1

# Stochastic Tools in Mathematics and Science

Second Edition

Springer

# Surveys and Tutorials in the Applied Mathematical Sciences

**Volume 1**

*Series Editors*

S.S. Antman, J.E. Marsden, L. Sirovich

For other titles published in this series, go to
www.springer.com/series/7219

## Surveys and Tutorials in the Applied Mathematical Sciences

Volume 1

*Editors*
S.S. Antman, J.E. Marsden, L. Sirovich

Mathematics is becoming increasingly interdisciplinary and developing stronger interactions with fields such as biology, the physical sciences, and engineering. The rapid pace and development of the research frontiers has raised the need for new kinds of publications: short, up-to-date, readable tutorials and surveys on topics covering the breadth of the applied mathematical sciences. The volumes in this series are written in a style accessible to researchers, professionals, and graduate students in the sciences and engineering. They can serve as introductions to recent and emerging subject areas and as advanced teaching aids at universities. In particular, this series provides an outlet for material less formally presented and more anticipatory of needs than finished texts or monographs, yet of immediate interest because of the novelty of their treatments of applications, or of the mathematics being developed in the context of exciting applications. The series will often serve as an intermediate stage of publication of materials which, through exposure here, will be further developed and refined to appear later in one of Springer's more formal series in applied mathematics.

Alexandre J. Chorin and Ole H. Hald

# Stochastic Tools in Mathematics and Science

Second Edition

Springer

Alexandre J. Chorin
Department of Mathematics
University of California at Berkeley,
970 Evans Hall
Berkeley, CA 94720-3840
USA
chorin@math.berkeley.edu

Ole H. Hald
Department of Mathematics
University of California at Berkeley,
970 Evans Hall
Berkeley, CA 94720-3840
USA
hald@math.berkeley.edu



ISBN 978-1-4419-1001-1 e-ISBN 978-1-4419-1002-8
DOI 10.1007/978-1-4419-1002-8
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009930418

Mathematics Subject Classification (2000): 34F05, 35-01, 35-04, 35K05, 35K20, 35Q7260-01, 60-04, 60H10, 60H30, 60H3580-01, 82-04, 82-08, 82B05, 82B26, 82B27, 82B28, 82B31, 82B8065-01, 65C05, 65C20, 65C35, 65C50



Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Prefaces

## Preface to the Second Edition

In preparing the second edition we have tried to improve and clarify the presentation, guided in part by the many comments we have received, and also to make the various arguments more precise, as far as we could while keeping this book short and introductory.

There are many dozens of small changes and corrections. The more substantial changes from the first edition include: a completely rewritten discussion of renormalization, and significant revisions of the sections on prediction for stationary processes, Markov chain Monte Carlo, turbulence, and branching random motion. We have added a discussion of Feynman diagrams to the section on Wiener integrals, a discussion of fixed points to the section on the central limit theorem, a discussion of perfect gases and the equivalence of ensembles to the section on entropy and equilibrium. There are new figures, new exercises, and new references.

We are grateful to the many people who have talked with us or written to us with comments and suggestions for improvement. We are also grateful to Valerie Heatlie for her patient help in putting the revised manuscript together.

Alexandre J. Chorin
Ole H. Hald
Berkeley, California
March, 2009

## Preface to the First Edition

This book started out as a set of lecture notes for a first-year graduate course on the "stochastic methods of applied mathematics" at the Department of Mathematics of the University of California at Berkeley. The course was started when the department asked a group of its former students who had gone into nonacademic jobs, in national labs and industry, what they actually did in their jobs, and found that most of them did stochastic things that had not appeared anywhere in our graduate course lineup; over the years the course changed as a result of the comments and requests of the students, who have turned out to be a mix of mathematics students and students from the sciences and engineering. The course has not endeavored to present a full, rigorous theory of probability and its applications, but rather to provide mathematics students with some inkling of the many beautiful applications of probability, as well as introduce the nonmathematical students to the general ideas behind methods and tools they already use. We hope that the book too can accomplish these tasks.

