitrary continuous functions in s.

A fourth distinctive aspect of multivariate approximation theory is the presence of unusual Banach spaces and unusual norms. More will be said about this in subsequent chapters, but one example here will be typical. In a linear integral equation, an operator of the following type will usually be present:

$$(Lx)(s) = \int_S f(s,\sigma)x(\sigma)\, d\sigma.$$

If the kernel function f is continuous, then L will be an operator from $C(S)$ into $C(S)$. (Here we assume that S is a compact Hausdorff space with a Borel measure.) The norm of L is given by the formula

$$\|L\| = \sup_s \int |f(s,\sigma)|\, d\sigma.$$

The expression on the right is a norm for f, and it is obviously a mixture of the supremum-norm and the L_1-norm. One approach to the numerical solution of integral equations such as

$$Lx = b \quad \text{or} \quad Lx - \lambda x = b$$

is to replace the kernel f by an approximate kernel of **separable** form:

$$k(s,\sigma) = \sum_{i=1}^{n} u_i(s)v_i(\sigma).$$

In doing so, the quantity that should be minimized is

$$\sup_s \int |f(s,\sigma) - k(s,\sigma)|\, d\sigma.$$

The appropriate setting for this approximation is the space $C(S, L_1(S))$ consisting of all continuous maps from S into $L_1(S)$.

This last observation leads to the subject of tensor-products of Banach spaces. Indeed, there is a natural identification (isometry)

$$C[S, L_1(S)] \simeq C(S) \otimes_\lambda L_1(S),$$

the object on the right being the tensor-product of two familiar spaces completed with a particular crossnorm, λ. Chapter II provides an introduction to this important topic.

Another type of space that occurs naturally is of the form $L_1(S, Y)$, where S is a measure space and Y is a Banach space. The new space consists of Bochner integrable maps from S into Y. The Bochner integrable maps, in turn, are defined to be limits of simple functions, and a simple function from S to Y is one of the form

$$f(s) = \sum_{i=1}^{n} \chi_{A_i}(s) y_i$$

where A_i are measurable sets in S, χ denotes a characteristic function, and $y_i \in Y$. More about these spaces will be found in later chapters.

Now we come to another distinctive feature of multivariate approximations: the crucial rôle played by the geometry of the domain. We illustrate with a two variable problem. Let D be a compact domain in $\mathbb{R}^2$. The projections of D onto the two coordinate axes are denoted by S and T. We contemplate the approximation of an $f \in C(D)$ by an element of $C(S)$ plus an element of $C(T)$:

$$f(s,t) \approx x(s) + y(t).$$

The existence of an optimal pair (x, y) in this problem depends strongly upon the geometry of D. If $D = S \times T$ then existence is assured. If, however, D is defined as

$$D = \{(s,t) \ : \ 0 \leq \tfrac{1}{2}s \leq t \leq s \leq 1\},$$

then some functions f will not possess best approximations. See von Golitschek [1984].

The strong influence of the geometry is also a familiar phenomenon in the numerical solution of partial differential equations. One has only to think of the Dirichlet problem on a circle or on an L-shaped region to appreciate the profound influence of the domain.

Another obvious difference between univariate and multivariate approximation is that in the latter, our "building blocks" will undoubtedly be univariate functions, combined in appropriate ways for the problem at hand. Among the various schemata that have been used for bivariate approximation are the following:

$$f(s,t) \approx \phi(x(s) + y(t)),$$

$$f(s,t) \approx \sum_{i=1}^{n} x_i(s) y_i(t),$$

$$f(s,t) \approx \sum_{i=1}^{n} g_i(s) y_i(t) + \sum_{i=1}^{m} h_i(t) x_i(s),$$

$$f(s,t) \approx \sum_{i=1}^{n} \phi_i(g(s) + h(t)).$$

All of these approximating functions are elementary combinations of univariate functions.

Often, the approximating forms that are useful are suggested by partial differential equations. Thus the general pseudopolynomial in two variables

$$v(s,t) = \sum_{i=0}^{m-1} x_i(s)t^i + \sum_{i=0}^{n-1} y_i(t)s^i$$

is the generic element in the null space of the partial differential operator

$$D \equiv \frac{\partial^m}{\partial t^m} \cdot \frac{\partial^n}{\partial s^n}.$$

This observation could be useful in solving a partial differential equation of the form $Du = b$ subject to boundary conditions, since one could start with *any* solution u of the equation $Du = b$ and add to it a pseudopolynomial of the form v in an attempt to satisfy the boundary conditions with $v + u$.

Another well-known example is the wave equation, $u_{ss} = u_{tt}$. After making the change of variables $\xi = s + t$ and $\eta = s - t$ we arrive at the equation $u_{\xi\eta} = 0$, whose solution is any function of the form

$$u = x(\xi) + y(\eta) = x(s + t) + y(s - t).$$

Thus the approximation problem

$$f(s,t) \approx x(s) + y(t)$$

arises when a function is to be approximated by a solution of the wave equation.

Another feature of multivariate approximation is that linear projection operators play a more important rôle. There are several reasons for this. First, algorithms for computing best approximations of multivariate functions are in a relatively primitive state, and consequently projections provide some of the best practical procedures for approximation. Second, the algebraic and lattice-theoretic structure of projections is well understood, and there is no difficulty in combining univariate projections in various ways to produce useful multivariate projections, especially when a tensor-product type of operator is practical. By contrast, it is not known how to combine univariate *proximity* maps (best approximation operators) in order to create multivariate proximity maps. Although the alternating algorithm (or Diliberto-Straus algorithm) is an exception to this statement, its applicability is strictly limited.

In this long catalog of differences between the univariate and multivariate theories, the final point is that *proximinality* (existence of best approximations) becomes a rather difficult question in multivariate approximation. An example of an open question in this area is this: Does each continuous function on the unit square of $\mathbb{R}^2$ possess a best approximation of the following form, with x_i and y_i continuous?

$$x_0(s) + x_1(s)t + y_0(t) + y_1(t)s$$

In the previous discussion, we have alluded to the use of univariate functions in approximating multivariate functions. Is there a sound basis for expecting this

to be possible? The answer to this is "Yes" and is to be found in some of the most enlightening theorems of analysis. Let us begin with the classical Weierstrass Approximation Theorem, which asserts that any continuous real-valued function defined on a compact interval of the real line can be approximated uniformly to arbitrary precision by a polynomial. This theorem can be said to motivate the study of polynomial approximations in one variable.

For functions of several variables, the Stone-Weierstrass Theorem is useful. If Q is a compact Hausdorff space, then a subalgebra $\mathcal{A}$ of the real space $C(Q)$ will be dense in $C(Q)$ provided that $\mathcal{A}$ separates the points of Q and contains the function 1. An immediate corollary of this is that the polynomials in two variables form a dense set in $C(Q)$ whenever Q is a compact subset of $\mathbb{R}^2$. Polynomials in two variables are built up in an obvious way from univariate functions. Similarly, the functions of the form $(s,t) \mapsto \sum_{i=1}^{n} x_i(s)y_i(t)$, with continuous functions x_i and y_i constitute a dense set in $C(Q)$.

When univariate functions are to be combined by *functional composition*, the theorem of Kolmogorov is pertinent. We discuss his theorem next, starting with the concept of a **nomographic function.** The latter is a bivariate function which can be expressed in the form

$$\phi(s,t) = g(x(s) + y(t))$$

for appropriate continuous functions g, x, y. Note that these three functions are *univariate*. Some common functions are nomographic; for example, if $0 < a < b < \infty$, then the function $f(s,t) = st$ is nomographic on $[a,b] \times [a,b]$. This is demonstrated by the equation

$$st = \exp(\log s + \log t).$$

On the other hand, the same function is *not* nomographic on $[0,1] \times [0,1]$. This nice example comes from Buck's paper [1968-a].

If we denote by $\mathcal{N}$ the set of nomographic functions on I^2, where $I = [0,1]$, then we can say that $\mathcal{N}$ is a *proper* subset of $C(I^2)$. As for *sums* of two or more nomographic functions, we obviously have

$$\mathcal{N} \ \subset \ \mathcal{N}+\mathcal{N} \ \subset \ \mathcal{N}+\mathcal{N}+\mathcal{N} \ \subset \ \cdots \ \subset \ C(I^2).$$

Kolmogorov's Theorem asserts that this chain of inclusions is finite, for

$$C(I^2) = \mathcal{N}+\mathcal{N}+\mathcal{N}+\mathcal{N}+\mathcal{N},$$

i.e., each continuous function on the square is a sum of 5 nomographic functions. This remarkable theorem provides a solution to the 13^{th} problem in Hilbert's famous list [1901]. For an up-to-date account of the status of Hilbert's problems, including the 13^{th}, see Lorentz's article in Browder's survey [1976].

Kolmogorov's Theorem has undergone various refinements since its publication [1957]. These can be found in Lorentz [1966]. Perhaps the most memorable refinement is this one:

THEOREM. *There exist five functions,* $g_1, \cdots, g_5$, *which are continuous on* $[0,1]$, *such that to each continuous function* f *defined on the square* $[0,1] \times [0,1]$ *there corresponds a continuous function* ϕ *satisfying*

$$f(s,t) = \sum_{i=1}^{5} \phi(g_i(s) + \sqrt{2}\, g_i(t)).$$

Thus, the functions $g_1, \cdots, g_5$ are *universal* functions, and it is only ϕ which depends on f.

CHAPTER 2
Tensor Products

Let us begin with the abstract definition, and then look at some concrete realizations that will be useful in approximation theory. We start with two Banach spaces, X and Y. From these we construct "expressions" of the form $\sum_{i=1}^{n} x_i \otimes y_i$, where $n \in \mathbb{N}$, $x_i \in X$, and $y_i \in Y$. (The integer n is *not* restricted.) In the set of all "expressions" we introduce an equivalence relation and then make the set into a vector space with appropriate definitions.

If the elements of X are functions defined on a set S and if the elements of Y are functions defined on a set T, then the "expression" $\sum x_i \otimes y_i$ can be interpreted as a function on $S \times T$, namely the function

$$(s,t) \mapsto \sum_{i=1}^{n} x_i(s) y_i(t).$$

For the purposes of approximation theory this is the principal interpretation given to "expressions." However, the abstract theory proceeds in a way that makes no assumptions about the nature of the elements in X and Y.

The set of all expressions is denoted by $X \otimes Y$; this is the **algebraic tensor product** of X and Y. An equivalence relation in $X \otimes Y$ is defined as follows: we write

$$\sum_{i=1}^{n} x_i \otimes y_i \simeq \sum_{j=1}^{m} u_j \otimes v_j \tag{1}$$

if and only if

$$\sum_{i=1}^{n} \langle \phi, x_i \rangle \langle \psi, y_i \rangle = \sum_{j=1}^{m} \langle \phi, u_j \rangle \langle \psi, v_j \rangle \tag{2}$$

for all $\phi \in X^*$ and for all $\psi \in Y^*$. (The notation X^* denotes the Banach-space conjugate of X.)

The perceptive reader will see that what we are doing is *forcing* the elements of X and Y to be functions, indeed functions on X^* and Y^*, by the mental exercise of thinking

$$x(\phi) = \langle \phi, x \rangle \qquad (x \in X, \phi \in X^*).$$

Thus the two expressions in (1) are to be regarded as equivalent if they represent the same *function* on $X^* \times Y^*$ as defined in (2).

If (1), and hence (2), are true for a pair of expressions then from (2) we can write

$$\left\langle \psi, \sum_{i=1}^{n} \langle \phi, x_i \rangle y_i \right\rangle = \left\langle \psi, \sum_{j=1}^{m} \langle \phi, u_j \rangle v_j \right\rangle. \tag{3}$$

Since this is true for all $\psi \in Y^*$ we infer that

$$\sum_{i=1}^{n} \langle \phi, x_i \rangle y_i = \sum_{j=1}^{m} \langle \phi, u_j \rangle v_j \qquad (\phi \in X^*). \tag{4}$$

This argument is reversible, and so (4) could have been adopted as the definition of equivalence. From (4) we perceive another interpretation to be given any expression $\sum x_i \otimes y_i$; namely, it can be thought of as a linear operator acting on X^* and taking values in Y. Thus the equation

$$\left(\sum_{i=1}^{n} x_i \otimes y_i \right)(\phi) = \sum_{i=1}^{n} \langle \phi, x_i \rangle y_i \qquad (\phi \in X^*) \tag{5}$$

defines the action of the expression when that expression is interpreted as an operator. This operator is linear and continuous. The continuity can be verified by showing that the operator is *bounded*:

$$\left\| \left(\sum x_i \otimes y_i \right)(\phi) \right\| = \left\| \sum \langle \phi, x_i \rangle y_i \right\| \leq \sum \|\phi\| \, \|x_i\| \, \|y_i\|.$$

Because of (4), we can say that two expressions are equivalent if they define the same operator from X^* to Y. Henceforth, $\simeq$ is written as an ordinary equality.

In the abstract theory, an expression $\sum_{i=1}^{n} x_i \otimes y_i$ is given the *basic* interpretation as a bounded linear operator from X^* to Y, hence an element of $\mathcal{L}(X^*, Y)$. The algebraic structure given to $X \otimes Y$ is derived from this interpretation. Thus addition *must* be defined by

$$\sum_{i=1}^{n} x_i \otimes y_i + \sum_{i=n+1}^{m} x_i \otimes y_i = \sum_{i=1}^{m} x_i \otimes y_i$$

because for any $\phi \in X^*$,

$$\sum_{i=1}^{n} \langle \phi, x_i \rangle y_i + \sum_{i=n+1}^{m} \langle \phi, x_i \rangle y_i = \sum_{i=1}^{m} \langle \phi, x_i \rangle y_i.$$

Scalar multiplication is defined by

$$\lambda \sum_{i=1}^{n} x_i \otimes y_i = \sum_{i=1}^{n} \lambda x_i \otimes y_i = \sum_{i=1}^{n} x_i \otimes (\lambda y_i).$$

The 0-element has many representations, among them $0 \otimes 0$, $x \otimes 0$, and $0 \otimes y$.

LEMMA. *In $X \otimes Y$, every expression which is not equivalent to $0 \otimes 0$ is equivalent to an expression $\sum_{i=1}^{n} x_i \otimes y_i$ in which $\{x_1, \cdots, x_n\}$ and $\{y_1, \cdots, y_n\}$ are linearly independent sets.*

Proof. Given any nonzero element z in $X \otimes Y$, let $\sum_{i=1}^{n} x_i \otimes y_i$ be an expression for z having a minimum number of summands. Then the sets of x_i's and y_i's are

independent. To prove this, suppose that one x_i is a linear combination of the others, say $x_1 = \sum_{i=2}^n \lambda_i x_i$. Then

$$\begin{aligned}\sum_{i=1}^n x_i \otimes y_i &= \left(\sum_{i=2}^n \lambda_i x_i\right) \otimes y_1 + \sum_{i=2}^n (x_i \otimes y_i) \\ &= \sum_{i=2}^n x_i \otimes \lambda_i y_1 + \sum_{i=2}^n (x_i \otimes y_i) \\ &= \sum_{i=2}^n x_i \otimes (\lambda_i y_1 + y_i).\end{aligned}$$

This equation contradicts the assumed minimality of n. The argument is valid for $n > 1$. If $n = 1$, and a dependency exists, then either $x_1 = 0$ or $y_1 = 0$, whence $z = 0 \otimes 0$. □

After the algebraic tensor product of two Banach spaces has been constructed, we can consider giving it a norm. There are many ways in which this can be done, and we begin with the simplest. Since $X \otimes Y$ can be interpreted as a subspace of $\mathcal{L}(X^*, Y)$, the norm on the latter space induces a norm on the former. This norm is denoted by λ because it turns out to be the *least* norm in a certain class. By the preceding remarks, its definition must be

$$\lambda\left(\sum_{i=1}^n x_i \otimes y_i\right) = \sup\left\{\left\|\sum_{i=1}^n \langle \phi, x_i\rangle y_i\right\| \;:\; \phi \in X^*,\ \|\phi\| = 1\right\}.$$

Let us illustrate these matters by taking the tensor product of two continuous function spaces, and giving it the norm λ. If S is a compact Hausdorff space, let $C(S)$ denote the Banach space of continuous real-valued functions on S, normed by $\|x\| = \max_S |x(s)|$. Let $C(T)$ be another such space. If $x_i \in C(S)$ and $y_i \in C(T)$ then, with ϕ running over the unit cell in $C(S)^*$, we have

$$\begin{aligned}\lambda\left(\sum_{i=1}^n x_i \otimes y_i\right) &= \sup_\phi \left\|\sum_{i=1}^n \langle \phi, x_i\rangle y_i\right\| \\ &= \sup_\phi \sup_t \left|\sum_{i=1}^n \langle \phi, x_i\rangle y_i(t)\right| \\ &= \sup_t \sup_\phi \left|\left\langle \phi, \sum_{i=1}^n y_i(t) x_i \right\rangle\right| \\ &= \sup_t \left\|\sum_{i=1}^n y_i(t) x_i\right\| \\ &= \sup_t \sup_s \left|\sum_{i=1}^n y_i(t) x_i(s)\right|.\end{aligned}$$

Hence, if the expression $z = \sum_{i=1}^n x_i \otimes y_i$ is interpreted as an element of $C(S \times T)$, its standard supremum norm is the same as $\lambda(z)$, which is in turn the norm of z when z is regarded as a linear operator from $C(S)^*$ to $C(T)$.

The norm λ has a useful property, called the **crossnorm** property: for any Banach spaces X and Y and any elements x, y,

$$\lambda(x \otimes y) = \|x\| \, \|y\|.$$

(This follows easily from the definition of λ.)

If a norm α has been introduced into $X \otimes Y$, the resulting normed linear space is usually incomplete. Its completion is denoted by $X \otimes_\alpha Y$. We illustrate again with $C(S) \otimes C(T)$.

THEOREM. $C(S) \otimes_\lambda C(T) = C(S \times T)$.

Proof. The equality asserted here means that $C(S) \otimes_\lambda C(T)$ is isometrically isomorphic to $C(S \times T)$. We have already seen that there is a natural linear embedding of $C(S) \otimes C(T)$ in $C(S \times T)$, and that when the λ-norm is used, this embedding is isometric. The completion, $C(S) \otimes_\lambda C(T)$, is then nothing but the closure of $C(S) \otimes C(T)$ in the space $C(S \times T)$. But an easy application of the Stone-Weierstrass Theorem shows that $C(S) \otimes C(T)$ is *dense* in $C(S \times T)$. Hence equality follows. □

A more general theorem of Grothendieck states that $C(S) \otimes_\lambda X = C(S, X)$ for any Banach space X. For the proof, see, for example, Semadeni [1971]. The space $C(S, X)$ has as its elements all continuous maps $f : S \to X$, normed by $\|f\| = \sup_S \|f(s)\|$. It is a Banach space.

If one of the spaces X and Y is finite-dimensional, then $X \otimes_\alpha Y$ is more easily described, and no completion is necessary. For definiteness suppose that X is finite-dimensional and that $\{b_1, \cdots, b_n\}$ is a basis. Then each expression in $X \otimes Y$ is equivalent to an expression $\sum_{i=1}^n b_i \otimes y_i$, as is seen with elementary algebraic manipulations.

LEMMA. *If X is finite-dimensional and if α is a crossnorm on $X \otimes Y$ which satisfies $\lambda \le \alpha$, then $X \otimes Y$ is complete in the metric defined by the α-norm.*

Proof. Select $\phi_1, \cdots, \phi_n$ in X^* so that $\phi_i(b_j) = \delta_{ij}$. Let $[z_k]$ be a sequence in $X \otimes Y$ which is α-Cauchy. Then write $z_k = \sum_{i=1}^n b_i \otimes y_{ki}$. For $1 \le j \le n$ we have

$$\|y_{kj}\| = \left\| \sum_{i=1}^n \langle \phi_j, b_i \rangle y_{ki} \right\| = \|z_k(\phi_j)\| \le \lambda(z_k) \|\phi_j\| \le \alpha(z_k) \|\phi_j\|.$$

In this calculation, z_k has been interpreted as an element of $\mathcal{L}(X^*, Y)$. A similar calculation shows that $[y_{kj}]_{k=1}^\infty$ is Cauchy for each j. Since Y is a Banach space, $\lim_k y_{kj} = y_j$ exists. Putting $z = \sum_{i=1}^n b_i \otimes y_i$ we shall see that $\lim x_k = z$. Indeed,

$$\alpha(z_k - z) = \alpha\left[\sum_{i=1}^n b_i \otimes (y_{ki} - y_i) \right] \le \sum_{i=1}^n \alpha[b_i \otimes (y_{ki} - y_i)] = \sum_{i=1}^n \|b_i\| \, \|y_{ki} - y_i\|. \ \square$$

Another important crossnorm is denoted by γ (because it turns out to be the greatest crossnorm). Its definition is

$$\gamma(z) = \inf \left\{ \sum_{i=1}^n \|x_i\| \, \|y_i\| \ : \ z = \sum_{i=1}^n x_i \otimes y_i \right\}.$$

The infimum is taken with respect to all possible representations of $z \in X \otimes Y$. It is clear that an infimum is essential, for otherwise $\gamma(z)$ would depend on the

representation of z. If $z \neq 0$, then $\gamma(z) > 0$ by the following reasoning. Let ϕ be any element of norm 1 in X^*. We shall interpret z as an operator in $\mathcal{L}(X^*, Y)$. If $z = \sum_{i=1}^n x_i \otimes y_i$, then

$$\|z(\phi)\| = \left\| \sum_{i=1}^n \langle \phi, x_i \rangle y_i \right\| \leq \sum_{i=1}^n \|x_i\| \, \|y_i\|.$$

By taking an infimum over all representations of z we get $\gamma(z) \geq \|z(\phi)\|$. By taking a supremum in ϕ we get $\gamma(z) \geq \lambda(z)$.