We have simplified the mathematical explanations as much as we could everywhere we could. On the other hand, we have not tried to present applications in any detail either. The book is meant to be an introduction, hopefully an easily accessible one, to the topics on which it touches.

The chapters in the book cover some background material on least squares and Fourier series, basic probability (with Monte Carlo methods, Bayes' theorem, and some ideas about estimation), some applications of Brownian motion, stationary stochastic processes (the Khinchin theorem, an application to turbulence, prediction for time series and data assimilation), equilibrium statistical mechanics (including Markov chain Monte Carlo), and time-dependent statistical mechanics (including optimal prediction). The leitmotif of the book is conditional expectation (introduced in a drastically simplified way) and its uses in approximation, prediction, and renormalization. All topics touched upon come with immediate applications; there is an unusual emphasis on time-dependent statistical mechanics and the Mori-Zwanzig formalism, in accordance with our interests and as well as our convictions. Each chapter is followed by references; it is, of course, hopeless to try to provide a full bibliography of all the topics included here; the bibliographies are simply lists of books and papers we have actually used in preparing notes and should be seen as acknowledgments as well as suggestions for further reading in the spirit of the text.

We thank Dr. David Bernstein, Dr. Maria Kourkina-Cameron, and Professor Panagiotis Stinis, who wrote down and corrected the notes on which this book is based and then edited the result; the book would not have existed without them. We are profoundly indebted to many wonderful collaborators on the topics covered in this book, in particular Professor G.I. Barenblatt, Dr. Anton Kast, Professor Raz Kupferman, and Professor Panagiotis Stinis, as well as Dr. John Barber, Dr. Alexander Gottlieb, Dr. Peter Graf, Dr. Eugene Ingerman, Dr. Paul Krause, Professor Doron Levy, Professor Kevin Lin, Dr. Paul Okunev, Dr. Benjamin Seibold, and Professor Mayya Tokman; we have learned from all of them (but obviously not enough) and greatly enjoyed their friendly collaboration. We also thank the students in the Math 220 classes at the University of California, Berkeley, and Math 280 at the University of California, Davis, for their comments, corrections, and patience, and in particular Ms. K. Schwarz, who corrected errors and obscurities. We are deeply grateful to Ms. Valerie Heatlie, who performed the nearly-Sisyphean task of preparing the various typescripts with unflagging attention and good will. Finally, we are thankful to the US Department of Energy and the National Science Foundation for their generous support of our endeavors over the years.

Alexandre J. Chorin
Ole H. Hald
Berkeley, California
September, 2005

# Contents

CHAPTER 1

# Preliminaries

## 1.1. Least Squares Approximation

Let $V$ be a vector space with vectors $u, v, w, \dots$ and scalars $\alpha, \beta, \dots$. The space $V$ is an inner product space if one has defined a function $(\cdot, \cdot)$ from $V \times V$ to the reals (if the vector space is real) or to the complex (if $V$ is complex) such that for all $u, v \in V$ and all scalars $\alpha$, the following conditions hold:

$$\begin{aligned} (u, v) &= \overline{(v, u)}, \\ (u + v, w) &= (u, w) + (v, w), \\ (\alpha u, v) &= \alpha (u, v), \\ (v, v) &\geq 0, \\ (v, v) &= 0 \Leftrightarrow v = 0, \end{aligned} \tag{1.1}$$

where the overbar denotes the complex conjugate. Two elements $u, v$ such that $(u, v) = 0$ are said to be orthogonal.

The most familiar inner product space is $\mathbb{R}^n$ with the Euclidean inner product. If $u = (u_1, u_2, \dots, u_n)$ and $v = (v_1, v_2, \dots, v_n)$, then

$$(u, v) = \sum_{i=1}^{n} u_i v_i.$$

Another inner product space is $C[0, 1]$, the space of continuous functions on $[0, 1]$, with $(f, g) = \int_0^1 f(x) g(x)\, dx$.

When you have an inner product, you can define a norm, the "$L_2$ norm", by

$$\|v\| = \sqrt{(v, v)}.$$

A.J. Chorin and O.H. Hald, *Stochastic Tools in Mathematics and Science*, Surveys and Tutorials in the Applied Mathematical Sciences 1, DOI: 10.1007/978-1-4419-1002-8_1,

This has the following properties, which can be deduced from the properties of the inner product:

$$\|\alpha v\| = |\alpha|\|v\|,$$
$$\|v\| \geq 0,$$
$$\|v\| = 0 \Leftrightarrow v = 0,$$
$$\|u+v\| \leq \|u\| + \|v\|.$$

The last, called the triangle inequality, follows from the Schwarz inequality

$$|(u,v)| \leq \|u\|\|v\|.$$

In addition to these three properties, common to all norms, the $L_2$ norm has the "parallelogram property" (so called because it is a property of parallelograms in plane geometry)

$$\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2),$$

which can be verified by expanding the inner products.

Let $\{u_n\}$ be a sequence in $V$.

DEFINITION. A sequence $\{u_n\}$ is said to converge to $\hat{u} \in V$ if $\|u_n - \hat{u}\| \to 0$ as $n \to \infty$ (i.e., for any $\epsilon > 0$, there exists some $N \in \mathbb{N}$ such that $n > N$ implies $\|u_n - \hat{u}\| < \epsilon$).

DEFINITION. A sequence $\{u_n\}$ is a Cauchy sequence if given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $m, n > N$ $\|u_n - u_m\| < \epsilon$.

A sequence that converges is a Cauchy sequence, although the converse is not necessarily true. If the converse is true for all Cauchy sequences in a given inner product space, then the space is called complete. All of the spaces we work with from now on are complete. Examples are $\mathbb{R}^n$, $\mathbb{C}^n$, $L_2$.

A few more definitions from real analysis:

DEFINITION. An open ball centered at $x$ with radius $r > 0$ is the set $B_r(x) = \{u : \|u - x\| < r\}$.

DEFINITION. A set $S$ is open if for all $x \in S$, there exists an open ball $B_r(x)$ such that $B_r(x) \subset S$.

DEFINITION. A set $S$ is closed if every convergent sequence $\{u_n\}$ such that $u_n \in S$ for all $n$ converges to an element of $S$.

An example of a closed set is the closed interval $[0,1] \subset \mathbb{R}$. An example of an open set is the open interval $(0,1) \subset \mathbb{R}$. The complement of an open set is closed, and the complement of a closed set is open. The empty set is both open and closed and so is $\mathbb{R}^n$.

Given a set $S$ and some point $b$ outside of $S$, we want to determine under what conditions there is a point $\hat{b} \in S$ closest to $b$. Let $d(b, S) = \inf_{x \in S} \|x - b\|$ be the distance from $b$ to $S$. The quantity on the right of this definition is the greatest lower bound of the set of numbers $\|x - b\|$, and its existence is guaranteed by the properties of the real number system. What is not guaranteed in advance, and must be proved here, is the existence of an element $\hat{b}$ that satisfies $\|\hat{b} - b\| = d(b, S)$. To see the issue, take $S = (0, 1) \subset \mathbb{R}$ and $b = 2$; then $d(b, S) = 1$, yet there is no point $\hat{b} \in (0, 1)$ such that $\|\hat{b} - 2\| = 1$.

THEOREM 1.1. *If $S$ is a closed linear subspace of $V$ and $b$ is an element of $V$, then there exists $\hat{b} \in S$ such that $\|\hat{b} - b\| = d(b, S)$.*

PROOF. There exists a sequence of elements $\{u_n\} \subset S$ such that $\|b - u_n\| \to d(b, S)$ by definition of the greatest lower bound. We now show that this sequence is a Cauchy sequence.