Let us prove that γ is a crossnorm. Suppose that $x \otimes y = \sum_{i=1}^n x_i \otimes y_i$. Select $\phi \in X^*$ such that $\|\phi\| = 1$ and $\langle \phi, x \rangle = \|x\|$. Then

$$\|x\| \, \|y\| = \langle \phi, x \rangle \|y\| = \|\langle \phi, x \rangle y\| = \left\| \sum_{i=1}^n \langle \phi, x_i \rangle y_i \right\| \leq \sum_{i=1}^n \|x_i\| \, \|y_i\|.$$

Hence $\gamma(x \otimes y) = \|x\| \, \|y\|$.

An important application of the γ-norm is contained in the next result, due to Dunford and Schatten [1946].

THEOREM. *If S and T are σ-finite measure spaces, then*

$$L_1(S) \otimes_\gamma L_1(T) = L_1(S \times T).$$

Proof. Let $z = \sum_{i=1}^n x_i \otimes y_i$, where $x_i \in L_1(S)$ and $y_i \in L_1(T)$. With z we associate an element $f \in L_1(S \times T)$ by writing

$$f(s,t) = \sum_{i=1}^n x_i(s) y_i(t).$$

The function f depends on z and not upon its representation – a fact which we do not stop to prove. By Fubini's Theorem,

$$\|f\| = \iint |f(s,t)| \, ds \, dt \leq \iint \sum_{i=1}^n |x_i(s)| \, |y_i(t)| \, ds \, dt = \sum_{i=1}^n \|x_i\| \, \|y_i\|.$$

This proves that $\|f\| \leq \gamma(z)$. In the special case that $\sum_{i=1}^n x_i = 1$, and $x_i x_j = 0$ for $i \neq j$, we can see that $\|f\| \geq \gamma(z)$. Indeed, in this case the x_i are characteristic functions of a measurable partition of S, and therefore

$$\begin{aligned} \iint \left| \sum_{i=1}^n x_i(s) y_i(t) \right| dt \, ds &= \sum_{i=1}^n \int x_i(s) \, ds \int |y_i(t)| \, dt \\ &= \sum_{i=1}^n \|x_i\| \, \|y_i\| \geq \gamma(z). \end{aligned}$$

The functions of the special type just described form a dense set in $L_1(S \times T)$, and this remark completes the proof. □

A more general theorem exists here as in the analogous situation involving $C(S \times T)$, namely, $L_1(S) \otimes_\gamma X = L_1(S, X)$ for any Banach space X. The space $L_1(S, X)$ is defined as the space of all Bochner integrable maps $f : S \to X$, with norm

$$\|f\| = \int_S \|f(s)\|_X \, ds.$$

The theory of these spaces is expounded in Diestel and Uhl [1977], and an introductory account is given in Light and Cheney [1985].

The theory of tensor product spaces has a number of isomorphism theorems as highlights; several of these have already been mentioned. Another one, of great importance in approximation theory, is this:

THEOREM. $(X \otimes_\gamma Y)^* = \mathcal{L}(X, Y^*)$.

Proof. We define $T : (X \otimes_\gamma Y)^* \to \mathcal{L}(X, Y^*)$ by putting

$$\langle (T\theta)(x), y \rangle = \langle \theta, x \otimes y \rangle, \qquad \theta \in (X \otimes_\gamma Y)^*,\ x \in X,\ y \in Y. \tag{6}$$

It is easy to see that for each θ, $T\theta$ is a linear map from X onto Y^*. Furthermore, $T\theta$ is bounded because

$$\begin{aligned} \|T\theta\| &= \sup_{\|x\|=1} \|(T\theta)(x)\| \\ &= \sup_{\|x\|=1} \sup_{\|y\|=1} |\langle (T\theta)x, y \rangle| \\ &= \sup_{\gamma(x \otimes y)=1} |\langle \theta, x \otimes y \rangle| \leq \|\theta\|. \end{aligned}$$

Hence $T\theta \in \mathcal{L}(X, Y^*)$. From Eq. (6) we see that T is injective, because if $T\theta = 0$ then the left side in (6) is 0 for all x and y. Consequently the right side is zero for all dyads $x \otimes y$, and so $\theta = 0$.

In order to prove that T is surjective, let $A \in \mathcal{L}(X, Y^*)$. Define θ on $X \otimes Y$ as follows. If $z \in X \otimes Y$, let one of its representations be $z = \sum_{i=1}^n x_i \otimes y_i$. Put $\langle \theta, z \rangle = \sum_{i=1}^n \langle Ax_i, y_i \rangle$. We shall see that $\langle \theta, z \rangle$ does not depend on this representation. We have

$$\begin{aligned} |\langle \theta, z \rangle| &\leq \left| \sum_{i=1}^n \langle Ax_i, y_i \rangle \right| \\ &\leq \sum_{i=1}^n |\langle Ax_i, y_i \rangle| \\ &\leq \|A\| \sum_{i=1}^n \|x_i\| \, \|y_i\|. \end{aligned}$$

By taking an infimum over all the representations of z we obtain

$$|\langle \theta, z \rangle| \leq \|A\| \gamma(z).$$

Thus if $z \simeq 0 \times 0$, then $\gamma(z) = 0$ and $\theta(z) = 0$. These considerations show that θ can be extended to $X \otimes_\gamma Y$ and that $\|\theta\| \leq \|A\|$. Since $T\theta = A$, we have proved that T is surjective. Putting the various inequalities together, we see that $\|T\theta\| = \|\theta\|$; hence T is an isometric isomorphism. □

The applications of the preceding theorem in approximation theory are especially interesting. Since the space $\mathcal{L}(X, Y^*)$ is the conjugate of a Banach space, it has a natural weak*-topology. In some contexts this topology is called the **weak*-operator** topology. Convergence of a net A_ν in this topology is described as follows: $A_\nu \to 0$ in the weak*-operator topology of $\mathcal{L}(X, Y^*)$ means that $\langle A_\nu x, y\rangle \to 0$ for all $x \in X$ and $y \in Y$. By the Alaoglu Theorem, each bounded subset of $\mathcal{L}(X, Y^*)$ is pre-compact in this topology. It follows that a set that is closed in this topology is automatically proximinal. As an example of this principle, we have:

THEOREM. *For every Banach space X and for each finite-dimensional subspace U in X there is a projection of least norm from X onto U.*

Proof. The family of all projections $P : X \twoheadrightarrow U$ is closed in the weak*-operater topology. Indeed, if P_ν is a net of such projections and if $P_\nu \to A$ in this topology then for each $u \in U$, $Au = \lim_\nu P_\nu u = u$. Here we reason that U is a conjugate space because it is finite-dimensional, and its weak*-topology coincides with its norm topology. See Isbell and Semadeni [1963] or Blatter and Cheney [1971]. □

If X and Y are Hilbert spaces, we can introduce an inner product in $X \otimes Y$ by the following definition: if $z = \sum_{i=1}^n x_i \otimes y_i \in X \otimes Y$ and $w = \sum_{j=1}^m u_j \otimes v_j \in X \otimes Y$, then

$$\langle z, w\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m}\langle x_i, u_j\rangle\langle y_i, v_j\rangle.$$

It must be proved, of course, that this definition is independent of the representations chosen for z and w, and that the resulting bilinear form is a genuine inner product. If we define

$$\beta(z) = \sqrt{\langle z, z\rangle}, \qquad z \in X \otimes Y,$$

then the function β is a crossnorm, and finally we have the following isomorphism theorem.

THEOREM. $L_2(S) \otimes_\beta L_2(T) = L_2(S \times T)$.

The details of this topic can be found in Light and Cheney [1985].

CHAPTER 3

Applications to Integral Equations and Matrix Scaling

Integral equations have motivated functional analysis since its inception, and they serve as motivation for certain branches of approximation theory as well.

Consider the following integral equation

$$x(t) = q(t) + \lambda \int_S f(s,t)x(s)\,ds \qquad (t \in S). \tag{1}$$

This is an example of a **linear Fredholm integral equation**. The functions q and f are prescribed, and the function x is to be determined. Various settings are possible, but let us start by assuming that $q \in C(S)$, $f \in C(S \times S)$, and $x \in C(S)$. The best way to think of Eq. (1) is in the form

$$x = q + \lambda A x \tag{2}$$

in which an integral operator A has been introduced by defining

$$(Ax)(t) = \int_S f(s,t)x(s)\,ds. \tag{3}$$

In specific instances, S is often an interval in $\mathbb{R}$, but it can be a more general space, such as any topological space in which a Borel measure has been introduced. Notice that Eq. (2) can be written as

$$(I - \lambda A)x = q. \tag{4}$$

A general theorem of functional analysis asserts that if $\|A\| < \infty$ and if $|\lambda|$ is sufficiently small, then a unique solution exists. It can be computed by the Neumann series:

$$x = (I - \lambda A)^{-1} q = \sum_{n=0}^{\infty} (\lambda A)^n q. \tag{5}$$

The exact hypothesis required is that $|\lambda| < \|A\|^{-1}$, where

$$\begin{aligned} \|A\| &= \sup\{\|Ax\| \ : \ x \in C(S),\ \|x\| \le 1\} \\ &= \sup_x \sup_t |(Ax)(t)| \\ &= \sup_t \sup_x \left| \int_S f(s,t)x(s)\,ds \right| \\ &= \sup_t \int_S |f(s,t)|\,ds. \end{aligned}$$

This calculation is valid if S is a compact Hausdorff space and if the measure employed is a regular Borel measure.

If the kernel, f, of the integral operator is of the special form

$$f(s,t) = \sum_{j=1}^{n} u_j(s)v_j(t), \tag{6}$$

then it is said to be "separable" or "degenerate." From our viewpoint, this means that f is already displayed as a simple combination of univariate functions. An elementary method for solving an integral equation with degenerate kernel proceeds as follows. Upon substituting (6) in the integral equation (1), we obtain

$$\begin{aligned} x(t) &= q(t) + \lambda \int_S \sum_{j=1}^{n} u_j(s)v_j(t)x(s)\,ds \\ &= q(t) + \lambda \sum_{j=1}^{n} v_j(t) \int_S u_j(s)x(s)\,ds. \end{aligned} \tag{7}$$

After introducing the abbreviation

$$\langle y, x\rangle = \int_S y(s)x(s)\,ds$$

we can write (7) in the form

$$x = q + \lambda \sum_{j=1}^{n} \langle u_j, x\rangle v_j. \tag{8}$$

This shows that if there exists a solution, x, then $x - q$ must lie in the subspace spanned by $\{v_1, \cdots, v_n\}$. We therefore put $c_j = \langle x, u_j\rangle$, and arrive at

$$c_i = \langle x, u_i\rangle = \langle q, u_i\rangle + \lambda \sum_{j=1}^{n} c_j \langle v_j, u_i\rangle \qquad (1 \le i \le n)$$

whence

$$\sum_{j=1}^{n} \big[\delta_{ij} - \lambda\langle v_j, u_i\rangle\big] c_j = \langle q, u_i\rangle \qquad (1 \le i \le n).$$

In matrix terms this equation has the form

$$(I - \lambda B)c = b \tag{9}$$

and for all sufficiently small λ a unique solution will exist. Finally, we note that if c is a solution of (9), then the function x defined by

$$x = q + \sum_{j=1}^{n} c_j v_j$$

will be a solution of the integral equation (9). Thus in the case of a separable kernel, the Fredholm integral equation (1) is equivalent to a system of n linear equations in n unknowns.

If the kernel f in Eq. (1) is not separable, we can approximate it by a separable kernel and solve the resulting equation by the method outlined above. This tactic produces an approximate solution of the original integral equation. The quality of the approximation is easily assessed in the following manner. We start with two invertible operators A and B (on any Banach space), and observe that

$$\|A^{-1} - B^{-1}\| = \|A^{-1}(B - A)B^{-1}\| \leq \|A^{-1}\| \, \|B - A\| \, \|B^{-1}\|.$$

From this we conclude that

$$\|(I - A)^{-1} - (I - B)^{-1}\| \leq \|(I - A)^{-1}\| \, \|A - B\| \, \|(I - B)^{-1}\|. \tag{10}$$

Now suppose that x is a solution of Eq. (4). Let $\tilde{A}$ be a perturbation of A, and let $\tilde{x}$ be a solution of

$$(I - \lambda\tilde{A})\tilde{x} = q. \tag{11}$$

Then from Eq. (10) we obtain

$$\begin{aligned} \|x - \tilde{x}\| &= \|(I - \lambda A)^{-1}q - (I - \lambda\tilde{A})^{-1}q\| \\ &\leq \|(I - \lambda A)^{-1}\| \, \|\lambda A - \lambda\tilde{A}\| \, \|(I - \lambda\tilde{A})^{-1}\| \, \|q\|. \end{aligned}$$

This proves that if $\tilde{A}$ converges to A, then $\tilde{x}$ will converge to x. It also suggests that, if we are going to replace A by $\tilde{A}$ and solve (11), then $\|A - \tilde{A}\|$ should be as small as possible. These remarks are valid for operator equations in *any* Banach space. If we work in the space $C(S)$ with integral operators, the original kernel f and the separable kernel $\tilde{f}$ should have the property that the quantity

$$\|A - \tilde{A}\| = \sup_{t \in S} \int_S |f(s,t) - \tilde{f}(s,t)| \, ds$$

is small. The appropriate setting for this problem is the space $C(S, L_1(S))$.

The preceding discussion suggests a related problem of approximating a given function $f \in C(S \times T)$ by a function of the form

$$\tilde{f}(s,t) = \sum_{i=1}^{n} u_i(s)v_i(t)$$

where $u_i \in C(S)$ and $v_i \in C(T)$. If n is fixed and if u_i and v_i are not further restricted, then this problem of best approximation is one that has not yet been solved. There do exist reasonable methods based upon projections that produce *good* approximations, but the underlying extremal problem remains unsolved.

Let us interrupt our discussion of integral equations to indicate how the best approximation problem

$$f(s,t) \approx \sum_{i=1}^{n} u_i(s)v_i(t)$$

leads to a problem in n-widths. Recall that the **n-width** of a set K in a Banach space X is the number

$$d_n(K) = \inf_{U} \sup_{x \in K} \operatorname{dist}(x, U)$$

in which the infimum is taken over all n-dimensional subspaces U in X. If a certain n-dimensional subspace U *achieves* the indicated infimum, then U is termed an **optimal** subspace for approximating the members of K. (See Pinkus [1985] for the theory of n-widths.)

In the problem at hand, the manifold of admissible approximating functions is

$$M_n = \left\{ \sum_{i=1}^{n} u_i \otimes v_i : u_i \in C(S),\, v_i \in C(T) \right\}.$$

It is *not* a linear manifold, but rather a Grassmann manifold. The distance from f to M_n can be expressed as follows:

$$\begin{aligned} \operatorname{dist}(f, M_n) &= \inf_{u_1 \cdots u_n} \inf_{v_1 \cdots v_n} \left\| f - \sum_{i=1}^{n} u_i \otimes v_i \right\| \\ &= \inf_{U} \operatorname{dist}\big[f, U \otimes C(T)\big] \\ &= \inf_{U} \sup_{t} \operatorname{dist}(f^t, U) \\ &= d_n(K), \end{aligned}$$

where K is the compact subset of $C(S)$ consisting of the sections f^t as t ranges over T. Thus the approximation problem will be solved if we can determine an optimal subspace U in the n-width problem. The converse is true also, as is shown by the above calculation.

If the setting of the approximation problem is shifted to $L_2(S \times S)$, the situation changes, and some early work of Schmidt [1907] becomes pertinent. Suppose then that we desire to approximate a function $f \in L_2(S \times S)$ by a function $\sum_{i=1}^{n} u_i(s) v_i(t)$, where n is fixed, $u_i \in L_2(S)$, and $v_i \in L_2(S)$. This is the problem that Schmidt solved, and we shall outline his solution here.

Before delving into the details of Schmidt's work, we recall some elementary approximation theory in Hilbert space. Suppose that z and v are elements of a Hilbert space, and that we desire to approximate z by a scalar multiple of v. We then calculate as follows:

$$\|z - \lambda v\|^2 = \langle z - \lambda v, z - \lambda v \rangle = \langle z, z \rangle - 2\lambda \langle z, v \rangle + \lambda^2 \langle v, v \rangle.$$

In this problem, no generality is lost if it be assumed that $\|v\| = 1$. Having made that assumption, we will have

$$\begin{aligned} \|z - \lambda v\|^2 &= \lambda^2 - 2\lambda \langle z, v \rangle + \langle z, v \rangle^2 + \langle z, z \rangle - \langle z, v \rangle^2 \\ &= \big[\lambda - \langle z, v \rangle\big]^2 + \|z\|^2 - \langle z, v \rangle^2. \end{aligned}$$

This shows that the best value of λ is $\lambda = \langle z, v \rangle$, and that when this optimal λ is used, the squared error is

$$\|z\|^2 - \langle z, v \rangle^2.$$

If v is the first element in an orthonormal sequence, $[v_1, v_2, v_3, \cdots]$ then we can repeat this process in the following way. In Step 2, we approximate $z - \langle z, v_1\rangle v_1$ as well as possible by a multiple of v_2. The correct coefficient is (because of orthogonality)

$$\lambda = \big\langle z - \langle z, v_1\rangle v_1, v_2\big\rangle = \langle z, v_2\rangle.$$

Thus, we can compute the optimal coefficients one by one or all at once. The recursive method is of course the only one available if the orthonormal system $[v_i]$ is not known at the beginning but is being determined one step at a time in the same process. The coefficients being developed are nothing but the generalized Fourier coefficients of z with respect to the given orthonormal system.

In order to understand Schmidt's result, let us examine the case when $n = 1$. We seek functions u and $v \in L^2(S)$ which minimize the squared-norm:

$$\iint \big[f(s,t) - u(s)v(t)\big]^2 \, dt\, ds. \tag{12}$$

Let us suppose for the moment that the function v is fixed. There is no loss of generality in assuming that $\|v\| = 1$; for the term uv can be written as $(\alpha u)(v/\alpha)$ with any nonzero real number α. Using the sections of f given by $f_s(t) = f(s,t)$, we interpret (12) as

$$\int \|f_s - u(s)v\|^2 \, ds. \tag{13}$$

Since v is fixed and of norm 1, the real number $u(s)$ that minimizes $\|f_s - u(s)v\|$ is the Fourier coefficient of f_s with respect to v :

$$u(s) = \langle f_s, v\rangle = \int f(s,t)v(t)\, dt = (Av)(s). \tag{14}$$

Here A is the integral operator defined by the kernel f; i.e.,

$$(Ax)(s) = \int f(s,t)x(t)\, dt, \qquad x \in L_2(S). \tag{15}$$

When u is defined by (14), the value of the squared error in (13) becomes

$$\int \big[\|f_s\|^2 - \langle f_s, v\rangle^2\big]\, ds = \|f\|^2 - \|Av\|^2. \tag{16}$$

This calculation shows us how to choose v; namely, v must *maximize* the expression $\|Av\|^2$ subject to the constraint that $\|v\| = 1$. Equivalently, we seek v, unconstrained, to maximize

$$\langle Av, Av\rangle / \langle v, v\rangle.$$

This quotient is the same as the Rayleigh quotient of A^*A :

$$\langle A^*Av, v\rangle / \langle v, v\rangle.$$

It is known that this quotient attains its supremum, that the supremum is the leading eigenvalue of A^*A, and that the maximizing vector v is an eigenvector of

A^*A corresponding to this largest eigenvalue. Since A^*A is self-adjoint, positive, and compact, its eigenvalues form a sequence tending downward to zero. We denote this sequence by $\lambda_1 \geq \lambda_2 \geq \cdots$, repeating each eigenvalue a number of times equal to its geometric multiplicity. The corresponding eigenvectors v_i form an orthonormal sequence. Putting $u_i = Av_i$, we can now state Schmidt's Theorem.

THEOREM. *The best L_2-approximation of f by a function of the form $\sum_{i=1}^{n} u_i(s)v_i(t)$, where n is given and $u_i \in L^2(S)$, $v_i \in L^2(T)$, is achieved by letting v_i be eigenvectors of A^*A, as described above, and by putting $u_i = Av_i$.*

Now it must be recognized that Schmidt's solution to the approximation problem is often impractical. Suppose, for example, that we were going to use Schmidt's method to approximate the kernel of an integral equation by a separable kernel, in order to obtain an approximate solution of the integral equation (by the technique outlined previously). In applying Schmidt's method we must find eigenvectors of the operator A^*A; this problem is at least as difficult as the original integral equation!

An alternative approach in the setting of Hilbert space is to start with any orthonormal sequence $[y_1, y_2, \cdots]$ in $L_2(S)$ and to search for an approximation of the form

$$f(s,t) \approx \sum_{i=1}^{n} x_i(s)y_i(t)$$

where the y_i are now prescribed. It is clear that for each s, we must have $x_i(s) = \langle f_s, y_i \rangle = Ay_i$, and so this procedure is quite straightforward. It should be noted, however, that for the *numerical* solution we must address the problem of integration buried in the innocent-looking inner product by which $x_i(s)$ is defined. If this integration is effected with a quadrature rule, it is not clear that we have gained any advantage over the traditional method of introducing a quadrature formula immediately in the integral equation. In this traditional procedure, the integral equation is transformed thereby to an approximate form:

$$x(s) = q(s) + \lambda \sum_{j=1}^{m} w_j f(s,t_j)x(t_j).$$

The weights w_j and the nodes t_j are specified by the quadrature formula. Evaluating both sides of the equation at t_i, we obtain

$$x(t_i) = q(t_i) + \lambda \sum_{j=1}^{m} w_j f(t_i,t_j)x(t_j) \qquad (1 \leq i \leq m).$$

The values of x at the nodes are recovered by solving m linear equations in m unknowns. Intermediate values of $x(t)$ are obtained by interpolation between the nodal values $x(t_i)$.

The problem of approximation in the L_1-norm for the schema

$$f(s,t) \approx \sum_{i=1}^{n} u_i(s)v_i(t)$$

has recently received some attention by Micchelli and Pinkus [1977]. They proved a beautiful theorem that is applicable when f is strictly totally positive. That means that f is continuous and that

$$\det f(s_i,t_j) > 0 \quad \text{if} \quad s_1 < \cdots < s_m \quad \text{and} \quad t_1 < \cdots < t_m.$$

THEOREM. *Let f be strictly totally positive in the square $0 \le s \le 1$, $0 \le t \le 1$. Then the miminum value of*

$$\int_0^1 \int_0^1 \left| f(s,t) - \sum_{i=1}^{n} u_i(s) v_i(t) \right| ds\, dt$$

as u_i and v_i range over $L_1[0,1]$ is attained by functions of the form $u_i(s) = f(s,t_i)$ and $v_i(t) = \sum_{j=1}^{n} c_{ij} f(s_j,t)$.

In the paper cited, Micchelli and Pinkus describe the special points (s_i,t_j) and the coefficients c_{ij}. Their theorem shows, incidentally, that the solution to the problem is continuous in spite of the fact that u_i and v_i are permitted to range over $L_1[0,1]$.

Further references on these best approximation problems are Brown [1982-b] and Deutsch, Mach and Saatkamp [1981].

We turn now to a problem in the scaling (or "preconditioning") of matrices that can be solved by turning it into an approximation problem and applying an effective algorithm. Let A be an $m \times n$ matrix. A **row scaling** of A consists of our dividing each row of A by a positive number:

$$a_{ij} \to a_{ij}/r_i \qquad (1 \le i \le m,\ 1 \le j \le n).$$

This can be effected by multiplying A on the left by the diagonal matrix

$$R = \operatorname{diag}(r_1^{-1}, r_2^{-1}, \cdots, r_m^{-1}).$$

Similarly, a **column scaling** divides the j^{th} column of A by a positive number c_j, for $j = 1, \cdots, n$. This is accomplished by multiplying A on the right by the diagonal matrix

$$C = \operatorname{diag}(c_1^{-1}, c_2^{-1}, \cdots, c_n^{-1}).$$

If both types of scaling are employed, the result is the matrix $B = RAC$ in which $b_{ij} = a_{ij}/r_i c_j$.

In many numerical problems where matrices are employed it is very easy to use B instead of A. For example, in solving a system of linear equations $Ax = u$, the row scaling corresponds to multiplying the i^{th} equation by r_i^{-1}; hence we compensate by multiplying the right-hand side u_i by r_i^{-1} also. The column scaling corresponds to a change of variable, $y_j = c_j x_j$.

In the numerical problems, the purpose of scaling is to improve the conditioning of the computation; hence the term "preconditioning." There are many measures of the conditioning of a matrix B. One of these is the quantity

$$\gamma = \max \left\{ |b_{ij}/b_{\mu\nu}| : b_{\mu\nu} \ne 0 \right\}.$$

If the scaling parameters r_i and c_j are to produce a matrix B with a minimal value of γ, and if this minimum γ is denoted by $e^{2\delta}$, then we are led to demand that

$$e^{-\delta} \le |b_{ij}| \le e^{\delta} \qquad (b_{ij} \ne 0).$$

In terms of the original matrix A, this means that

$$e^{-\delta} \le |a_{ij}/r_i c_j| \le e^{\delta} \qquad (a_{ij} \ne 0).$$

Using overbars to denote natural logarithms of the quantities, we have

$$-\delta \le |\bar{a}_{ij}| - \bar{r}_i - \bar{c}_j \le \delta \qquad (a_{ij} \ne 0).$$

Here we are seeking to approximate the bivariate function $|\bar{a}_{ij}|$ by a sum of two univariate functions $\bar{r}_i$ and $\bar{c}_j$. The appropriate norm is a maximum norm:

$$\left\| |\bar{a}| - \bar{r} - \bar{c} \right\|_D = \max_{(i,j)\in D} \left| |\bar{a}_{ij}| - \bar{r}_i - \bar{c}_j \right|$$

where $D = \{(i,j) : a_{ij} \ne 0\}$.

Von Golitschek [1980-a] has developed an efficient algorithm for solving such problems. It is essentially the discrete version of his algorithm for the approximation of functions on continua. A simple case of this algorithm is described later (Chapter 6). The reader should consult von Golitschek [1980-a, 1980-b, 1980-c], Fulkerson and Wolfe [1962], Bauer [1963, 1969], Rothblum and Schneider [1980, 1982], Osborne [1960], Curtis and Reid [1972], Eaves et al. [1985], von Golitschek, Rothblum, and Schneider [1983], von Golitschek and Schneider [1984], Marshall and Olkin [1968], and Tomlin [1975].

CHAPTER 4

Approximation by Projections

We have alluded previously to the difficulties inherent in finding best approximations, especially in the case of multivariate functions. More of these difficulties will appear in later chapters, and they will give added urgency to the question of whether, in general, we can find "good" approximations with modest effort. In a wide variety of problems the answer is "Yes," and linear projection operators are the principal weapons in our arsenal for solving this problem.

We define a **projection** to be a bounded linear idempotent operator P defined on a normed linear space X. **Idempotent** means that the operator satisfies the equation $P^2 = P$. In order that this equation be meaningful, the range of P must be a subspace of X. Let V be the range of P. If $v \in V$ then $v = Px$ for some $x \in X$, and therefore

$$Pv = P^2x = Px = v.$$

Thus every point of V is a fixed point of P. Conversely if x is a fixed point of P then $x = Px$ and obviously x is in the range, V. To summarize:

$$P(X) = \{x \ : \ Px = x\}.$$

This equation shows that the range of P is automatically closed (a property not possessed by bounded linear operators generally).

A projection $P : X \twoheadrightarrow V$ can be used to provide approximations. The basic inequality governing the error when Px is used as an approximation to x is

$$\|x - Px\| \leq \|I - P\| \operatorname{dist}(x, V). \tag{1}$$

This inequality is proved by letting v be an arbitrary element of V and writing

$$\|x - Px\| = \|(x - v) - P(x - v)\| = \|(I - P)(x - v)\| \leq \|I - P\| \, \|x - v\|.$$

Then we take an infimum as v ranges over V.

An important type of projection is the **Lagrange** interpolation operator. Let S be a compact Hausdorff space, and V an n-dimensional subspace in $C(S)$. Let $s_1, \cdots, s_n$ be n points of S with the property that the set of point evaluation functionals $\{\hat{s}_1, \cdots, \hat{s}_n\}$ is **total** over V. We mean by this that if v is an element of V and $v(s_i) = 0$ for $1 \leq i \leq n$, then $v = 0$. With this assumption in effect, there exist elements $v_1, \cdots, v_n$ in V such that $v_i(s_j) = \delta_{ij}$. Then the "Lagrange" interpolating projection P is defined by

$$Px = \sum_{i=1}^{n} x(s_i) v_i, \qquad x \in C(S). \tag{2}$$

It is clear that $Px \in V$ and that Px interpolates x at $s_1, \cdots, s_n$. That is, $(Px)(s_i) = x(s_i)$. We prefer to write this as $\hat{s}_i(Px) = \hat{s}_i(x)$, or better still as

$$\hat{s}_i \circ P = \hat{s}_i \qquad (1 \leq i \leq n).$$

We shall return to this phenomenon presently.

Recall now the definition of the **adjoint** of a linear operator. If $A : X \to Y$ is a bounded linear operator acting between normed spaces X and Y, then $A^* : Y^* \to X^*$ is the **adjoint** of A, defined by the equation

$$A^*\phi = \phi \circ A, \qquad \phi \in Y^*.$$

This means of course that $(A^*\phi)(x) = \phi(Ax)$ for all $x \in X$ and for all $\phi \in Y^*$. This relationship is more easily remembered in the notation of duality:

$$\langle Ax, \phi \rangle = \langle x, A^*\phi \rangle.$$

Returning to the Lagrange operator, we have observed that

$$Pv = v \quad \text{for all } v \in V,$$

$$P^*\phi = \phi \quad \text{for all } \phi \text{ in the linear span of } \{\hat{s}_1, \cdots, \hat{s}_n\}.$$

Since P^* is also a projection with an n-dimensional range, its range is exactly the linear span of $\{\hat{s}_1, \cdots, \hat{s}_n\}$. This is the subspace of $C(S)^*$ consisting of all functionals which have the form

$$\phi = \sum_{i=1}^{n} \lambda_i \hat{s}_i, \qquad (\lambda_1, \cdots, \lambda_n) \in \mathbb{R}^n$$

where $s_1, \cdots, s_n$ are the points fixed at the beginning of the discussion.

The equation $P^*\phi = \phi$ asserts that the value of the functional ϕ at x is the same as its value at Px. This is the *interpolation* property. The situation that has just been described for the Lagrange operator is true of all projections. Here is the formal result:

THEOREM. *A projection of a normed space onto a subspace is uniquely determined by the range of its adjoint.*

Proof. Suppose that $P : X \twoheadrightarrow V$ and $Q : X \twoheadrightarrow V$ are projections and that $\text{range}\,(P^*) = \text{range}\,(Q^*)$. We show that $P = Q$. Observe that for any x, Qx is in the range of Q, which is the range of P. Therefore Qx is a fixed point of P. Thus $PQx = Qx$ and $PQ = Q$. Similarly, $Q^*P^* = P^*$. Hence

$$Q^* = (PQ)^* = Q^*P^* = P^*.$$

If $x \in X$ and $\phi \in X^*$, then

$$\phi(Px) = (P^*\phi)(x) = (Q^*\phi)(x) = \phi(Qx).$$

Since ϕ and x are arbitrary, $P = Q$. □

As another illustration of this theory, consider an "orthogonal" projection of $L_2[a,b]$ onto an n-dimensional subspace V. Select an orthonormal base $\{v_1, \cdots, v_n\}$ for V. Thus

$$\langle v_i, v_j \rangle = \int_a^b v_i(s) v_j(s)\, ds = \delta_{ij}.$$

The orthogonal projection of $L_2[a,b]$ onto V can be defined by

$$Px = \sum_{i=1}^{n} \langle x, v_i \rangle v_i. \tag{3}$$

It is clear that V is the range of P. The range of P^* is spanned by the n functionals ϕ_i defined by

$$\phi_i(x) = \langle x, v_i \rangle.$$

Each of the foregoing examples is subsumed under the following general considerations. Let V be an n-dimensional subspace of a normed linear space X, and let Φ be an n-dimensional subspace of X^*. If Φ is total over V then there is a unique projection P such that $P(X) = V$ and $P^*(X^*) = \Phi$. The projection P has the interpolation property $\phi(Px) = \phi(x)$ for all $x \in X$ and all $\phi \in \Phi$.

Because of inequality (1), the projections of X onto V which are most useful for approximation are those for which $\|I - P\|$ is small. Since $\|I - P\| \leq 1 + \|P\|$, attention usually focuses on the problem of making $\|P\|$ as small as possible. In the case of finite-dimensional subspaces in $C(S)$, assuming that S is free of isolated points, the minimizations of $\|P\|$ and of $\|I - P\|$ are the same because of Daugavet's Equation:

$$\|I - P\| = 1 + \|P\|. \tag{4}$$

See Daugavet [1963], Cheney and Price [1970], or Foias and Singer [1965]. The same equation is valid for projections of finite rank on the space $L_1[a,b]$. See Babenko and Pichugov [1981]. Equation (4) is not just a curiosity; it implies, for example, that a finite-dimensional proper subspace in $C[a,b]$ cannot possess a linear proximity map.

If a projection P is defined from $C(S)$ onto a subspace V, then there is an elementary procedure for extending P to $C(S \times T)$. The formula for the extended projection $\overline{P}$ is

$$(\overline{P}z)(s,t) = (Pz^t)(s). \tag{5}$$

Here $z \in C(S \times T)$, and its *sections* are defined by

$$z^t(s) = z(s,t) = z_S(t).$$

As an example of how this works, consider the Lagrange interpolating projection from Eq. (2). The extended map is defined in this concrete case by

$$(\overline{P}z)(s,t) = \sum_{i=1}^{n} z(s_i, t) v_i(s). \tag{6}$$

Thus $\overline{P}$ maps $C(S \times T)$ into the subspace of all functions of the form

$$(s,t) \mapsto \sum_{i=1}^{n} y_i(t) v_i(s), \qquad y_i \in C(T).$$

This subspace is properly denoted by $V \otimes C(T)$, and indeed the construction of $\overline{P}$, in either the general formula (5) or the specific case (6), can be best described in the language of tensor products.

Consider then two bounded linear maps $A : X \to X$ and $B : Y \to Y$. There is a natural construction of an operator $A \otimes B$ on $X \otimes Y$. We define

$$(A \otimes B)\Big(\sum_{i=1}^{m} x_i \otimes y_i \Big) = \sum_{i=1}^{m} Ax_i \otimes By_i. \tag{7}$$

This definition is independent of the representation of elements in $X \otimes Y$, as is easily proved. So far, this is purely *algebraic*. If a crossnorm α has been prescribed on $X \otimes Y$, then $X \otimes_\alpha Y$ (being the completion of the normed space) is a Banach space, and we would like to extend $A \otimes B$ to be a continuous linear map of $X \otimes_\alpha Y$ into $X \otimes_\alpha Y$. The existence of such an extension depends upon the boundedness of $A \otimes B$ on the dense set $X \otimes Y$. To make this work smoothly, we assume that α has an additional property, viz.,

$$\alpha\Big(\sum_{i=1}^{m} Ax_i \otimes By_i \Big) \le \|A\| \, \|B\| \alpha\Big(\sum_{i=1}^{m} x_i \otimes y_i \Big) \tag{8}$$

for all operators $A : X \to X$, all operators $B : Y \to Y$, and all elements of $X \otimes Y$. If the crossnorm α has the property (8), it is said to be a **uniform** crossnorm. In the presence of the uniform property of α, it is clear that as an operator on $X \otimes Y$ (with norm α), $A \otimes B$ is bounded (continuous) and

$$\|A \otimes B\| = \sup\big\{ \alpha\big[(A \otimes B)z\big] : z \in X \otimes Y, \, \alpha(z) \le 1 \big\} \le \|A\| \, \|B\|.$$

In fact, equality holds here, as we can see by selecting $x_n \in X$ and $y_n \in Y$ such that $\|x_n\| = \|y_n\| = 1$, $\lim \|Ax_n\| = \|A\|$, and $\lim \|By_n\| = \|B\|$. Then

$$\alpha\big[(A \otimes B)(x_n \otimes y_n)\big] = \alpha(Ax_n \otimes By_n) = \|Ax_n\| \, \|By_n\| \to \|A\| \, \|B\|.$$

Finally, the extension of $A \otimes B$ to a continuous linear map $A \otimes_\alpha B$ on $X \otimes_\alpha Y$ is accomplished with a standard theorem of functional analysis. See Treves [1967].

Returning to the Lagrange projection P and its extension $\overline{P}$ in Eq. (5) we see that $\overline{P} = P \otimes_\lambda I$ because these operators have the same effect on each dyad $x \otimes y$. Indeed,

$$\big[\overline{P}(x \otimes y)\big](s,t) = \big[P(y(t)x)\big](s) = y(t)(Px)(s),$$

$$\big[(P \otimes I)(x \otimes y)\big](s,t) = \big[(Px) \otimes (Iy)\big](s,t) = y(t)(Px)(s).$$

A fact that we do not stop to prove is that λ is a uniform crossnorm. The same is true of γ and β, which were introduced in Chapter 2.

THEOREM. *Let P be a projection of X onto U and let Q be a projection of Y onto V. If α is a uniform crossnorm on $X \otimes Y$ then $P \otimes_\alpha Q$ is a projection of $X \otimes_\alpha Y$ onto $U \otimes_\alpha V$.*

Proof. Clearly $P \otimes Q$ maps $X \otimes Y$ onto $U \otimes V$. Hence $P \otimes_\alpha Q$ maps $X \otimes_\alpha Y$ onto $U \otimes_\alpha V$. The latter is the closure of $U \otimes V$ in $X \otimes_\alpha Y$. If $u \in U$ and $v \in V$ then $(P \otimes Q)(u \otimes v) = u \otimes v$. Hence $P \otimes_\alpha Q$ is a projection onto $U \otimes_\alpha V$. □

This theorem is nicely illustrated by two interpolating projections, $P : C(S) \twoheadrightarrow U$ and $Q : C(T) \twoheadrightarrow V$. If P interpolates functions at nodes $s_1, \cdots, s_n$ and if Q interpolates functions at points $t_1, \ldots, t_m$, then $P \otimes Q$ interpolates (bivariate) functions at all the mesh points (s_i, t_j).

Now we wish to consider another method for combining projections to form new ones. If $P : X \twoheadrightarrow U$ and $Q : X \twoheadrightarrow V$ are projections defined on a linear space X, we construct their Boolean sums by the equations

$$P \oplus Q = P + Q - PQ, \qquad Q \oplus P = Q + P - QP. \tag{9}$$

These are (in general) different operators and are (in general) *not* projections. However, we have this result:

THEOREM. *In order that $P \oplus Q$ be a projection of X onto $U + V$ it is necessary and sufficient that $PQP = QP$.*

Proof. Obviously $P \oplus Q$ maps X into $U + V$. If $v \in V$ then

$$(P \oplus Q)v = Pv + Qv - PQv = Pv + v - Pv = v. \tag{10}$$

If $PQP = QP$ then for any x,

$$(P \oplus Q)Px = P^2x + QPx - PQPx = Px. \tag{11}$$

By (10) and (11), $P \oplus Q$ leaves invariant each element of $U + V$. Conversely, if $P \oplus Q$ leaves invariant each element of U then $QPx = PQPx$ by (11). □

In the literature, the ranges of P and P^* are sometimes referred to as the "function invariance set" and the "interpolation precision set." In applications, one of these sets may be prescribed, and then a projection may be sought conforming to that prescription. As an example of this, suppose that we wish to construct a projection whose interpolation precision set is the vector sum of two subspaces F and G in X^*. The solution is the Boolean sum $P \oplus Q$, provided that $P^*(X^*) = F$, $Q^*(X^*) = G$, and $QPQ = QP$. Many theorems of this type are given in Gordon and Cheney [1978].