From the parallelogram law we have

$$\left\|\frac{1}{2}(b - u_m)\right\|^2 + \left\|\frac{1}{2}(b - u_n)\right\|^2 = \frac{1}{2}\left\|b - \frac{1}{2}(u_n + u_m)\right\|^2 + \frac{1}{8}\|u_n - u_m\|^2. \tag{1.2}$$

$S$ is a vector space; therefore,

$$\frac{1}{2}(u_n + u_m) \in S \Rightarrow \left\|b - \frac{1}{2}(u_n + u_m)\right\|^2 \geq d^2(b, S).$$

Then since $\|b - u_n\| \to d(b, S)$, we have

$$\left\|\frac{1}{2}(b - u_n)\right\|^2 \to \frac{1}{4} d^2(b, S).$$

From (1.2),

$$\|u_n - u_m\| \to 0,$$

and thus $\{u_n\}$ is a Cauchy sequence by definition; our space is complete and therefore this sequence converges to an element $\hat{b}$ in this space. $\hat{b}$ is in $V$ because $V$ is closed. Consequently

$$\|\hat{b} - b\| = \lim \|u_n - b\| = d(b, S).$$

∎

We now wish to describe further the relation between $b$ and $\hat{b}$.

THEOREM 1.2. *Let $S$ be a closed linear subspace of $V$, let $x$ be any element of $S$, $b$ any element of $V$, and $\hat{b}$ an element of $S$ closest to $b$. Then*

$$(x - \hat{b}, b - \hat{b}) = 0.$$

PROOF. If $x = \hat{b}$ we are done. Else set

$$\theta(x - \hat{b}) - (b - \hat{b}) = \theta x + (1 - \theta)\hat{b} - b = y - b.$$

Since $y$ is in $S$ and $\|y - b\| \geq \|\hat{b} - b\|$, we have

$$\begin{aligned}\|\theta(x - \hat{b}) - (b - \hat{b})\|^2 &= \theta^2\|x - \hat{b}\|^2 - 2\theta(x - \hat{b},\, b - \hat{b}) + \|b - \hat{b}\|^2 \\ &\geq \|b - \hat{b}\|^2.\end{aligned}$$

Thus $\theta^2\|x - \hat{b}\|^2 - 2\theta(x - \hat{b},\, b - \hat{b}) \geq 0$ for all $\theta$. The left hand side attains its minimum value when $\theta = (x-\hat{b},\, b-\hat{b})/\|x-\hat{b}\|^2$ in which case $-(x - \hat{b},\, b - \hat{b})^2/\|x - \hat{b}\|^2 \geq 0$. This implies that $(x - \hat{b},\, b - \hat{b}) = 0$. ■

THEOREM 1.3. *$(b - \hat{b})$ is orthogonal to $x$ for all $x \in S$.*

PROOF. By Theorem 1.2, $(x - \hat{b}, b - \hat{b}) = 0$ for all $x \in S$. When $x = 0$ we have $(\hat{b}, b - \hat{b}) = 0$. Thus $(x, b - \hat{b}) = 0$ for all $x$ in $S$. ■

COROLLARY 1.4. *If $S$ is a closed linear subspace, then $\hat{b}$ is unique.*

PROOF. Let $b = \hat{b} + n = \hat{b}_1 + n_1$, where $n$ is orthogonal to $\hat{b}$ and $n_1$ is orthogonal to $\hat{b}_1$. Therefore,

$$\begin{aligned}\hat{b} - \hat{b}_1 \in S &\Rightarrow (\hat{b} - \hat{b}_1, n_1 - n) = 0 \\ &\Rightarrow (\hat{b} - \hat{b}_1, \hat{b} - \hat{b}_1) = 0 \\ &\Rightarrow \hat{b} = \hat{b}_1.\end{aligned}$$

■

One can think of $\hat{b}$ as the orthogonal projection of $b$ on $S$ and write $\hat{b} = \mathbb{P}b$, where the projection $\mathbb{P}$ is defined by the foregoing discussion.