To illustrate these methods, let us find a function which interpolates a given function on the perimeter of the unit square in the st-plane. Lagrange interpolation with two nodes, 0 and 1, is given by

$$(Px)(s) = x(0)(1 - s) + x(1)s.$$

Here x is any element of $C(S)$, and $S = \{s : 0 \leq s \leq 1\}$. Similarly, in $C(T)$, with $T = \{t : 0 \leq t \leq 1\}$ we have

$$(Qy)(t) = y(0)(1 - t) + y(1)t.$$

The extensions of these maps to $C(S \times T)$ are then defined by

$$(\overline{P}z)(s,t) = z(0,t)(1-s) + z(1,t)s,$$
$$(\overline{Q}z)(s,t) = z(s,0)(1-t) + z(s,1)t.$$

The projections $\overline{P}$ and $\overline{Q}$ commute with each other, and their Boolean sum provides an interpolant on the perimeter of the square

$$\begin{aligned}\big[(\overline{P} \oplus \overline{Q})z\big](s,t) = z(0,t)(1-s) + z(1,t)s + z(s,0)(1-t) \\ + z(s,1)t - \big[z(0,0)(1-t) + z(0,1)t\big](1-s) \\ - \big[z(1,0)(1-t) + z(1,1)t\big]s.\end{aligned}$$

The operator used here can also be described as $(P \otimes I) \oplus (I \otimes Q)$, where the identity operators are defined on $C(T)$ and $C(S)$ respectively.

This example illustrates one of the general results about projections, namely the following one.

THEOREM. *Let $P : X \twoheadrightarrow U$ and $Q : Y \twoheadrightarrow V$ be projections of normed spaces onto subspaces. For any uniform crossnorm α on $X \otimes Y$, $P \otimes_\alpha I$ commutes with $I \otimes_\alpha Q$. The Boolean sum $(P \otimes_\alpha I) \oplus (I \otimes_\alpha Q)$ is a projection of $X \otimes_\alpha Y$ onto $U \otimes_\alpha Y + X \otimes_\alpha Q$.*

Proof. It is only necessary to prove the commutativity. Testing this on a dyad $z = x \otimes y$, we have

$$(P \otimes I)(I \otimes Q)z = (P \otimes I)(x \otimes Qy) = Px \otimes Qy.$$

The result is the same if we reverse the order of multiplication. Now we appeal to the fact that the dyads form a fundamental set in $X \otimes_\alpha Y$. (That is, their linear combinations are dense in the space.) □

It is an interesting consequence of this theorem that under the hypotheses given, $U \otimes_\alpha Y + X \otimes_\alpha V$ is a closed (even complemented) subspace in $X \otimes_\alpha Y$. The vector sum of two closed subspaces in a Banach space is not closed, in general.

The preceding theorem provides the best practical means for obtaining approximations to elements of $X \otimes_\alpha Y$ by elements of the subspace

$$W = U \otimes_\alpha Y + X \otimes_\alpha V.$$

As has been pointed out previously, projections P having the property $\|I - P\| = 1$ are *ideal* since they produce *best* approximations. However, it is very rare for the proximity map ("best approximation operator") onto a subspace to be linear, and the expectation that $\|I - P\|$ will be close to 1 is usually unrealistic. In Hilbert space, the orthogonal projection onto a subspace is at the same time a proximity map, and $I - P$ is also an orthogonal projection having norm 1. For Boolean sums we have the following very satisfactory theorem which includes the Hilbert space case.

THEOREM. *Let $P : X \twoheadrightarrow U$ and $Q : X \twoheadrightarrow V$ be linear proximity maps mapping a Banach space X onto subspaces U and V. If $PQP = QP$ then $P \oplus Q$ is a linear proximity map of X onto $U + V$.*

Proof. We noted previously that when $PQP = QP$, $P \oplus Q$ will be a projection onto $U + V$. The fact that the Boolean sum is a proximity map follows when we write

$$\|I - (P \oplus Q)\| = \|(I - P)(I - Q)\| \leq \|I - P\| \, \|I - Q\| = 1. \qquad \square$$

For tensor product spaces, we have the following consequence of the preceding results.

THEOREM. *Let $P : X \to\!\!\!\!\to U$ and $Q : Y \to\!\!\!\!\to V$ be linear proximity maps. If α is a uniform crossnorm on $X \otimes Y$, then the projections $P \otimes_\alpha I$, $I \otimes_\alpha Q$, and their Boolean sum are linear proximity maps of $X \otimes_\alpha Y$ onto $U \otimes_\alpha Y$, $X \otimes_\alpha V$, and $U \otimes_\alpha Y + X \otimes_\alpha V$ respectively.* (*These three subspaces are therefore proximinal.*)

Another example of the Boolean sum construction which is important in applications is interpolation on a triangle. The problem is to define an interpolating projection for the perimeter of a triangle. We take a standard triangle as shown in Figure 1.

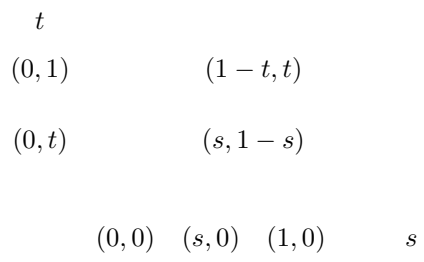

Fig. 1.

We need two operators, which interpolate linearly on horizontal and vertical segments. These are defined as follows:

$$(Pz)(s,t) = z(s,0)\,\frac{1-s-t}{1-s} + z(s,1-s)\,\frac{t}{1-s},$$

$$(Qz)(s,t) = z(0,t)\,\frac{1-t-s}{1-t} + z(1-t,t)\,\frac{s}{1-t}.$$

The function Pz interpolates z on the hypotenuse and horizontal leg of the triangle. The graph of Pz is a ruled surface. Similar remarks apply to Q. One can verify that although $PQ \neq QP$, the weaker condition $QPQ = QP$ holds. Hence by a theorem analogous to the one proved above, $P \oplus Q$ is a projection which accomplishes interpolation on the entire perimeter of the triangle. These ideas originated in Barnhill, Birkhoff, and Gordon [1973]. See also the survey Barnhill [1976]. Projections that interpolate other data (such as derivatives) on the boundary of a triangle can also be found in Barnhill's survey, which lists 50 references on these topics.

The study of *minimal* projections on tensor-product subspaces has only recently begun. A result of Jameson and Pinkus [1983] is the following theorem:

THEOREM. *Let S and T be compact Hausdorff spaces containing infinitely many points. Then the projection constant of $C(S) + C(T)$ in $C(S \times T)$ is* 3.

A minimal projection in this situation can be constructed as follows. Select $(s_0, t_0) \in S \times T$ and define P on $C(S)$ by $Px = x(s_0) \cdot 1$. (Here 1 denotes the unit function.) Similarly, let $Qy = y(t_0) \cdot 1$ for $y \in C(T)$. Then the Boolean sum $(P \otimes I) \oplus (I \otimes Q)$ is a projection of norm 3 from $C(S \times T)$ onto $C(S) + C(T)$.

Franchetti and Cheney [1984] proved this extension:

THEOREM. *If S and T are as before and G and H are finite-dimensional subspaces containing constants in $C(S)$ and $C(T)$ respectively then the projection constant of $G \otimes C(T) + C(S) \otimes H$ in $C(S \times T)$ is at least* 3.

Halton and Light [1985-a] proved this result:

THEOREM. *Let S and T be finite and nonatomic measure spaces. Let G and H be finite-dimensional subspaces containing the constant functions in $L_1(S)$ and $L_1(T)$, respectively. Then the projection constant of $L_1(S) \otimes H + G \otimes L_1(T)$ in $L_1(S \times T)$ is at least* 3.

A minimal projection of $L_p(S \times T)$ onto $L_p(S) + L_p(T)$ can be constructed as follows. Define P on $L_p(S)$ and Q on $L_p(T)$ by

$$Px = \int x(s)ds, \qquad Qy = \int y(t)dt.$$

Then $(P \otimes I) \oplus (I \otimes Q)$ is the minimal projection sought. Here we have assumed for simplicity that S and T have measure 1. This result has been proved by Halton and Light [1985b].

An interesting open problem here is whether spaces of the type

$$W = X \otimes H + G \otimes Y$$

always possess minimal projections from $X \otimes_\alpha Y$, it being assumed that G and H are finite-dimensional subspaces.

For practical purposes it would be very advantageous if the following were true:

CONJECTURE. *For every projection $L : X \otimes_\alpha Y \twoheadrightarrow W$ (as above) there are projections $P : X \twoheadrightarrow G$ and $Q : Y \twoheadrightarrow H$ such that*

$$\|(P \otimes I) \oplus (I \otimes Q)\| \leq \|L\|.$$

Further references on the use of projections to obtain multivariate approximations are Delvos [1975], Delvos and Kösters [1975], Delvos and Malinka [1974], Delvos and Schempp [1983], Gordon [1969-a, 1969-b, 1971], Gordon and Hall [1973], and Haussman and Zeller [1980, 1983, 1984].

CHAPTER 5

The Problem of Proximinality

A subspace U in a Banach space X is said to be **proximinal** if each $x \in X$ possesses at least one best approximation in U. If U is proximinal, then there exists a map $A : X \to U$, called a **proximity** map, such that

$$\|x - Ax\| = \operatorname{dist}(x, U).$$

One must expect in general that A will be nonlinear and discontinuous.

Every finite-dimensional subspace in an arbitrary Banach space is proximinal, and therefore the interesting problems emerge only when one contemplates infinite-dimensional subspaces. In multivariate approximation, the obvious starting point is the investigation of tensor product subspaces. Let us begin with $C(S \times T)$ and consider subspaces of the form $C(S) \otimes V$, in which V is a proximinal subspace of $C(T)$. We ask, is $C(S) \otimes V$ proximinal? The answer is generally "No," even if V is one-dimensional. One hypothesis on V that ensures the proximinality of $C(S) \otimes V$ is that there exist a continuous proximity map $A : C(T) \twoheadrightarrow V$. A more general theorem is true, however, and it reads as follows. (See Franchetti and Cheney [1981-a].)

THEOREM. *Let S be a compact Hausdorff space, and let V be a subspace in a normed space Y. If there is a continuous proximity map from Y onto V, then there exists a continuous proximity map of $C(S, Y)$ onto $C(S, V)$. In particular, the latter is proximinal.*

Proof. If $A : Y \twoheadrightarrow V$ is a continuous proximity map, define $\bar{A} : C(S, Y) \to C(S, V)$ by the equation

$$\bar{A}f = A \circ f.$$

Then for all $s \in S$ and for all $g \in C(S, V)$ we have

$$\|f(s) - (\bar{A}f)(s)\| = \|f(s) - A(f(s))\| \leq \|f(s) - g(s)\|.$$

This shows that for all $g \in C(S, V)$,

$$\|f - \bar{A}f\| \leq \|f - g\|.$$

The continuity of $\bar{A}$ is most easily proved by contradiction. Suppose that a net $f_\alpha \to f$ exists in $C(S, Y)$ with $\|A \circ f_\alpha - A \circ f\| \geq \epsilon$. By compactness we may assume that $s_\alpha \to s$. By the continuity of A at $f(s)$, there is a positive δ such that for $y \in Y$,

$$\|y - f(s)\| < \delta \Longrightarrow \|Ay - A(f(s))\| < \epsilon/2.$$

By the continuity of f at s, there is an index α such that

$$\|f(s_\alpha) - f(s)\| < \delta/2 \quad \text{and} \quad \|f_\alpha - f\| < \delta/2.$$

By the triangle inequality, $\|f_\alpha(s_\alpha) - f(s)\| < \delta$. Therefore

$$\|A(f_\alpha(s_\alpha)) - A(f(s))\| < \epsilon/2 \quad \text{and} \quad \|A(f(s_\alpha)) - A(f(s))\| < \epsilon/2.$$

Hence we arrive at the contradiction $\|A(f_\alpha(s_\alpha)) - A(f(s_\alpha))\| < \epsilon$. □

Because of the isometric isomorphism $C(S, Y) = C(S) \otimes_\lambda Y$, this theorem applies to $C(S) \otimes_\lambda V$ as a subspace of $C(S) \otimes_\lambda Y$. In particular, we have the following important corollary.

THEOREM. *If V is a finite-dimensional Haar subspace in $C(T)$, then there exists a continuous proximity map of $C(S \times T)$ onto $C(S) \otimes V$. The latter is therefore proximinal.*

Proof. Use the preceding theorem and the old result that a finite-dimensional Haar subspace in $C(T)$ has a continuous proximity map. □

It has been recently proved by Li Wu of Hangzhou University (private communication) that the existence of a continuous proximity map from $C(T)$ onto the finite-dimensional subspace V is necessary and sufficient for the proximinality of $C(S) \otimes V$ in $C(S \times T)$. An example showing the non-proximinality of $C(S) \otimes V$ can be found in Franchetti and Cheney [1981-a]. It seems to be an open problem whether the proximinality of $C(S, V)$ in $C(S, Y)$ implies the existence of a continuous proximity map from Y onto V. A theorem having a strong negative character is as follows:

THEOREM. *If S and T are compact intervals in $\mathbb{R}$, then there exists no finite-dimensional subspace $V \subset L_1(T)$ for which $V \neq 0$ and $C(S, V)$ is proximinal in $C(S, L_1(T))$.*

This theorem means, for example, that in general we cannot find best approximations of the form

$$z(s,t) \approx \sum_{i=1}^{n} x_i(s) v_i(t)$$

if $v_1, \cdots, v_n$ are prescribed in $L_1(T)$ and $x_1, \cdots, x_n$ are sought in $C(S)$ so as to minimize the expression

$$\sup_{a \le s \le b} \int_a^b \left| z(s,t) - \sum_{i=1}^{n} x_i(s) v_i(t) \right| dt.$$

In order to recover a positive result in this setting, we confine our attention to continuous functions, as in the following theorem.

THEOREM. *Let S and T be compact real intervals. Let H be an n-dimensional Haar subspace in $C(T)$. Then there exists a continuous map $\bar{A} : C(S \times T) \to C(S) \otimes H$ such that $N(z - \bar{A}z) \le N(z - w)$ for all $z \in C(S \times T)$ and $w \in C(S) \otimes H$. Here N is the norm $N(z) = \sup_{s \in S} \int_T |z(s,t)|\, dt$.*

Proof. By a theorem of Dunham Jackson (Singer [1970, p. 236] or Cheney [1965]), each element of $C(T)$ has a unique best L_1-approximation in H. By a general theorem in approximation theory (Singer [1970, p. 251]), the L_1-proximity map $A : C(T) \to H$ is continuous in the topology of the L_1-norm in $C(T)$. Since the supremum norm is stronger than the L_1-norm on $C(T)$ and equivalent to the L_1-norm on H, A is continuous in the supremum norm. The extended map $\bar{A} : C(S \times T) \to C(S) \otimes H$ defined by $(\bar{A}z)(s,t) = (Az_S)(t)$ is continuous by arguments used above, and produces the approximations as described. □

The preceding theorem guarantees the existence of a solution to the approximation problem of minimizing

$$\sup_{s\in S}\int_T \left| z(s,t) - \sum_{i=1}^{n} h_i(t)x_i(s)\right| dt$$

by an appropriate choice of $x_1, \cdots, x_n$ in $C(S)$, provided that $z \in C(S \times T)$ and $h_i \in C(T)$. A more general result along the same lines is proved in the same way.

THEOREM. *Let S and T be compact Hausdorff spaces. Let Δ be a norm on $C(T)$ such that $\Delta(x) \leq \|x\|_\infty$ for all $x \in C(T)$. Let H be a finite-dimensional subspace of $C(T)$ which has the Chebyshev property with respect to the norm Δ. Then there is a mapping $\bar{A} : C(S \times T) \to\!\!\!\!\to C(S) \otimes H$ which is* sup-norm *continuous and is a proximity map for the norm*

$$\|z\|_\Delta = \sup_{s\in S} \Delta(z_s), \qquad z \in C(S \times T).$$

Further references on approximation in spaces $C(S, X)$ are Buck [1974], Amir and Deutsch [1979], Singer [1970, 1974], and Brown [1982-a].

An approximating subspace in $C(S \times T)$ recommended to us by the study of projections is of the form

$$W = C(S) \otimes H + G \otimes C(T) \tag{1}$$

in which H and G are finite-dimensional subspaces in $C(T)$ and $C(S)$ respectively. We saw in Chapter 4 that it is easy to construct projections onto W starting with projections $P : C(S) \to\!\!\!\!\to G$ and $Q : C(T) \to\!\!\!\!\to H$. The Boolean sum projection constructed for W has norm no greater than

$$\|P\| + \|Q\| + \|P\|\,\|Q\|.$$

If this bound is small, then the problem of finding *good* approximations in W is solved. The more difficult problem of the existence of *best* approximations has not yet been completely solved. The most satisfactory result to date is the following sufficient condition given by Respess and Cheney [1982-a]. (It is known as the "Sitting Duck Theorem.")

THEOREM. *If there exist a continuous proximity map of $C(S)$ onto G and a Lipschitzian proximity map of $C(T)$ onto H, then the subspace W in Eq.* (1) *is proximinal in $C(S \times T)$.*

The first interesting case of this type of approximation was studied by Diliberto and Straus [1951]. They considered

$$W_0 = C(S) + C(T).$$

More properly, we should write

$$W_0 = C(S) \otimes \Pi_0(T) + \Pi_0(S) \otimes C(T)$$

where Π_0 denotes the subspace of constant functions on the given domain. This problem is of special importance because it illustrates the simplest nontrivial case

of proximinality and because a completely constructive proof of proximinality is possible by means of the algorithm which Diliberto and Straus devised. (See the next chapter for a discussion of this algorithm.) We note also that the elements of W are basic functions from which *nomographic* functions

$$(s,t) \mapsto h\big(x(s)+y(t)\big), \qquad h \in C(\mathbb{R})$$

are constructed.

If we pass now to more general subspaces W, the first case in which proximinality is still an open question is

$$W_1 = C(S) \otimes \Pi_1(T) + \Pi_1(S) \otimes C(T).$$

The elements of W are then pseudopolynomials of the form

$$w(s,t) = x_0(s) + x_1(s)t + y_0(t) + y_1(t)s.$$

It is not known whether this subspace is proximinal in $C(S \times T)$, if we take S and T to be compact real intervals. What *is* relatively easy to prove is that the subspace

$$\ell_\infty(S) \otimes G + H \otimes \ell_\infty(T)$$

is proximinal in $\ell_\infty(S \times T)$ whenever G and H are finite-dimensional subspaces in $\ell_\infty(S)$ and $\ell_\infty(T)$, respectively. The space $\ell_\infty(S)$ is the Banach space of all bounded functions on the set S, normed by $\|x\| = \sup_{s \in S} |x(s)|$. This theorem can be found in von Golitschek and Cheney [1983-a]. A consequence of it is that each $z \in C(S \times T)$ has a best approximation in

$$\ell_\infty(S) \otimes \Pi_m(T) + \Pi_n(S) \otimes \ell_\infty(T).$$

Among these best approximations one would hope to find one that is continuous, but it is not known whether this is possible, even when $m = n = 1$. A positive result from the paper just cited states that among the best approximations to z in

$$\ell_\infty(S) \otimes \Pi_1(T) + \Pi_1(S) \otimes \ell_\infty(T),$$

there exists one that is continuous on the *interior* of $S \times T$. A review of the work in this field would strongly suggest a conjecture that the subspace W_1 defined above is *not* proximinal in $C(S \times T)$.

For practical purposes, the following theorem gives important information about these problems.

THEOREM. *Let S and T be compact spaces, and let G and H be finite-dimensional subspaces in $C(S)$ and $C(T)$, respectively. For any $f \in C(S \times T)$, the distances to*

$$G \otimes \ell_\infty(T) + \ell_\infty(S) \otimes H \quad \text{and} \quad G \otimes C(T) + C(S) \otimes H$$

are the same.

Thus, no improvement of the accuracy in this approximation is obtained by allowing discontinuous coefficient functions, if the function being approximated is continuous.

The proof of the theorem just cited is an easy consequence of the following lemma from von Golitschek and Cheney [1983-a].