We will now give a few applications of the above results.

EXAMPLE. Consider a matrix equation $Ax = b$, where $A$ is an $m\times n$ matrix and $m > n$. This kind of problem arises when one tries to fit a large set of data by a simple model. Assume that the columns of $A$ are linearly independent. Under what conditions does the system have a solution? To clarify ideas, consider the $3 \times 2$ case:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

Let $A_1$ denote the first column vector of $A$, $A_2$ the second column vector, etc. In this case,

$$A_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix}, \quad A_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix}.$$

If $Ax = b$ has a solution, then one can express $b$ as a linear combination of $A_1, A_2, \ldots A_m$; for example, in the $3 \times 2$ case, $x_1A_1 + x_2A_2 = b$. If $b$ does not lie in the column space of $A$ (the set of all linear combinations of the columns of $A$), then the problem has no solution. It is often reasonable to replace the unsolvable problem by the solvable problem $A\hat{x} = \hat{b}$, where $\hat{b}$ is as close as possible to $b$ and yet does lie in the column space of $A$. We know from the foregoing that the "best $\hat{b}$" is such that $b - \hat{b}$ is orthogonal to the column space of $A$. This is enforced by the $m$ equations

$$(A_1, \hat{b} - b) = 0, \quad (A_2, \hat{b} - b) = 0, \quad \ldots \quad , (A_m, \hat{b} - b) = 0.$$

Since $\hat{b} = A\hat{x}$, we obtain the equation

$$A^T(A\hat{x} - b) = 0 \quad \Rightarrow \quad \hat{x} = (A^TA)^{-1}A^Tb.$$

One application of the above is to "fit" a line to a set of points on the Euclidean plane. Given a set of points, $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ that come from some experiment and that we believe would lie on a straight line if it were not for experimental error, what is the line that "best approximates" these points? We hope that if it were not for the errors, we would have $y_i = ax_i + b$ for all $i$ and for some fixed $a$ and $b$; so we seek to solve a system of equations

$$\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

Example. Consider the system of equations given by $Ax = b$, where $A$ is an $n \times m$ matrix and $n < m$ (there are more unknowns than equations). The system has infinitely many solutions. Suppose you want the solution of smallest norm; this problem arises when one tries to find the most likely solution to an underdetermined problem.

Before solving this problem, we need some preliminaries.

Definition. $S \subset V$ is an affine subspace if $S = \{y : y = x + c, c \neq 0, x \in X\}$, where $X$ is a closed linear subspace of $V$. Note that $S$ is not a linear subspace.

LEMMA 1.5. *If $S$ is an affine subspace and $b' \notin S$, then there exists $\hat{x} \in X$ such that $d(b', S) = \|\hat{x} + c - b'\|$. Furthermore, $\hat{x} - (b' - c)$ is orthogonal to $x$ for all $x \in X$. (Note that here we use $b'$ instead of $b$, to avoid confusion with the system's right-hand side.)*

PROOF. We have $S = \{y : y = x + c, c \neq 0, x \in X\}$, where $X$ is a closed linear subspace of $V$. Now,

$$\begin{aligned} d(b', S) &= \inf_{y \in S} \|y - b'\| = \inf_{x \in X} \|x + c - b'\| \\ &= \inf_{x \in X} \|x - (b' - c)\| = d(b' - c, X) \\ &= \|\hat{x} - (b' - c)\| = \|\hat{x} + c - b'\|. \end{aligned}$$

The point $\hat{x} \in X$ exists since $X$ is a closed linear subspace. It follows from Theorem 1.3 that $\hat{x} - (b' - c)$ is orthogonal to $X$. Note that the distance between $S$ and $b'$ is the same as that between $X$ and $b' - c$. ∎

From the proof above, we see that $\hat{x} + c$ is the element of $S$ closest to $b'$. For the case $b' = 0$, we find that $\hat{x} + c$ is orthogonal to $X$.