LEMMA. *Let S be a paracompact space, and H a closed subspace in a Banach space X. For any $f \in C(S, X)$,*

$$\operatorname{dist}(f, C(S, H)) = \operatorname{dist}(f, \ell_\infty(S, H)).$$

Proof. The space $\ell_\infty(S, H)$ consists of all maps $u : S \to H$ for which

$$\|u\| = \sup_s \|u(x)\| < \infty.$$

Since $C(S, H) \subset \ell_\infty(S, H)$, the inequality

$$\operatorname{dist}(f, \ell_\infty(S, H)) \leq \operatorname{dist}(f, C(S, H))$$

is obvious. Now let $\rho > \operatorname{dist}(f, \ell_\infty(S, H))$. We will find $u \in C(S, H)$ such that $\|f - u\| \leq \rho$. For each $s \in S$, define

$$\Phi(s) = \big\{h \in H : \|f(s) - h\| \leq \rho\big\}.$$

It is clear that $\Phi(s)$ is nonvoid, closed, and convex in H. It will be shown that the set-valued map Φ is lower semicontinuous. To this end, let $\mathcal{O}$ be an open set in H, and put

$$\mathcal{O}^* = \{s \in S : \Phi(s) \cap \mathcal{O} \neq \square\}.$$

In order to prove that $\mathcal{O}^*$ is open, let $s \in \mathcal{O}^*$. Then there is an $h \in \mathcal{O}$ such that $\|f(s) - h\| \leq \rho$. By the definition of ρ there exists $v \in \ell_\infty(S, H)$ such that $\|f - v\| < \rho$. By convexity, there is a point h' in $\mathcal{O}$ (on the line segment joining h to $v(s)$) such that $\|f(s) - h'\| < \rho$. Since f is continuous, there is a neighborhood $\mathcal{N}$ of s such that $\|f(\sigma) - f(s)\| < \rho - \|f(s) - h'\|$ for all $\sigma \in \mathcal{N}$. Then $\|f(\sigma) - h'\| < \rho$ and $h' \in \Phi(\sigma)$ when $\sigma \in \mathcal{N}$. Hence $\sigma \in \mathcal{O}^*$, $\mathcal{N} \subset \mathcal{O}^*$, and $\mathcal{O}^*$ is open. It follows that Φ is lower semicontinuous. By the Michael Selection Theorem (Holmes [1972]), there is a continuous map $u : S \to H$ for which $u(s) \in \Phi(s)$ for all s. Thus $\|f(s) - u(s)\| \leq \rho$. Since u is obviously bounded, $u \in C(S, H)$. □

Presumably the existence of best approximations in a subspace such as W_1 can be proved if the function being approximated is subject to stronger hypotheses than that of simple continuity. This idea has yet to be fully explored.

Proximinality questions like the ones just raised are interesting in spaces other than $C(S \times T)$. As is well known, proximinality of closed subspaces becomes a triviality in the presence of uniform convexity. Thus in $L_p(S \times T)$, with $1 < p < \infty$, the subspace

$$L_p(S) \otimes H + G \otimes L_p(T)$$

will be proximinal provided that there is a uniform crossnorm α_p for which $L_p(S) \otimes_{\alpha_p} L_p(T) = L_p(S \times T)$. (The uniform crossnorm is needed in order to prove that the subspace in question is closed.) Fortunately, there exist such crossnorms. They are called the "p-nuclear" norms, and their practical properties were developed by Saphar [1970], Chevet [1969], and Persson [1969]. See Light and Cheney [1985] for the parts of this theory needed in these approximation problems.

Another space of special importance in approximation theory is $L_1(S\times T)$. As a sample of the proximinality theorems involving this space, we give a result of Light and Cheney [1981]. For more general results, particularly one of Khalil [1983], see Chapter 2 of Light and Cheney [1985].

THEOREM. *If S and T are σ-finite measure spaces, and if H is a finite-dimensional subspace of $L_1(T)$, then $L_1(S)\otimes H$ is proximinal in $L_1(S\times T)$.*

Proof. Let $f\in L_1(S\times T)$. By the Fubini Theorem, $f_s\in L_1(T)$ for almost all $s\in S$. We redefine f on a set of measure 0 so that $f_s\in L_1(T)$ for all s. Define

$$\Phi(s)=\big\{h\in H:\|f_s-h\|_1=\operatorname{dist}_1(f_s,H)\big\}.$$

Clearly $\Phi(s)$ is a nonempty closed subset of H. In order to use the measurable selection theorem, we must prove that Φ is measurable, as a set-valued map. This means that for any compact set K in H, the set

$$K^*=\{s\in S:\Phi(s)\cap K\neq\square\}$$

should be measurable in S. Another form for K^* is

$$K^*=\big\{s\in S:\inf_{h\in K}\|f_s-h\|_1=\operatorname{dist}_1(f_s,H)\big\}.$$

Thus the measurability of K^* will follow from the measurability of all mappings of the following type, where $B\subset H$:

$$s\mapsto\inf_{h\in B}\|f_s-h\|_1.$$

To prove that this is measurable, select a countable dense set $[h_i]$ in B. Then

$$\inf_{h\in B}\|f_s-h\|_1=\inf_i\|f_s-h_i\|_1.$$

The map is measurable since it is the infimum of a countable family of measurable maps. By the measurable selection theorem quoted below, there exist measurable maps $x_j:S\to\mathbb{R}$, $j=1,\cdots,n$ such that $\sum_{j=1}^n x_j(s)h_j$ is a best L_1-approximation to f_s in H for each s. Here $\{h_1,\cdots,h_n\}$ is any basis for H. In order to verify that $x_j\in L_1(S)$, select elements ψ_i in $L_\infty(T)$ such that $\int\psi_i(t)h_j(t)dt=\delta_{ij}$. Put $v(s,t)=\sum_{j=1}^n x_j(s)h_j(t)$. Then

$$\|x_i\|_1=\int|x_i(s)|ds=\int\left|\int v(s,t)\psi_i(t)dt\right|ds\le\|v\|_1\,\|\psi_i\|_\infty.$$

Thus $v\in L_1(S)\otimes H$, and we have for any $u\in L_1(S)\otimes H$,

$$\begin{aligned}\|f-v\|_1&=\iint|f(s,t)-v(s,t)|ds\,dt\\&=\int\|f_s-v_s\|_1\,dt\\&\le\int\|f_s-u_s\|_1\,dt=\|f-u\|_1.\end{aligned}$$

□

MEASURABLE SELECTION THEOREM. *If Φ is a measurable set-valued map of a measurable space S into the family of closed nonvoid subsets of a finite-dimensional Banach space X, then there exists a function $f : S \to X$ such that $f(s) \in \Phi(s)$ for all $s \in S$ and $f^{-1}(\mathcal{O})$ is measurable for each open $\mathcal{O}$ in X.* (See Parthasarathy [1972] and references cited therein.)

We state without proof some further results on proximinality. The first three can be found in Light and Cheney [1985].

THEOREM. *Let S be a measure space of finite measure. Let H be a reflexive subspace in a Banach space X. Then $L_1(S, H)$ is proximinal in $L_1(S, X)$.*

This is the theorem of Khalil previously alluded to. The space $L_1(S, X)$ consists of all Bochner integrable maps from S to X. The elementary theory of these spaces is developed in Chapter 10 of Light and Cheney [1985]. The book by Diestel and Uhl [1977] gives more complete information.

THEOREM. *Let S and T be finite measure spaces. Let G and H be finite-dimensional subspaces in $L_1(S)$ and $L_1(T)$, respectively. Then $G \otimes L_1(T) + L_1(S) \otimes H$ is proximinal in $L_1(S \times T)$.*

THEOREM. *Let S and T be σ-finite measure spaces. Let G and H be finite-dimensional subspaces in $L_\infty(S)$ and $L_\infty(T)$, respectively. Then $G \otimes L_\infty(T) + L_\infty(S) \otimes H$ is proximinal in $L_\infty(S \times T)$.*

The following theorem is due to Darst, Legg, and Townsend [1983]. Theorems of this type are applicable in certain problems of prediction theory.

THEOREM. *Let $(S, \mathcal{A}, \mu)$ be a finite measure space, and let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Let X be any uniformly convex space. Then $L_\infty(S, \mathcal{B}, \mu, X)$ is proximinal in $L_\infty(S, \mathcal{A}, \mu, X)$.*

CHAPTER 6
Algorithms

In this subject, we mean by an **algorithm** any procedure or "recipe" for producing approximations. In this broad sense, Lagrange interpolation, Shepard interpolation, and orthogonal expansions are algorithms. Indeed, any projection operator can be regarded as an algorithm. Projections, of course, do not generally produce *best* approximations; algorithms for the latter are usually more difficult to invent, and are less efficient computationally.

At the practical level, an algorithm must be judged by such criteria as ease of programming, speed of computation, stability and accuracy of computation. Two algorithms that are equally attractive in theory may differ markedly in their numerical properties. For an example of this, consider two projections into Π_n, one given by Lagrange interpolation and another given by an orthogonal expansion:

$$Px = \sum_{i=0}^{n} x(s_i)\ell_i, \qquad \ell_i(s_j) = \delta_{ij}, \tag{1}$$

$$Qx = \sum_{i=0}^{n} \langle x, u_i \rangle u_i, \qquad \langle u_i, u_j \rangle = \delta_{ij}. \tag{2}$$

Theoretically, these are quite similar: each involves a basis for Π_n and an accompanying system of "biorthogonal" functionals. Notice, however, that a functional of the type $x \mapsto \langle x, u_i \rangle$ is much more difficult to compute than one of the type $x \mapsto x(s_i)$, since the former involves an integration. Numerically such an integration will be effected with a quadrature formula, which, in turn, will be based upon functionals of the form $x \mapsto x(s_i)$.

Algorithms for best approximation will usually be iterative in nature, and each step in the procedure will often involve roughly as much work as a single linear projection process. Thus, the Remez Second Algorithm can be interpreted as a sequence of Lagrange interpolation processes, or as a sequence of linear projections similar to ordinary interpolation. The major exception to these general observations occurs in the case of quadratic norms – the Hilbert-space case. Here, the best approximation operator onto a linear subspace is a linear projection. In the finite-dimensional case, it is of the type illustrated by the operator Q in Eq. (2).

Let us proceed now to subspaces in tensor product spaces. We have already seen that, for producing best approximations in a subspace $C(S) \otimes H$, when H is a subspace of $C(T)$, there will be no theoretical difficulties if H is a finite-dimensional *Haar* subspace (see Chapter 5). Thus, if A is the proximity map of $C(T)$ onto H, then one proximity map of $C(S \times T)$ onto $C(S) \otimes H$ is defined by

$$(\bar{A}f)(s,t) = (Af_s)(t).$$

In terms of a basis $\{h_1, \cdots, h_m\}$ for H, $\bar{A}$ is given by

$$(\bar{A}f)(s,t) = \sum_{j=1}^{m} x_j(s)\, h_j(t)$$

where, for each s, $\sum_{j=1}^{m} x_j(s)h_j$ is the unique best approximation of f_S in H. The theory assures us that the functions x_j determined point-by-point in this manner are continuous.

In a practical realization of this algorithm, one could select a discrete subset $\{s_1, \cdots, s_N\}$ in S, and determine the functions x_j only on this discrete set. Then, by a spline-interpolation procedure, the x_j-functions could be extended to all of S. The resulting function will not be a best approximation, but a near substitute for it. If this strategy has been adopted, then the resulting approximation will lie in a subspace $G \otimes H$, where G is now a finite-dimensional subspace of splines having knots $s_1, \cdots, s_N$. In this roundabout way, then, we are led again to consider the general question of best approximation in $C(S \times T)$ by elements of a prescribed subspace $G \otimes H$, both G and H being finite-dimensional. Since $G \otimes H$ is also finite-dimensional, the exchange algorithm or the Remez First Algorithm can be used.

It should be noted that the operator $\bar{A}$ defined in the previous paragraphs is not the only proximity map of $C(S \times T)$ onto $C(S) \otimes H$. The subspace $C(S) \otimes H$ never has the Chebyshev property (except in degenerate cases), and so there is much latitude in the choice of best approximations. The map $\bar{A}$ selects the optimal one of these, since it minimizes the norm of each section $f_S - (Af)_S$. It is not necessary to do this. The theorem governing this situation is as follows (Franchetti and Cheney [1981-a]).

THEOREM. *Let H be a subspace of a Banach space Y, f an element of $C(S,Y)$, and g an element of $C(S,H)$. In order that g be a best approximation of f in $C(S,H)$ it is necessary and sufficient that there exist a point s_0 such that*

$$\|f - g\| = \operatorname{dist}\big(f(s_0), H\big).$$

Proof. A lemma of Buck [1974] asserts that

$$\operatorname{dist}\big(f, C(S,H)\big) = \sup_{s} \operatorname{dist}\big(f(s), H\big). \tag{3}$$

Let s_0 be a point of S such that

$$\operatorname{dist}\big(f(s_0), H\big) = \operatorname{dist}\big(f, C(S,H)\big).$$

If g is a best approximation of f then

$$\|f - g\| = \operatorname{dist}\big(f, C(S,H)\big) = \operatorname{dist}\big(f(s_0), H\big).$$

For the converse, let σ be a point such that $\|f - g\| = \operatorname{dist}\big(f(\sigma), H\big)$. By (3), $\|f - g\| \leq \operatorname{dist}\big(f, C(S,H)\big)$. □

It is clear that if g and s_0 have the property

$$\|f - g\| = \operatorname{dist}\big(f(s_0), H\big),$$

then $g(s_0)$ is a best approximation of $f(s_0)$ in H. Thus, in principle, there need be only one s_0 where $g(s_0)$ is a solution to an extremal problem. For all the other points it is merely necessary to secure the weaker requirement

$$\|f(s) - g(s)\| \leq \|f(s_0) - g(s_0)\|.$$

For the more interesting and versatile subspaces

$$C(S) \otimes H + G \otimes C(T),$$

there exist no general algorithms for producing minimizing sequences. The sole exception to this is the Diliberto-Straus algorithm for producing best approximations in

$$W_0 = C(S) + C(T).$$

Their algorithm proceeds as follows. Let f be an element of $C(S \times T)$ whose best approximation in $C(S)+C(T)$ is sought. A process of "alternating proximity maps" is used. Observe that we can construct best approximations of f in $C(S)$ and $C(T)$ by these processes:

$$(Af)(s,t) = \tfrac{1}{2} \max_{\sigma \in S} f(\sigma, t) + \tfrac{1}{2} \min_{\sigma \in S} f(\sigma, t), \tag{4}$$

$$(Bf)(s,t) = \tfrac{1}{2} \max_{\tau \in T} f(s, \tau) + \tfrac{1}{2} \min_{\tau \in T} f(s, \tau). \tag{5}$$

It is clear that Af is really a function of t alone, or, more precisely, is constant in s. Thus $Af \in C(T)$. Similarly, $Bf \in C(S)$. The alternating process constructs the sequence

$$f_0 = f, \quad f_{2n+1} = f_{2n} - Af_{2n}, \quad f_{2n+2} = f_{2n+1} - Bf_{2n+1}. \tag{6}$$

By successively subtracting from f certain elements of $C(S)$ and $C(T)$ we can expect to reduce the norm of the residual functions f_n to an absolute minimum. The first significant result in the theory asserts that, indeed,

$$\lim_{n \to \infty} \|f_n\| = \operatorname{dist}(f, W_0).$$

This was proved already by Diliberto and Straus. The second significant result was proved by Aumann [1958], who was apparently unaware of Diliberto and Straus' work. He established the convergence of the sequence $[f_n]$. Since the definition of the algorithm ensures that $f - f_n \in W_0$ for all n, and since W_0 is closed, it follows that $f - \lim f_n$ is a best approximation of f in W_0. The algorithm therefore provides a constructive proof that W_0 is proximinal.

Golomb [1959] noticed that the theory of Diliberto and Straus could be applied to any pair of proximity maps A and B in any normed space provided that a certain property called "centrality" was possessed by A and B. That property, for A, is expressed by the equation

$$\|x - Ax + Ay\| = \|x - Ax - Ay\|,$$

assumed valid for all x and y. The particular maps A and B in Eq. (4) and (5) are indeed central proximity maps. Also, any orthogonal projection on an inner-product space is a central proximity map. If a central proximity map $B : Y \twoheadrightarrow H$ is given, then a central proximity map $\bar{B} : C(S, Y) \twoheadrightarrow C(S, H)$ is constructed by the definition

$$\bar{B}f = B \circ f.$$

This construction is, in fact, used in Eqs. (4) and (5).

The main theorem in Golomb's theory states that if $A : X \twoheadrightarrow U$ and $B : X \twoheadrightarrow V$ are central proximity maps and if $U + V$ is closed, then the Diliberto-Straus algorithm given in Eq. (6) produces a sequence $[f_n]$ such that

$$\|f_n\| \downarrow \operatorname{dist}(f, U + V).$$

This is valid in any Banach space X. In order to draw the conclusion that the sequence $[f_n]$ converges, one can add uniform convexity of X as a hypothesis. These matters are expounded in Chapter 4 of Light and Cheney [1985].

A paper by Light [1980] reveals that central proximity maps are rare. For example, a Haar subspace of dimension 2 or more in $C[a, b]$ does not have a central proximity map. The constant functions in $L_p[a, b]$ have a central proximity map only if $p = 2$.

Further negative results were established by Dyn [1980] and by von Golitschek and Cheney [1983-b]. The work of these three authors shows that if G and H are Haar subspaces in $C(I)$ with $I = [a, b]$, $\dim G \geq 1$, and $\dim H \geq 2$, then the Diliberto-Straus algorithm fails when applied to the subspaces $G \otimes C(I)$ and $H \otimes C(I)$. This means that the sequence $[f_n]$ obtained by alternately subtracting best approximations may have the property

$$\lim \|f_n\| > \operatorname{dist}(f, W)$$

where W is the vector sum of $G \otimes C(I)$ and $C(I) \otimes H$.

If the topological space T is a disjoint union of m closed sets, then the functions which are constant on these closed sets form an m-dimensional subspace H in $C(T)$. One can visualize this as a subspace of 0-degree spline functions. Such a subspace possesses a central proximity map. Figure 2 shows how this map is defined.

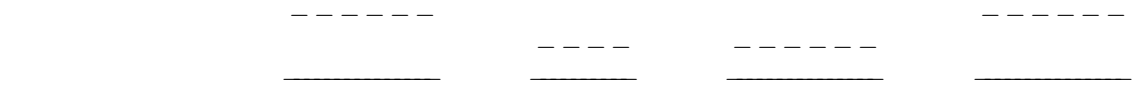

Fig. 2.

If S is also disconnected and has, say, n components, then a similar subspace G of dimension n exists in $C(S)$. The generalized Diliberto-Straus algorithm now is completely successful. With each $f \in C(S \times T)$ we associate a convergent sequence $[f_k]$ such that $f - \lim f_k$ is a best approximation of f in

$$Z = C(S) \otimes H + G \otimes C(T).$$

Therefore the algorithm gives a constructive proof of proximinality of Z. These matters have been dealt with by Respess and Cheney [1982-b].

Because of its natural and elegant character, the algorithm we have called by the names of Diliberto and Straus has been used in many diverse situations. Its use in Hilbert space predates the work of Diliberto and Straus by almost 20 years, having been exploited by von Neumann to obtain the orthogonal projection onto the closure of the sum of two subspaces. Von Neumann proved that if P and Q are orthogonal projections of a Hilbert space onto subspaces U and V, respectively, then the projection of an element f onto $\overline{U+V}$ is given by $f - \lim f_n$, where $[f_n]$ comes from the algorithm. In this situation, the algorithm has been known as the **alternating algorithm.** The formulae are

$$f_0 = f, \quad f_{2n+1} = f_{2n} - Pf_{2n}, \quad f_{2n+2} = f_{2n+1} - Qf_{2n+1}.$$

These are, of course, the same as those used in the Diliberto-Straus algorithm. In Hilbert space, the algorithm is effective for any pair of closed subspaces, while in $C(S)$ or $C(S \times T)$ it is effective only for very special pairs of subspaces. More recent work by Sullivan [1975], Atlestam and Sullivan [1976], Deutsch [1984], and Franchetti and Light [1984] has led to various generalizations of von Neumann's theorem, such as this one:

THEOREM. *Let X and its conjugate be uniformly convex Banach spaces. If U, V and $U+V$ are closed subspaces in X, then the alternating algorithm produces best approximations in $U+V$.*

A new algorithm having many favorable features has recently been developed by von Golitschek [1984]. His algorithm finds best approximations to elements of $C(D)$, $D \subset S \times T$, by functions of the form

$$z(s,t) = \phi\big[x(s)h(t) + y(t)g(s)\big] \tag{7}$$

in which ϕ, g, and h are prescribed continuous functions, and the functions x and y are at our disposal. It is assumed that

$$g \in C(S), \quad h \in C(T), \quad \phi \in C(\mathbb{R}), \quad g > 0, \quad h > 0, \quad \text{and} \quad \phi^{-1} \in C(\mathbb{R}).$$

We will outline von Golitschek's results as they apply to the simpler problem

$$f(s,t) \approx x(s) + y(t),$$

for in this case the analogy to the Diliberto-Straus algorithm will be more apparent. This procedure is iterative and uses the following formulae, in which α is a parameter lying in the interval $0 \le \alpha \le \|f\|$. Ideally, we would set $\alpha = \operatorname{dist}(f, W_0)$, where $W_0 = C(S) + C(T)$.

$$\begin{cases} x_0(s) = 0 & y_0(t) = \inf\limits_s \big[f(s,t) + \alpha\big] \\ x_n(s) = x_{n+1}(s) \vee \sup\limits_t \big[f(s,t) - \alpha - y_{n-1}(t)\big]. & \text{If } \|x_n\| > 4\|f\|, \text{ STOP.} \\ y_n(t) = y_{n-1}(t) \wedge \inf\limits_s \big[f(s,t) + \alpha - x_n(s)\big]. & \text{If } y_n = y_{n-1}, \text{ STOP.} \end{cases}$$

In these formulae $\vee$ and $\wedge$ denote the larger and the smaller of two numbers. It is already clear from these formulae that

$$0 = x_0 \leq x_1 \leq x_2 \leq \cdots \quad \text{and} \quad y_0 \geq y_1 \geq y_2 \geq \cdots .$$

Another favorable property of the functions x_n and y_n is that they have the same modulus of continuity as f; thus the sequences $[x_n]$ and $[y_n]$ are equicontinuous. The *uniform convergence* of these sequences will therefore depend solely upon whether they are bounded in $C(S)$ and $C(T)$. Here are the three principal results concerning the algorithm:

THEOREM 1. *If the inequality* $\alpha > \operatorname{dist}(f, W_0)$ *is true then the iteration terminates with* $y_n = y_{n-1}$ *for some* n. *When this occurs,* $\|f - x_n - y_n\| \leq \alpha$.

THEOREM 2. *The inequality* $\alpha < \operatorname{dist}(f, W_0)$ *is true if and only if the iteration terminates with* $\|x_n\| > 4\|f\|$ *for some* n.

THEOREM 3. *If* $\alpha = \operatorname{dist}(f, W_0)$, *then either the algorithm terminates as in Theorem 1 or else it produces a sequence* $[x_n + y_n]$ *that converges uniformly to a best approximation of* f *in* W_0.

CHAPTER 7

Simultaneous Approximation and Chebyshev Centers

Simultaneous approximation is concerned with the optimal approximation of *sets* of functions, rather than just single functions. In the literature, the problem of approximating just two functions simultaneously has been considered at length, and even this case is of great interest. In fact, the approxmation of a set of elements can sometimes be reduced to the approximation of a pair of elements, as will be seen later.

As before, we begin with a normed linear space X and a subspace U of "approximants." (In the problem now to be described, the case $U = X$ is interesting and important.) Let a subset K be given in X. We desire to approximate all the elements of K simultaneously by a single element of U. For any $u \in U$, the quality of this approximation is measured by $\sup_{x\in K} \|x - u\|$. The optimization problem is then to select $u \in U$ so as to make this deviation as small as possible. The quantity

$$\inf_{u\in U} \sup_{x\in K} \|x - u\|$$

measures the best that we can hope to achieve; it is called the **Chebyshev radius of K relative to** U, and is denoted by $r(K;U)$. We see at once that the Chebyshev radius of K relative to U is the infimum of the radii of all the cells with centers in U that contain K.

The existence problem for optimal simultaneous approximations has received much attention, starting with Garkavi [1961]. The (possibly empty) set of solutions to our problem is

$$E(K;U) = \big\{u \in U : \sup_{x\in K} \|x - u\| = r(K;U)\big\}.$$

This is called the **Chebyshev center of K relative to** U. The *existence* question is now whether $E(K;U)$ is empty or not.

In the theory of simultaneous approximation, the case when $U = X$ is of fundamental importance. In this special situation, the simpler notation $r(K)$, $E(K)$ is used. These are called the **Chebyshev radius** and the **Chebyshev center** of K, respectively.

As a sample of what can be proved without much effort, we cite the following theorem:

THEOREM 1. *The Chebyshev center of a compact set K in a space $C(S)$ is the same as the Chebyshev center of a pair of functions, namely, $v_1(s) = \max_{x\in K} x(s)$ and $v_2(s) = \min_{x\in K} x(s)$.*

Since the Chebyshev center of this pair $\{v_1, v_2\}$ contains their average, we see that the center of K is the set of all $u \in C(S)$ that satisfy

$$\max\{\|u - v_1\|, \|u - v_2\|\} \le \tfrac{1}{2}\|v_1 - v_2\|.$$

This condition, in turn, can be simplified to read

$$v_1 - r \le u \le v_2 + r, \qquad u \in C(S) \tag{1}$$

where r is the Chebyshev radius and equals $\frac{1}{2}\|v_1 - v_2\|$.

For a *bounded* subset B in $C(S)$ similar results are known. The definitions of v_1 and v_2 must be modified to read

$$\begin{aligned} v_1(s) &= \inf_{\mathcal{N}} \sup_{t\in\mathcal{N}} \sup_{x\in B} x(t), \\ v_2(s) &= \sup_{\mathcal{N}} \inf_{t\in\mathcal{N}} \inf_{x\in B} x(t), \end{aligned} \tag{2}$$

in which $\mathcal{N}$ runs over all the neighborhoods of s. The functions v_1 and v_2 are upper and lower semicontinuous, respectively. The reader is referred to Holmes [1972, p. 186], and Franchetti and Cheney [1981-b] for the following theorem, in which $C(S)$ denotes the space of all bounded continuous real-valued functions on an arbitrary topological space S.

THEOREM 2. *The Chebyshev center of a bounded set in $C(S)$ is nonempty. If S is a normal space, the center of a bounded set B is defined by inequality* (1) *above, with v_1 and v_2 defined in* (2).

The proof is an interesting application of the Hahn-Tong Theorem, which states that if f_1 and f_2 are, respectively, upper semicontinuous and lower semicontinuous real-valued functions on a normal topological space, and if $f_1 \le f_2$, then there is a continuous f between f_1 and f_2. See Semadeni [1971, p. 100]. Now, letting $f_1 = v_1 - r$ and $f_2 = v_2 + r$, we conclude that inequality (1) is valid for some continuous u. The assertion that $E(B)$ is nonempty whenever B is bounded in $C(S)$ can thus be proved when S is normal. Then one uses Kakutani's Theorem which asserts that every abstract M-space with unit is (isometric to) a $C(T)$ with T compact and Hausdorff. Hence any Banach space property of spaces $C(T)$, with T compact Hausdorff, is shared by all abstract M-spaces, in particular, spaces $C(S)$ where S is not compact.

Now we consider one source of simultaneous approximation problems. Let $z \in C(S \times T)$. We ask, "How well can z be approximated by an element of $C(S)$?" If $x \in C(S)$ then the deviation between z and x is

$$\|z - x\| = \sup_{t\in T} \sup_{s\in S} |z(s,t) - x(s)| = \sup_t \|z^t - x\|.$$

The notation z^t indicates the element of $C(S)$ whose value at s is $z(s,t)$. Obviously, x should be chosen as an element in the Chebyshev center of the set

$$K = \{z^t : t \in T\}.$$

This is a compact set in $C(S)$. According to Theorem 1 and the comments accompanying it, one best approximant is the function u defined by

$$u(s) = \tfrac{1}{2} \max_{t\in T} z(s,t) + \tfrac{1}{2} \min_{t\in T} z(s,t).$$

If one desires to approximate z by an element of a prescribed subspace U in $C(S)$, then the best approximations (if they exist) are precisely the elements of the relative Chebyshev center $E(K;U)$, if it is nonvoid.

Before we proceed too far into this subject, it is important to point out that every simultaneous approximation problem can be turned into an ordinary approximation problem (in a different space). One method for doing so, given in Franchetti and Cheney [1981-b], is as follows. Suppose that we seek the simultaneous approximations of a bounded set S in a Banach space X by elements chosen from a subset U in X. Consider the Banach space $C(S, X)$ of all bounded continuous maps f from S into X, normed by

$$\|f\| = \sup_{s \in S} \|f(s)\|.$$

For each $u \in U$, define $\overline{u} \in C(S, X)$ by putting $\overline{u}(s) = u$. Let $\overline{U} = \{\overline{u} : u \in U\}$. Define also $e \in C(S, X)$ by putting $e(s) = s$. For each $u \in U$ we have

$$\|e - \overline{u}\| = \sup_{s \in S} \|e(s) - \overline{u}(s)\| = \sup_{s \in S} \|s - u\|.$$

Consequently, the simultaneous approximation of S by U in X is equivalent to the approximation of e by $\overline{U}$ in $C(S, X)$. These remarks do not make the problem any easier but perhaps add to the motivation for studying approximation theory in the spaces $C(S, X)$.

Returning now to relative Chebyshev centers in $C(S)$, we mention another way of turning the simultaneous approximation problem into an approximation involving a single function. Let B be a bounded set in $C(S)$, with S compact Hausdorff. In addition to the functions v_1 and v_2 defined previously in Eq. (2), we put

$$w = \tfrac{1}{2}(v_1 - v_2), \qquad c = \tfrac{1}{2}(v_1 + v_2).$$

It was proved by Diaz and McLaughlin [1969-b] that

$$r(B; U) = \inf_{u \in U} \|w + |c - u|\|. \tag{3}$$

Furthermore, the elements of U that solve the minimization problem (3) are in $E(B; U)$, and each element of $E(B; U)$ solves (3). The problem in (3) is a Chebyshev approximation problem with an additive weight function.

Within the same sphere of ideas is a result of Smith and Ward [1975]. They proved that

$$r(B; U) = r(B) + \operatorname{dist}\big(E(B), U\big)$$

and that $E(B; U)$ is nonempty if and only if $\operatorname{dist}(x, U)$ attains its infimum as x ranges over $E(B)$. Since this is a Banach-space result, it must be valid in all abstract M-spaces. Hence it is true for all spaces $C(S)$, whether S is compact or not. (Smith and Ward proved the theorem for S paracompact.) An example in $\mathbb{R}^2$ using the ℓ_∞-metric is shown in Figure 3.

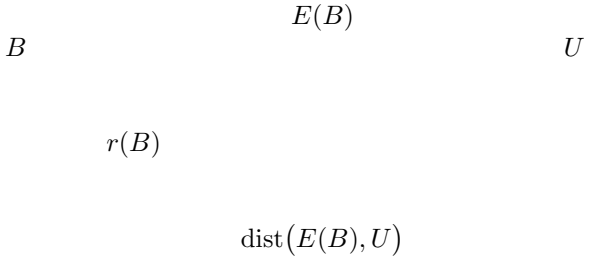

Fig. 3. *Relative Chebyshev center.*

An important existence theorem for relative Chebyshev centers was proved by Smith and Ward [1975-a]:

THEOREM. *If S is compact Hausdorff, if U is a closed subalgebra of $C(S)$ and if B is a bounded subset of $C(S)$, then the relative Chebyshev center $E(B;U)$ is nonempty.*

For a generalization of this theorem, see Amir, Mach and Saatkamp [1982].

In spaces other than $C(S)$, a number of results have been obtained, of which a few are listed here.

THEOREM. *The Chebyshev center of a bounded set in a uniformly convex Banach space consists of precisely one point* (Garkavi [1962]).

THEOREM. *Let X be a Banach space whose canonical image in X^{**} is norm-1 complemented. Then every bounded set in X has a nonempty Chebyshev center* (Garkavi [1962]).

THEOREM. *In Hilbert space, the Chebyshev center of a hypercircle is its element of least norm.* (A **hypercircle** is the intersection of a closed linear manifold with a ball centered at the origin) Golomb and Weinberger [1959].

THEOREM. *If X is any normed linear space, then every bounded set in X^* has a nonempty Chebyshev center* (Garkavi [1962]).

THEOREM. *If X is an abstract L-space, then every bounded set in X has a nonempty Chebyshev center* (Prolla [1983]).

THEOREM. *If X is a Banach space in which each bounded set has a nonempty center, then the same is true of any norm-1 complemented subspace of X.*

In much of the theory of Chebyshev centers, the results concerning centers of bounded sets are weaker than those for compact sets. If U is a subspace in a Banach space X such that $E(K;U) \neq \square$ for all compact sets K in X, then U is said to have property "EK." Property "EO" is similar but for all bounded sets B in K. These properties are like ordinary proximinality for U, except that they pertain to the simultaneous approximation problem. We have these implications, for a subspace U:

$$EO \Longrightarrow EK \Longrightarrow \text{proximinality.}$$

In general these implications are not reversible. Zamyatin [1973] proved that the first of these is not reversible, and Feng [1986] proved that even in $C(S)$ the second is not. Here is a theorem characterizing the EK-subspaces in $C(S)$. (See Franchetti and Cheney [1986].)

THEOREM. *In order that a subspace U in $C(S)$ have property EK it is necessary and sufficient that U be proximinal in $C(S \times T)$ for every compact Hausdorff space T.*

A similar theorem is true for property EO; in this case U must be proximinal in $C(S \times T)$ for all topological spaces T. In both theorems, S is compact Hausdorff.

Another characterization from the same paper is this:

THEOREM. *A proximinal subspace U in $C(S)$ has property EK if and only if* $\operatorname{dist}(x, U)$ *attains its infimum on each "interval" $[a, b]$ where $a \in C(S)$, $b \in C(S)$, and* $\min_S \big[b(s) - a(s)\big] = 0$.

This theorem follows from the result of Smith and Ward cited earlier, together with the remark that the Chebyshev centers of compact sets in $C(S)$ are precisely the intervals of the type described in the theorem.

In order to illustrate how the theory of Chebyshev centers can assist in proving theorems about ordinary proximinality, we consider the common technique of introducing changes of variable in approximation problems. For example, practical problems may lead to the use of approximating functions of the form $\sum_{i=0}^{n} a_i \big[f(s)\big]^i$, which is nothing but a polynomial in the variable $f(s)$. A general theorem about proximinality in such problems is this:

THEOREM. *Let S and T be compact Hausdorff spaces. Let V be a subspace having property EK in $C(T)$. Let $f : S \to T$ be continuous and open. Then the set of functions $\{v \circ f : v \in V\}$ is proximinal in $C(S)$.*

The subspaces of finite co-dimension have received special attention in spaces $C(S)$. Here is a beautiful theorem of Garkavi [1973]:

THEOREM. *A subspace of finite co-dimension in $C(S)$ has property EK if and only if it is proximinal.*

In contrast to this, Zamyatin [1973] has proved:

THEOREM. *In order that a subspace U of finite co-dimension in $C(S)$ have property EO it is necessary and sufficient that each element of $U^{\perp}$ have finite support.*

From these two theorems it is clear that some hyperplanes in $C(S)$ have property EK but not property EO.

Further references on the problem of Chebyshev centers are Amir and Ziegler [1980-a, 1980-b, 1981], Blatt [1973], Bosznay [1978], Brondsted [1976], Carroll [1972], Chui, Rahman, Sahney and Smith [1978], Diaz and McLaughlin [1969-a], Dierieck [1976], Dunham [1967], Franchetti [1977], Gillotte and McLaughlin [1976], Hall [1976], Holland, Sahney, and Tzimbalario [1976], Kadets and Zamyatin [1968], Laurent and Tuan [1970], Lin [1974], Mach [1979-a, 1979-b, 1982], Milman [1977], Sahney and Singh [1982], and Owens [1983].

CHAPTER 8
Multivariate Interpolation

In Chapter 1, we pointed out some differences between univariate and multivariate interpolation, and we described the method of Shepard interpolation as being representative of the procedures available for the multivariate case. Here we will describe some other methods that are useful and have been the subject of recent research. The discussion relies heavily on Micchelli's survey [1986], which is recommended for further reading.

First, it should be pointed out that interpolation on a Cartesian grid presents no challenge, since the *tensor product* of two univariate interpolation operators is a bivariate interpolation operator. Thus, if an operator A interpolates at points $s_1, \cdots, s_n$ and if an operator B interpolates at $t_1, \cdots, t_m$ then $A \otimes B$ will interpolate at all grid points (s_i, t_j). This statement generalizes to the interpolation of functionals other than point-evaluations. The formal result is as follows.

THEOREM 1. *Let A and B be linear operators in $\mathcal{L}(X, X)$ and $\mathcal{L}(Y, Y)$, respectively. Let α be a uniform crossnorm. If $A^*\phi = \phi$ and $B^*\psi = \psi$ for a pair of functionals $\phi \in X^*$, $\psi \in Y^*$, then for all $z \in X \otimes_\alpha Y$,*

$$(\phi \otimes \psi)(A \otimes_\alpha B)(z) = (\phi \otimes \psi)(z) \ . \tag{1}$$

Specializing to ordinary polynomials, we see that interpolation on a Cartesian grid

$$\big\{(s_i, t_j) : 0 \le i \le n,\ 0 \le j \le m\big\}$$

is uniquely possible by a polynomial of the form

$$p(s,t) = \sum_{\nu=0}^{n} \sum_{\mu=0}^{m} a_{\nu\mu} s^\nu t^\mu \ . \tag{2}$$

The uniqueness follows from two facts: that the interpolation on the gridwork of $(n+1)(m+1)$ points is *possible*, and that the dimension of the subspace being used for interpolation is also $(n+1)(m+1)$. Let us elaborate this second point.

The polynomial subspace that figures in this tensor-product scheme is $\Pi_n \otimes \Pi_m$. According to the definition of a tensor product, the elements of this space are finite sums of the form

$$p(s,t) = \sum u_i(s) v_i(t) \qquad (u_i \in \Pi_n \ ,\ v_i \in \Pi_m) \ .$$

By simple linear algebra these elements are representable as in Eq. (2). The dimension of this space is $(n+1)(m+1)$, by the general theory of tensor products.

The polynomial subspace which is usually denoted by $\Pi_m(\mathbb{R}^2)$ consists precisely of all polynomials having the form

$$p(s,t) = \sum_{i+j \le m} a_{ij} s^i t^j = \sum_{i=0}^{m} s^i \sum_{j=0}^{m-i} a_{ij} t^j \ . \tag{3}$$

The following theorem and proof are adapted from Micchelli [1986].

THEOREM 2. *Interpolation to arbitrary data by an element of* $\Pi_m(\mathbb{R}^2)$ *is uniquely possible on a set* $\mathcal{N}$ *of* $(m+1)(m+2)/2$ *nodes if there exist* $m+1$ *lines* $L_0, \cdots, L_m$ *whose union contains* $\mathcal{N}$ *and that have the property that each* L_i *contains exactly* $i+1$ *nodes.*

Proof. The sets $\mathcal{N} \cap L_i$ are pairwise disjoint, for if they were not, we would have a strict inequality in the following calculation, where $n = \frac{1}{2}(m+1)(m+2)$:

$$n = \#\mathcal{N} \leq \sum_{i=0}^{m} \#(\mathcal{N} \cap L_i) = 1 + 2 + \cdots + (m+1) = n \ .$$

Since the number of coefficients in the polynomial

$$p(s,t) = \sum_{i+j \leq m} a_{ij} s^i t^j$$

is n and the number of nodes is n, it suffices to prove that if $p \mid \mathcal{N} = 0$ then $p = 0$. For each i, let ℓ_i be a element of $\Pi_1(\mathbb{R}^2)$ such that $L_i = \{x : \ell_i(x) = 0\}$. Observe that $p^2 + \ell_m^2$ has $m+1$ zeros because $\#(\mathcal{N} \cap \ell_m) = m+1$. By Bézout's Theorem (below) we conclude that p contains ℓ_m as a factor. In the same way, $(p/\ell_m)^2 + \ell_{m-1}^2$ has m zeros, and by Bézout's Theorem, p/ℓ_m contains ℓ_{m-1} as a factor. Continuing the argument, we conclude that p is a scalar multiple of $\ell_1 \cdots \ell_m$. Since p vanishes on $\mathcal{N} \cap \ell_0$ while $\ell_1 \cdots \ell_m$ does not, $p = 0$. □

BÉZOUT'S THEOREM. *If* $p \in \Pi_n(\mathbb{R}^2)$ *and* $q \in \Pi_m(\mathbb{R}^2)$*, and if* $p^2 + q^2$ *has more than* nm *zeros, then* p *and* q *have a nonconstant factor in common.* (See Hartshorne [1977].)

Another special distribution of nodes for which interpolation is possible is given in this theorem of Chung and Yao [1977]:

THEOREM. *Interpolation to arbitrary data on a node set* $\mathcal{N}$ *in* $\mathbb{R}^n$ *is uniquely possible with elements from* $\Pi_m(\mathbb{R}^n)$ *if* $\#\mathcal{N} = \binom{n+m}{m}$ *and if there is associated with each node* x *a set* $S(x)$ *that is a union of* m *hyperplanes and has the property* $x \in S(y) \Leftrightarrow x \neq y$*, for* x *and* y *in* $\mathcal{N}$*.*

For computing interpolating functions it is important to have efficient algorithms that are easily programmed. The Newton algorithm for univariate polynomial interpolation is the epitome of such algorithms. We can summarize it by saying that if a polynomial p interpolates to a function f at nodes $x_1, \cdots, x_{r-1}$ then a polynomial that interpolates f at nodes $x_1, \cdots, x_r$ can be found in the form

$$p(x) + c \prod_{i=1}^{r-1} (x - x_i) \ .$$

The abstract embodiment of this principle can be stated in the following terms. Let X be a set and f a real-valued function on X. Let $\mathcal{N}$ be a set of points ("nodes") in X. Let h be a real-valued function on X, and put

$$Z = \{x \in X : h(x) = 0\} \ .$$

If p interpolates f on $\mathcal{N} \cap Z$ and if q interpolates $(f - p)/h$ on $\mathcal{N} \setminus Z$, then $p + qh$ interpolates f on $\mathcal{N}$.

The Newton algorithm as described above becomes an instance of this abstract formulation when $X = \mathbb{R}$, $h = \prod_{i=1}^{r-1}(x - x_i)$, and q is the constant

$$q = \big[f(x_r) - p(x_r)\big]/h(x_r) \ .$$

This abstract version of the Newton algorithm enables us to divide an interpolation problem into two smaller problems. (Here "smallness" refers to the number of interpolation conditions.) In the above discussion, the subproblems involve the determination of p and q. Procedures such as this are often fundamental to successful algorithms.

As Micchelli points out, this strategy is completely successful in the interpolation problem of Theorem 2. We let ℓ_m denote an element of $\Pi_1(\mathbb{R}^2)$ whose zero set, $Z(\ell_m)$, is the line L_m. Let p_m be an element of $\Pi_m(\mathbb{R}^2)$; this solves the easy problem of interpolating f on $\mathcal{N} \cap L_m$. Let q_{m-1} be an element of $\Pi_{m-1}(\mathbb{R}^2)$ which interpolates $(f - p_m)/\ell_m$ on $\mathcal{N} \setminus Z(\ell_m)$. If this process is applied m times, the solution takes on successively the following forms:

$$\begin{aligned} p_m + \ell_m q_{m-1} &= p_m + \ell_m(p_{m-1} + \ell_{m-1}q_{m-2}) \\ &= p_m + \ell_m\big(p_{m-1} + \ell_{m-1}(p_{m-2} + \ell_{m-2}q_{m-3})\big) \\ &\cdots\cdots\cdots \\ &= \sum_{i=1}^{m} p_i \prod_{j=i+1}^{m} \ell_j \ . \end{aligned}$$

The interpolation problem of Theorem 2 can be solved by another type of polynomial in two variables. Using the previous notation, we set up an interpolating function of the form

$$r_o + \ell_0 r_1 + \ell_0\ell_1 r_2 + \ell_0\ell_1\ell_2 r_3 + \cdots + (\ell_0 \cdots \ell_{m-1})r_m$$

where $\ell_i \in \Pi_1(\mathbb{R}^2)$ and $r_i \in \Pi_i(\mathbb{R}^2)$. The polynomials $r_0, r_1, \cdots$ are determined (in that order) by imposing the interpolation conditions on $\mathcal{N} \cap L_0$, $\mathcal{N} \cap L_1$, etc. Each r_i can be of the special form $r_i = g_i \circ \ell$ for an appropriate $\ell \in \Pi_1(\mathbb{R}^2)$ and $g_i \in \Pi_i(\mathbb{R})$. This is because each r_i is called upon to solve an interpolation only on the $i + 1$ points which compose $\mathcal{N} \cap L_i$. The affine function ℓ should separate the points of $\mathcal{N}$ (i.e., should be one-to-one from $\mathcal{N}$ to $\mathbb{R}$).

For polynomial interpolation at n nodes not regularly distributed in $\mathbb{R}^2$, it can argued that the probability of encountering a singular system of linear equations is zero. Hence one can select a set of n monomials $s^\mu t^\nu$ and attempt to solve for the coefficients in the equation

$$\sum_{\mu}\sum_{\nu} a_{\mu\nu} s_i^\mu t_i^\nu = \lambda_i, \quad 1 \le i \le n \tag{4}$$

where (s_i, t_i) are the nodes. With probability 1, this system is nonsingular. However, it may be nearly-singular, and in any event, for large n the solution of Eq. (4) may be a formidable task. Furthermore, polynomials are not suitable for *all* problems involving the representation of data. Hence we require other, more general, interpolation schemes.

Interpolation at nodes having no exploitable pattern is referred to as the case of "scattered data." The method of Shepard, outlined in Chapter 1, is applicable to this problem.

Another important method was introduced by Hardy [1971]; it uses functions called "multiquadrics." Franke [1982-b] and others have found this method to be very satisfactory. Here, the interpolating function has the form

$$g(s,t) = \sum_{j=1}^{n} c_j \left[(s-s_j)^2 + (t-t_j)^2 + r^2 \right]^{1/2},$$

in which (s_j, t_j) are the nodes of interpolation and r is a parameter at our disposal. There is one basis function for each node. In order to interpolate a function f at the nodes we have to solve the linear system

$$\sum_{j=1}^{n} c_j \left[(s_i-s_j)^2 + (t_i-t_j)^2 + r^2 \right]^{1/2} = f(s_i, t_i) \quad (1 \le i \le n) \ .$$

Micchelli proved [1985] that the coefficient matrix in this system is non-singular, provided that all the nodes are different from each other.

Determining an optimal value of the parameter r (practically or theoretically) is an interesting problem. Hardy [1971] suggested $r = 0.815d$, where d is the mean distance between nodes. Other rules-of-thumb have been suggested, all depending upon the geometry of the node set.

A closely related type of interpolating function, called a "thin-plate spline," was introduced by Harder and Desmarais [1972]. It has the form

$$g(s,t) = \sum_{j=1}^{n} c_j \rho_k(s,t) \log \rho_k(s,t) \quad \rho_k(s,t) = (s-s_k)^2 + (t-t_k)^2 \ .$$

See also Duchon [1976], Meinguet [1979-a] and the references cited by Franke [1982-b].

Several "local" versions of Shepard's method have been proposed. This term has an informal meaning that the subspace of functions being used for interpolation has a basis consisting of functions with small supports. When this is the case, the linear system of equations that must be solved to obtain the interpolant will have a sparse coefficient matrix. Furthermore, and this is more important, the evaluation of the interpolant at an arbitrary point will be much simpler because only a few of the basis functions will be different from zero at that point.

A simple local Shepard method for nodes $x_1, \cdots, x_n$ in $\mathbb{R}^2$ (or higher dimensional space) employs basis functions $g_1, \cdots, g_n$ of the form

$$g_i(x) = r_i^{-\mu} \big(r_i - |x - x_i| \big)_+^{\mu} \ .$$

(The subscript + indicates that $g_i(x) = 0$ if $|x - x_i| \le r_i$.) The numbers μ and r_i are parameters at our disposal. If r_i is defined to be the distance from x_i to its nearest neighboring node, then $g_i(x_j) = \delta_{ij}$, and the Shepard interpolant for a function f is simply $\sum_{i=1}^{n} f(x_i) g_i$. Increased smoothness of g_i is obtained by raising the value of μ. See Franke [1985] for further remarks about this topic.

Further references on multivariate interpolation are Carlson and Fritsch [1985], Ciarlet and Raviart [1971, 1972], Franke [1977, 1982-a, 1982-b, 1985], Hakopian [1982], Haussmann [1970, 1972], Kergin [1980], Meinguet [1970, 1979-a, 1979-b], Micchelli [1980], Schumaker [1976], and Tarwater [1985].

REFERENCES

P. ALFELD [1985], *Multivariate perpendicular interpolation*, SIAM J. Numer. Anal. **22**, p. 95.

D. AMIR [1978], *Chebyshev centers and uniform convexity*, Pacific J. Math. **77**, pp. 1–6.

D. AMIR AND F. DEUTSCH [1979], *Approximation by certain subspaces in the Banach space of continuous vector-valued functions*, J. Approx. Theory **27**, pp. 254–270.

D. AMIR, J. MACH, AND K. SAATKAMP [1982], *Existence of Chebyshev centers, best n-nets, and best compact approximants*, Trans. Amer. Math. Soc. **271**, pp. 513–524.

D. AMIR AND Z. ZIEGLER [1980-a], *Relative Chebyshev centers in normed linear spaces*, J. Approx. Theory **29**, pp. 235–252.

——— [1980-b], *Characterizations of relative Chebyshev centers*, Approximation Theory III, E.W. Cheney, ed., Academic Press, New York, pp. 157–162.

——— [1981], *Construction of elements of the relative Chebyshev center*, Approximation Theory and Applications, Z. Ziegler, ed., Academic Press, New York, pp. 1–11.

B. ATLESTAM AND F. SULLIVAN [1976], *Iteration with best approximation operators*, Rev. Roumaine Math. Pures Appl. **21**, 125–131.

G. AUMANN [1958], *Uber approximative Nomographie I, II, III*, Bayer. Akad. Wiss. Math.-Natur. Kl. S.B., pp. 137–155; ibid. (1959), pp. 103–109; ibid. (1960), pp. 27–34.

V.F. BABENKO AND S.A. PICHUGOV [1981], *A property of compact operators in the space of integrable functions*, Ukrain. Mat. Zh. **33**, pp. 491–492. [English translation: Ukranian Math. J.]

R. BANK [1979], *An automatic scaling procedure for a d'Yakanov-Gunn iteration scheme*, Linear Algebra Appl. **28**, pp. 17–33.

R.E. BARNHILL [1976], *Blending function interpolation: a survey and some new results*, Numerische Methoden der Approximationstheorie, ISNM, Vol.**30**, Birkhäuser, Basel, pp. 43–89.

——— [1977], *Representation and approximation of surfaces*, Mathematical Software II, J.R. Rice, ed., Academic Press, New York, pp. 69–120.

R.E. BARNHILL, G. BIRKHOFF AND W.J. GORDON [1973], *Smooth interpolation in triangles*, J. Approx. Theory **8**, pp. 114–128.

R.E. BARNHILL, R.P. DUBE AND F.F. LITTLE [1983], *Properties of Shepard's surfaces*, Rocky Mountain J. Math. **13**, pp. 365–382.

R.B. BARRAR AND H.L. LOEB [1984], *Lifting a distribution*, Approximation Theory IV, C.K. Chui, L.L. Schumaker and J.D. Ward, eds., Academic Press, New York, pp. 323–327.

F.L. BAUER [1963], *Optimally scaled matrices*, Numer. Math. **5**, pp. 73–87.

——— [1969], *Remarks on optimally scaled matrices*, Numer. Math **13**, pp. 1–3.

H.-P. BLATT [1973], *Nicht-lineare gleichmässig Simultanapproximation*, J. Approx. Theory **8**, pp. 210–248.

J. BLATTER AND E.W. CHENEY [1971], *On the existence of extremal projections*, J. Approx. Theory **6**, pp. 72-79.

A.P. BOSZNAY [1978], *A remark on simultaneous approximation*, J. Approx. Theory **23**, pp. 296–298.

A. BRONDSTED [1976], *A note on best simultaneous approximation in normed linear spaces*, Canad. Math. Bull. **19**, pp. 359–360.

F.E. BROWDER, ed. [1976], *Mathematical Developments Arising from Hilbert's Problems*, Proc. Sympos. in Pure Math. **28**, Amer. Math. Soc.

A.L. BROWN [1982-a], *An extension to Mairhuber's theorem on metric projections and discontinuity of multivariate best uniform approximation*, J. Approx. Theory **36**, pp. 156–172.

——— [1982-b], *Finite rank approximations to integral operators which satisfy certain total positivity conditions*, J. Approx. Theory **34**, pp. 42–90.

Ju. A. BRUDNYI [1970], *Approximation of functions of n variables by quasipolynomials*, Izv. Akad. Nauk SSSR Ser. Mat. **34** (Russian). [English translation: Math. USSR-Izv. **4**, pp. 568–586.]

R.C. BUCK [1968-a], *Functional decomposition and metric entropy*, Institute for Defense Analysis Report No.**8**, Princeton, N.J.

——— [1968-b], *Alternation theorems for functions of several variables*, J. Approx. Theory **1**, pp. 325–334.

——— [1972], *On approximation theory and functional equations*, J. Approx. Theory **5**, pp. 228–237.

——— [1974], *Approximation properties of vector-valued functions*, Pacific J. Math. **53**, pp. 85–94.

——— [1975], *Approximate functional complexity*, Bull. Amer. Math. Soc. **81**, pp. 1112–1114.

——— [1976], *Approximate functional complexity*, Approximation Theory II, G.G. Lorentz, C.K. Chui and L.L. Schumaker, eds., Academic Press, New York, pp. 303–307.

——— [1979], *Approximate complexity and functional representation*, J. Math. Anal. Appl. **70**, pp. 280–298.

R.E. CARLSON AND F.N. FRITSCH [1985], *Monotone piecewise bicubic interpolation*, SIAM J. Numer. Anal. **22**, pp. 386–400.

M.P. CARROLL [1972], *Simultaneous L_1 approximation of a compact set of real-valued functions*, Numer. Math. **19**, pp. 110–115.

A.S. CAVARETTA, C.A. MICCHELLI AND A. SHARMA [1980], *Multivariate interpolation and the Radon transform*, Math. Zit. **174**, pp. 263–279.

E.W. CHENEY [1965], *An elementary proof of Jackson's theorem on mean approximation*, Math. Mag. **38**, pp. 189–191.

E.W. CHENEY AND K.H. PRICE [1970], *Minimal projections*, Approximation Theory, A. Talbot, ed., Academic Press, New York, pp. 261–289.

S. CHEVET [1969], *Sur certains produits tensoriels topologiques d'espaces de Banach*, Z. Wahrscheinlichkitstheorie Verw. Gebiete **11**, pp. 120–138.

C.K. CHUI, Q.I. RAHMAN, B.N. SAHNEY AND P.W. SMITH [1978], *On simultaneous Chebyshev approximation*, J. Approx. Theory **22**, pp. 350–351.

K.C. CHUNG AND T.H. YAO [1977], *On lattices admitting unique Lagrange interpolations*, SIAM J. Numer. Anal. **14**, pp. 735–743.

P.G. CIARLET AND P.-A. RAVIART [1971], *Interpolation de Lagrange dans $\mathbb{R}^n$*, C.R. Acad. Sci. Paris **273**, pp. 578–581.

——— [1972], *General Lagrange and Hermite interpolation in $\mathbb{R}^n$ with applications to finite element method*, Arch. Rational Mech. Anal. **46**, pp. 177–199.

L. COLLATZ [1956], *Approximation von Funktionen bei einer und bei mehreren unabhängigen Veränderlichen*, Z. Angew Math. Mech.**36**, pp. 198–211.

——— [1972-a], *Approximation by functions of fewer variables*, Conference on the theory of ordinary and partial differential equations, Dundee, pp. 16–31. Lecture Notes in Mathematics, Vol.**280**, W.N. Everitt and B.D. Sleeman, eds., Springer, Berlin.

——— [1972-b], *Zur Tschebyschew-Approximation bei Funktionen mehrerer unabhängiger Veränderlicher*, Constructive Theory of Functions, Academiei Kiado, Budapest, pp. 89–99.

A.R. CURTIS AND J.K. REID [1972], *On the automatic scaling of matrices for Gaussian elimination*, J. Inst. Math. Appl. **10**, pp.118–124.

R.B. DARST, D.A. LEGG AND D.W. TOWNSEND [1983], *The Polya algorithm in L_∞ approximation*, J. Approx. Theory **38**, pp. 209–220.

I.K. DAUGAVET [1963], *A property of compact operators in the space C*, Uspekhi Mat. Nauk **18**, pp. 157–158.

V.A. DAUGAVET [1981], *Approximation of a function of two variables by a sum of products of functions of one variable* (Russian), Metody Vycisl No. **12**, pp. 174–186, 235.

F.J. DELVOS [1975], *On surface interpolation*, J. Approx. Theory **15**, pp. 209–213.

F.J. DELVOS AND H.-W. KÖSTERS [1975], *On the variational characterization of bivariate interpolation methods*, Math. Z. **145**, pp. 129–137.

F.J. DELVOS AND G. MALINKA [1974], *Das Blending-Schema von Spline Systemen*, Spline-Funktionen, K. Böhmer, G. Meinardus and W. Schempp, eds, Bibliographisches Institute, Mannheim, pp. 47–57.

F.J. DELVOS AND W. SCHEMPP [1983], *The method of parametric extension applied to right invertible operators*, Numer. Funct. Anal. Opt. **6**, pp. 135–148.

F. DEUTSCH [1976], *Review of the paper by Goel, Holland, Nasim and Sahney*, Math. Reviews **51**, #13556.

——— [1979], *The alternating method of von Neumann*, Multivariate Approximation Theory, W. Schempp and K. Zeller, eds., ISNM Vol.**51**, Birkhäuser, Basel, pp. 83–96.

——— [1984], *Von Neumann's alternating method: the rate of convergence*, Approximation Theory IV, C.K. Chui, L.L. Schumaker and J.D. Ward, eds., Academic Press, New York, pp. 427–434.

F. DEUTSCH, J. MACH AND K. SAATKAMP [1981], *Approximation by finite rank operators*, J. Approx. Theory **33**, pp. 199–213.

J.B. DIAZ AND H.W. MCLAUGHLIN [1969-a], *Simultaneous Chebyshev approximation of a set of bounded complex-valued functions*, J. Approx. Theory **2**, pp. 419–432.

——— [1969-b], *Simultaneous approximation of a set of bounded real functions*, Math. Comp. **23**, pp. 538-584.

——— [1972], *On simultaneous Chebyshev approximation and Chebyshev approximation with an additive weight function*, J. Approx. Theory **6**, pp. 68–71.

C. DIERIECK [1976], *Simultaneous approximation of vector-valued functions*, J. Approx. Theory **18**, pp. 74–85.

J. DIESTEL AND J.J. UHL [1977], *Vector Measures*, Amer. Math. Soc., Providence, R.I.

S.P. DILIBERTO AND E.G. STRAUS [1951], *On the approximation of a function of several variables by the sum of functions of fewer variables*, Pacific J. Math. **1**, 195-210.

J. DUCHON [1976], *Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces*, RAIRO Anal. Numér. **10**, pp. 5–12.

N. DUNFORD AND R. SCHATTEN [1946], *On the associate and conjugate space for the direct product of Banach spaces*, Trans. Amer. Math. Soc. **59**, pp. 430–436.

C.B. DUNHAM [1967], *Simultaneous Chebyshev approximation of functions on an interval*, Proc. Amer. Math. Soc. **18**, pp. 472–477.

——— [1977], *Simultaneous approximation of a set*, J. Approx. Theory **21**, pp. 205-208.

N. DYN [1980], *A straightforward generalization of Diliberto and Straus' algorithm does not work*, J. Approx. Theory **30**, pp. 247–250.

N. DYN, D. LEVIN AND S. RIPPA [1983], *Surface interpolation and smoothing by thin plate splines*, Approximation Theory IV, C.K. Chui, L.L. Schumaker and J.D. Ward, eds., Academic Press, New York, pp. 445–449.

B.C. EAVES, A.J. HOFFMAN, U.G. ROTHBLUM AND H. SCHNEIDER [1985], *Line-sum-symmetric scalings of square nonnegative matrices*, Math. Programming Stud. **25**, pp. 124–141.

H. EHLICH AND W. HAUSSMAN [1970], *Cebysev-Approximation stetiger Funktionen in zwei Veränderlichen*, Math. Z. **117**, pp. 21–34.

R. FARWIG [1984], *Rate of convergence of Shepard's global interpolation formula*, Sonderforschungsbereich **72**, Preprint no. 642, University of Bonn.

G. FENG [1986], *A counterexample on global Chebyshev approximation* J. Approx. Theory, to appear.

C. FOIAS AND I. SINGER [1965], *Points of diffusion of linear operators*, Math. Z. **87**, pp. 434–450.

A.R. FORREST [1972], *On Coons' and other methods for the representation of curved surfaces*, Computer Graphics and Image Processing **1**, pp. 341–359.

C. FRANCHETTI [1977], *Restricted centers and best approximation in $C(Q)$*, Ann. Fac. Sci. Univ. Nat. Zaïre (Kinshasa) Sect. Math.-Phys. **3**, pp. 35–45.

C. FRANCHETTI AND E.W. CHENEY [1981-a], *Best approximation problems for multivariate functions*, Bollettino Unione Mat. Ital. **18**-B, pp. 1003–1015.

——— [1981-b], *Simultaneous approximation and restricted Chebyshev centers in function spaces*, Approximation Theory and Applications, Z. Ziegler, ed., Academic Press, New York, pp. 65–88.

——— [1984], *Minimal projections in tensor-product spaces*, J. Approx. Theory **41**, 367–381.

——— [1986], *The embedding of proximinal sets*, J. Approx. Theory, to appear.

C. FRANCHETTI AND W.A. LIGHT [1984], *The alternating algorithm in uniformly convex spaces*, J. London Math. Soc. (2), **29**, pp. 545–555.

——— [1986], *On the von Neumann alternating algorithm in Hilbert spaces*, J. Math. Analysis and Applications, to appear.

R. FRANKE [1977], *Locally determined smooth interpolation at irregularly spaced points in several variables*, J. Inst. Math. Appl. **19**, pp. 471–482.

——— [1982-a], *Smooth interpolation of scattered data by local thin plate splines*, Comput. Math. Appl. **8**, pp. 273–281.

——— [1982-b], *Scattered data interpolation: tests of some methods*, Math. Comp. **38**, pp. 181–200.

——— [1985], *Thin plate splines with tension*, Comput. Aided Geom. Des. **2**, pp. 87–95.

D.R. FULKERSON AND P. WOLFE [1962], *An algorithm for scaling matrices*, SIAM Rev. **4**, pp. 142–146.

A.L. GARKAVI [1961], *On the Chebyshev center of a set in a normed space*, Investigations of Contemporary Problems in the Constructive Theory of Functions, Moscow, pp. 328–331.

——— [1962], *The best possible net and the best possible cross-section of a set in a normed linear space*, Izv. Akad. Nauk SSSR **26**, pp. 87–106 (Russian). [English translation: Amer. Math. Soc. Translations, Ser.2, **39** (1964), pp. 111–132.]

——— [1964], *The Chebyshev center and the convex hull of a set*, Uspekhi Mat. Nauk **18**, pp. 139–145.

——— [1973], *The conditional Chebyshev center of a compact set of continuous functions*, Mat. Zametki **14**, pp. 469–478 (Russian). [English translation: Math. Notes of the USSR (1973), pp. 827–831.]

A.L. GARKAVI AND V.N. ZAMYATIN [1975], *Conditional Chebyshev center of a bounded set of continuous functions*, Mat. Zametki **18**, pp. 67–76 (Russian). [English translation: Math. Notes of the USSR.]

W.B. GEARHART [1973], *Some Chebyshev approximations by polynomials in 2 variables*, J. Approx. Theory **8**, pp. 195– 209.

M.J. GILLOTTE AND H.W. MCLAUGHLIN [1976], *On additive weight approximation and simultaneous Chebyshev approximation using varisolvent families*, J. Approx. Theory **17**, pp. 35–43.

M. V. GOLITSCHEK [1980-a], *Optimal cycles in weighted and doubly weighted digraphs and applications to approximation problems*, Preprint no. 61, Math. Inst., University of Würzburg.

——— [1980-b], *An algorithm for scaling matrices and computing the minimum cycle mean in a digraph*, Numer. Math. **35**, pp. 45–55.

———— [1980-c], *Approximating bivariate functions and matrices by nomographic functions*, Quantitative Approximation, R.A. Devore and K. Scherer, eds., Academic Press, New York, pp. 143–151.

———— [1980-d], *Approximation of functions of two variables by the sum of two functions of one variable*, Numerical Methods of Approximation Theory, L. Collatz, G. Meinardus, and H. Werner, eds., Birkhäuser Verlag, ISNM Vol.**52**, pp. 117–124.

———— [1982], *Optimal cycles in doubly weighted graphs and approximation of bivariate functions by univariate ones*, Numer. Math. **39**, pp. 65–84.

———— [1984], *Shortest path algorithms for the approximation by nomographic functions*, Approximation Theory and Functional Analysis (Proceedings of Conference, Oberwolfach, July 30 – August 6, 1983), P.L. Butzer, R.L. Stens and B. Sz-Nagy, eds., ISNM Vol.**65**, Birkäuser Verlag, Basel.

M. V. GOLITSCHEK AND E.W. CHENEY [1979], *On the algorithm of Diliberto and Straus for approximating bivariate functions by univariate ones*, Numer. Funct. Anal. Optim. **1**, pp. 341–363.

———— [1983-a], *The best approximation of bivariate functions by separable functions*, Contemporary Math. **21**, pp. 125–136.

———— [1983-b], *Failure of the alternating algorithm for best approximation of multivariate functions*, J. Approx. Theory **38**, pp. 139–143.

M.V. GOLITSCHEK, U.G. ROTHBLUM AND H. SCHNEIDER [1983], *A conforming decomposition theorem, a piecewise linear theorem of the alternative, and scalings of matrices satisfying lower and upper bounds*, Math. Programming **27**, pp. 291–306.

M.V. GOLITSCHEK AND H. SCHNEIDER [1984], *Applications of shortest path algorithms to matrix scalings*, Numer. Math. **44**, pp. 111–126.

M. GOLOMB [1959], *Approximation of functions of fewer variables*, On Numerical Approximation, R. Langer, ed., University of Wisconsin Press, Madison, Wisconsin, pp. 275–327.

M. GOLOMB AND H.F. WEINBERGER [1959], *Optimal approximation and error bounds*, On Numerical Approximation, R.E. Langer, ed., University of Wisconsin Press, Madison, Wisconsin, pp. 117-190.

H. GONSKA AND K. JETTER [1986], *Jackson-type theorems on approximation by trigonometric and algebraic pseudopolynomials*, J. Approx. Theory, to appear.

W.J. GORDON [1969-a], *Spline-blended surface interpolation through curve networks*, J. Math. Mech. **18**, pp. 931–952.

———— [1969-b], *Distributive lattices and the approximation of multivariate functions*, Approximation with Special Emphasis on Spline Functions, I.J. Schoenberg, ed., Academic Press, New York, pp. 223–277.

———— [1971], *Blending-function methods of bivariate and multivariate interpolation and approximation*, SIAM J. Numer. Anal. **8**, pp. 158–177.

W.J. GORDON AND E.W. CHENEY [1978], *Bivariate and multivariate interpolation with noncommutative projectors*, Linear Spaces and Approximation, P.L. Butzer and B. Sz.-Nagy, eds., ISNM **40**, Birkhäuser Verlag, Basel, pp. 381–387.

W.J. GORDON AND C.A. HALL [1973], *Transfinite element methods. Blending-function interpolation over arbitrary curved element domains*, Numer. Math. **21**, pp. 109–129.

W.J. GORDON AND J.A. WIXOM [1978], *Shepard's method of metric interpolation to bivariate and multivariate interpolation*, Math. Comp. **32**, pp. 253–264.

P.R. GRAVES-MORRIS AND R.H. JONES [1976], *An analysis of two variable rational approximants*, J. Comput. Appl. Math. **2**, p. 41.

R.B. GUENTHER AND E.L. ROETMAN [1970], *Some observations on interpolation in higher dimensions*, Math. Comp. **24**, pp. 517–522.

H. HAKOPIAN [1982], *Multivariate divided differences and multivariate interpolation of Lagrange and Hermite type*, J. Approx. Theory **34**, pp. 286-305.

C.A. HALL [1976], *Transfinite interpolation and applications to engineering problems*, Theory of Approximation with Applications, A.G. Law and B.N. Sahney, eds., Academic Press, New York, pp. 308–331.

E.J. HALTON AND W.A. LIGHT [1985-a], *Projections on tensor product spaces*, Trans. Amer. Math. Soc. **287**, pp. 161–165.

——— [1985-b], *Minimal projections in L_p-spaces*, Math. Proc. Cambridge Philos. Soc. **97**, pp. 127–136.

R.L. HARDER AND R.N. DESMARAIS [1972], *Interpolation using surface splines*, J. Aircraft **9**, pp. 189–191.

R.L. HARDY [1971], *Multiquadric equations of topography and other irregular surfaces*, J. Geophysical Research **76**, pp. 1905–1915.

P.H. HARTLEY [1976], *Tensor product approximations to data defined on rectangular meshes in n-space*, Comput. J. **19**, pp. 348–352.

R. HARTSHORNE [1977], *Algebraic Geometry*, Springer, New York.

W. HAUSSMANN [1970], *Tensorprodukte und mehrdimensionale Interpolation*, Math. Z. **113**, pp. 17–23. ——— [1972], *Tensorproduktmethoden bei mehrdimensionaler Interpolation*, Math. Z. **124**, pp. 191–198.

W. HAUSSMANN AND K. ZELLER [1980], *Multivariate approximation and tensor products*, Approximation Theory III, E.W. Cheney, ed., Academic Press, New York, pp. 495–500.

——— [1983], *Blending interpolation and best L^1-approximation*, Arch. Math. **40**, pp. 545–552.

——— [1984], *Mixed norm multivariate approximation with blending functions*, Proceedings of the Conference on Constructive Theory of Functions, Varna, Bulgaria, Bulgarian Academy of Sciences, Sofia, pp. 403–408.

S.J. HAVISON [1969], *A Chebyshev theorem for the approximation of a function of two variables by sums $\phi(x)+\psi(y)$*, Izv. Akad. Nauk SSSR Ser. Mat. **33**, pp. 650–666 (Russian). [English translation: Math. USSR-Izv. **3** (1969), pp. 617–632.

T. HEDBERG [1971], *The Kolmogorov superposition theorem*, Appendix II in Topics in Approximation Theory, H.S. Shapiro, Springer-Verlag Lecture Notes in Mathematics Vol.**187**, pp. 267–275.

D. HILBERT [1901], *Mathematical problems*, Bull. Amer. Math. Soc. **8**, pp. 437–479.

A.S.B. HOLLAND, B.N. SAHNEY AND J. TZIMBALARIO [1976], *On best simultaneous approximation*, J. Indian Math. Soc. **40**, pp. 69–73.

S.M. HOLLAND, W.A. LIGHT AND L.J. SULLEY [1982], *On proximinality in $L_1(T\times S)$*, Proc. Amer. Math. Soc. **86**, pp. 279–282.

R.B. HOLMES [1972], *A course in optimization and best approximation*, Lecture Notes in Math. Vol.**257**, Springer-Verlag, New York.

J.R. ISBELL AND Z. SEMADENI [1963], *Projection constants and spaces of continuous functions*, Trans. Amer. Math. Soc. **107**, pp. 38–48.

G.J.O. JAMESON AND A. PINKUS [1983], *Positive and minimal projections in function spaces*, J. Approx. Theory **37**, pp. 182–195.

M. KADETS AND V. ZAMYATIN [1968], *Chebyshev centers in the space $C[a,b]$*, Teor. Funsktsiĭ Funkstional. Anal. i Prilozhen **7**, pp. 20–26 (Russian).

P. KERGIN [1980], *A natural interpolation of C^k-functions*, J. Approx. Theory **19**, pp. 278–293.

R. KHALIL [1983], *Best approximation in $L^p(I,X)$*, Math. Proc. Cambridge Philos. Soc. **94**, pp. 277–279.

A.N. KOLMOGOROV [1957], *On the representation of continuous functions of several variables by superpositions of continuous functions of one variable and addition*, Dokl. Akad. Nauk SSSR **114**, pp. 679–681.

P.J. LAURENT AND P.D. TUAN [1970], *Global approximation of a compact set by elements of a convex set in a normed space*, Numer. Math. **15**, pp. 137–150.

W.A. LIGHT [1980], *Central proximity maps*, Approximation Theory III, E.W. Cheney, ed., Academic Press, New York, pp. 589–594.

——— [1981], *Convergence of the Diliberto-Straus algorithm in* $L_1(X\times Y)$, Numer. Funct. Anal. Optim. **3**, pp. 137–146.

——— [1983], *The Diliberto-Straus algorithm in* $L_1(X\times Y)$, J. Approx. Theory **38**, 1–8.

——— [1986], *A note on proximinality in* $C(S\times T)$ *with the* L_1*-norm*, Proc. Edinburgh Math. Soc., to appear.

W.A. LIGHT AND E.W. CHENEY [1980], *On the approximation of a bivariate function by the sum of univariate functions*, J. Approx. Theory **29**, pp. 305–322.

——— [1981], *Some best-approximation theorems in tensor-product spaces*, Math. Proc. Cambridge Philos. Soc. **89**, pp. 385–390.

——— [1985], *Approximation theory in tensor product spaces*, Lecture Notes in Math. **1169**, Springer, New York.

K.-P. LIN [1974], *Simultaneous approximation of compact sets by elements of convex sets in normed linear spaces*, J. Approx. Theory **12**, pp. 332–351.

G.G. LORENTZ [1966], *Approximation of Functions*, Holt, Rinhart and Winston, New York. Reprinted by Chelsea Publishing Co., New York, 1986.

J. MACH [1979-a], *Best simultaneous approximation*, Math. Ann. **240**, pp. 157–164.

——— [1979-b], *On the existence of best simultaneous approximation*, J. Approx. Theory **25**, pp. 258–265.

——— [1982], *Chebyshev centers of compact sets with respect to Stone-Weierstrass subspaces*, CAT Report #**15**, Texas A & M University.

J.C. MAIRHUBER [1956], *On Haar's theorem concerning Chebyshev approximation problems having unique solutions*, Proc. Amer. Math. Soc. **7**, pp. 609–615.

A.W. MARSHALL AND I. OLKIN [1968], *Scaling of matrices to achieve specified row and column sums*, Numer. Math. **12**, pp. 83–90.

J. MEINGUET [1970], *Optimal approximation and interpolation in normed spaces*, Numerical Approximation to Functions and Data, J. Hayes, ed., Athlone Press, London, pp. 143–157.

——— [1979-a], *An intrinsic approach to multivariate spline interpolation at arbitrary points*, Polynomial and Spline Approximation, B.H. Sahney, ed., Reidel, Dordrecht, pp. 163–190.

——— [1979-b], *Multivariate interpolation at arbitrary points made simple*, Z. Angew. Math. Phys. **30**, pp. 292–304.

C.A. MICCHELLI [1980], *A constructive approach to Kergin interpolation*, Rocky Mountain J. Math. **10**, pp. 485–497.

——— [1985], *Interpolation of scattered data: distance matrices and conditionally positive definite functions*, Research Report, IBM Laboratories, Yorktown Heights, N.Y.

——— [1986], *Algebraic aspects of interpolation* in Approximation Theory Short Course, Amer. Math. Soc.

C.A. MICCHELLI AND A. PINKUS [1977], *Best mean approximation to a* 2*-dimensional kernel by tensor products*, Bull. Amer. Math. Soc. **83**, pp. 400–402.

P.D. MILMAN [1977], *On best simultaneous approximation in normed linear spaces*, J. Approx. Theory **20**, pp. 223–238.

J. VON. NEUMANN [1950], *Functional Operators*, Vol. II, Annals of Math. Studies **22**, Princeton University Press, Princeton, N.J.

JU. P. OFMAN [1961], *Best approximation of functions of two variables by functions of the form* $\phi(x)+\psi(y)$, Izv. Akad. Nauk SSSR Ser. Mat. **25**, pp. 239–252. [English translation: Amer. Math. Soc. Translations Series 2, Vol.**44**, pp. 12-28].

E.E. OSBORNE [1960], *On preconditioning of matrices*, J. Assoc. Comput. Mach.**7**, pp. 338–345.

R.W. OWENS [1983], *A characterization of an element of best simultaneous approximation*, Numer. Funct. Anal. Optim. **6**, pp. 447–456.

K.R. PARTHASARATHY AND K. SCHMIDT [1972], *Positive Definite Kernels, Continuous Tensor Products, and Central Limit Theorems of Probability Theory*, Lecture Notes in Math. **272**, Springer-Verlag, Berlin.

T. PARTHASARATHY [1972], *Selection Theorems and Their Applications*, Lecture Notes in Math. **263**, Springer-Verlag, New York.

A. PERSSON [1969], *On some properties of p-nuclear and p-integral operators*, Studia Math. **33**, pp. 213–222.

A. PINKUS [1985], *n-Widths in Approximation Theory*, Springer-Verlag, New York.

J.B. PROLLA [1983], *Nonexpansive mappings and Chebyshev centers in (AL)-spaces*, Approximation Theory, C.K. Chui, L.L. Schumaker, and J.D. Ward, eds., Academic Press, New York, pp. 663–666.

J.R. RESPESS, JR. AND E.W. CHENEY [1982-a], *Best approximation problems in tensor-product spaces*, Pacific J. Math. **102**, pp. 437–446.

——— [1982-b], *On Lipschitzian proximity maps*, in Nonlinear Analysis and Application, S.P. Singh and J.H. Burry, eds., pp. 73–85, Lecture Notes in Pure and Applied Mathematics Vol.**80**, Marcel Dekker, New York.

T.J. RIVLIN AND R.J. SIBNER [1965], *The degree of approximation of certain functions of two variables by the sum of functions of one variable*, Amer. Math. Monthly **72**, pp. 1101–1103.

U.G. ROTHBLUM AND H. SCHNEIDER [1980], *Characterization of optimal scalings of matrices*, Math. Programming **19**, pp. 121–136.

——— [1982], *Characterizations of extreme normalized circulations satisfying linear constraints*, Linear Algebra Appl. **46**, pp. 61–72.

B.N. SAHNEY AND S.P. SINGH [1982], *On best simultaneous Chebyshev approximation with additive weight functions and related results*, Nonlinear Analysis and Applications, S.P. Singh and J.H. Burry, eds., Marcel Dekker, Inc., New York, pp. 443–463.

P. SAPHAR [1970], *Produits tensoriels d'espaces de Banach et classes d'applications linéaires*, Studia Math. **38**, pp. 71–100.

JU.A. SASKIN [1967], *Interpolation families of functions and imbeddings of sets in Euclidean and projective spaces*, Dokl. Akad. Nauk SSSR **174**, pp. 1030–1032 (Russian). [English translation: Sov. Math. Dokl. **8**, pp. 722–725.]

E. SCHMIDT [1907], *Zur theorie der linearen und nichtlinearen integralgleichungen I Teil*, Math. Ann. **63**, pp. 433–476.

L.L. SCHUMAKER [1976], *Fitting surfaces to scattered data*, Approximation Theory, III, G.G. Lorentz, C.K. Chui, and L.L. Schumaker, eds., Academic Press, New York, pp. 203–268.

Z. SEMADENI [1971], "Banach Spaces of Continuous Functions", Polish Scientific Publishers, Warsaw.

D. SHEPARD [1968], *A two-dimensional interpolation function for irregularly spaced data*, Proc. 23rd Nat. Conf. Assoc. for Computing Machinery, pp. 517–524.

I. SINGER [1970], "Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces", Springer-Verlag, Berlin.

——— [1974], "The Theory of Best Approximation and Functional Analysis", Soc. for Industrial and Appl. Math., Philadelphia.

P.W. SMITH AND J.D. WARD [1975-a], *Restricted centers in subalgebras of $C(X)$*, J. Approx. Theory **15**, pp. 54–59.

——— [1975-b], *Restricted centers in $C(\Omega)$*, Proc. Amer. Math. Soc. **48**, pp. 165–172.

B. STEINHAUS [1984], *L_1-Approximation with blending functions*, preprint.

F. SULLIVAN [1975], *A generalization of best approximation operators*, Ann. Mat. Pura Appl. **107**, pp. 245–261.

A.E. TARWATER [1985], *A parameter study of Hardy's multiquadric method for scattered data interpolation*, Research Report UCRL-53670, Lawrence Livermore National Laboratory.

H.C. THACHER AND W.E. MILNE [1960], *Interpolation in several variables*, J. Soc. Indust. Appl. Math. **8**, pp. 33–42.

J.A. TOMLIN [1975], *On scaling linear programming problems*, Math. Programming Study **4**, pp. 146–166.

F. TREVES [1967], "Topological Vector Spaces", Academic Press, New York.

G.A. WATSON [1982], *Some characterizations of best mixed norm approximation*, J. Approx. Theory **36**, pp. 32–42.

H. WERNER [1980], *Remarks on Newton type multivariate interpolation for subsets of grids*, Computing **25**, pp. 181–191.

V.N. ZAMYATIN [1973], *Relative Chebyshev centers in the space of continuous functions*, Soviet Math. Dokl. **14**, pp. 610–614.

(continued from inside front cover)

MORTON E. GURTIN, *Topics in Finite Elasticity*
THOMAS G. KURTZ, *Approximation of Population Processes*
JERROLD E. MARSDEN, *Lectures on Geometric Methods in Mathematical Physics*
BRADLEY EFRON, *The Jackknife, the Bootstrap, and Other Resampling Plans*
M. WOODROOFE, *Nonlinear Renewal Theory in Sequential Analysis*
D. H. SATTINGER, *Branching in the Presence of Symmetry*
R. TEMAM, *Navier-Stokes Equations and Nonlinear Functional Analysis*
MIKLÓS CSÖRGO, *Quantile Processes with Statistical Applications*
J. D. BUCKMASTER and G. S. S. LUDFORD, *Lectures on Mathematical Combustion*
R. E. TARJAN, *Data Structures and Network Algorithms*
PAUL WALTMAN, *Competition Models in Population Biology*
S. R. S. VARADHAN, *Large Deviations and Applications*
KIYOSI ITÔ, *Foundations of Stochastic Differential Equations in Infinite Dimensional Spaces*
ALAN C. NEWELL, *Solitons in Mathematics and Physics*
PRANAB KUMAR SEN, *Theory and Applications of Sequential Nonparametrics*
LÁSZLÓ LOVÁSZ, *An Algorithmic Theory of Numbers, Graphs and Convexity*
E. W. CHENEY, *Multivariate Approximation Theory: Selected Topics*