Now we return to the problem of finding the "smallest" solution of an underdetermined problem. Assume $A$ has "maximal rank"; that is, $m$ of the column vectors of $A$ are linearly independent. We can write the solutions of the system as $x = x_0 + z$, where $x_0$ is a particular solution and $z$ is a solution of the homogeneous system $Az = 0$. So the solutions of the system $Ax = b$ form an affine subspace. As a result, if we want to find the solution with the smallest norm (i.e., closest to the origin) we need to find the element of this affine subspace closest to $b' = 0$. From the above, we see that such an element must satisfy two properties. First, it has to be an element of the affine subspace (i.e., a solution to the system $Ax = b$) and second, it has to be orthogonal to the linear subspace $X$, which is the null space of $A$ (the set of solutions of $Az = 0$). Now consider $x' = A^T(AA^T)^{-1}b$; this vector lies in the affine subspace of the solutions of $Ax = b$, as one can check by multiplying it by $A$. Furthermore, it is orthogonal to every vector in the space of solutions of $Az = 0$ because $(A^T(AA^T)^{-1}b, z) = ((AA^T)^{-1}b, Az) = 0$. This is enough to make $x'$ the unique solution of our problem.

## 1.2. Orthonormal Bases

The problem presented in the previous section, of finding an element in a closed linear space that is closest to a vector outside the space, lies in the framework of approximation theory, where we are given a function (or a vector) and try to find an approximation to it as a linear combination of given functions (or vectors). This is done by requiring that the norm of the error (difference between the given function and the approximation) be minimized. In what follows, we shall find coefficients for this optimal linear combination.

DEFINITION. Let $S$ be a linear vector space. A collection of $m$ vectors $\{u_i\}_{i=1}^m$ belonging to $S$ are linearly independent if and only if $\lambda_1 u_1 + \cdots + \lambda_m u_m = 0$ implies $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 0$.

DEFINITION. Let $S$ be a linear vector space. A collection $\{u_i\}_{i=1}^m$ of vectors belonging to $S$ is called a basis of $S$ if $\{u_i\}$ are linearly independent and any vector in $S$ can be written as a linear combination of them.

Note that the number of elements of a basis can be finite or infinite depending on the space.

THEOREM 1.6. *Let $S$ be an $m$-dimensional linear inner-product space with $m$ finite. Then any collection of $m$ linearly independent vectors of $S$ is a basis.*

DEFINITION. A set of vectors $\{e_i\}_{i=1}^m$ is orthonormal if the vectors are mutually orthogonal and each has unit length (i.e., $(e_i, e_j) = \delta_{ij}$, where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise).

The set of all the linear combinations of the vectors $\{u_i\}$ is called the span of $\{u_i\}$ and is written as $\text{Span}\{u_1, u_2, \ldots, u_m\}$.

Suppose we are given a set of vectors $\{e_i\}_{i=1}^m$ that are an orthonormal basis for a subspace $S$ of a real vector space. If $b$ is an element outside the space, we want to find the element $\hat{b} \in S$, where $\hat{b} = \sum_{i=1}^m c_i e_i$ such that $\|b - \sum_{i=1}^m c_i e_i\|$ is minimized. Specifically, we have

$$\begin{aligned}
\left\| b - \sum_{i=1}^m c_i e_i \right\|^2 &= \left( b - \sum_{i=1}^m c_i e_i \,, b - \sum_{j=1}^m c_j e_j \right) \\
&= (b, b) - 2\sum_{i=1}^m c_i (b, e_i) + \left( \sum_{i=1}^m c_i e_i \,, \sum_{j=1}^m c_j e_j \right) \\
&= (b, b) - 2\sum_{i=1}^m c_i (b, e_i) + \sum_{i,j=1}^m c_i c_j (e_i, e_j) \\
&= (b, b) - 2\sum_{i=1}^m c_i (b, e_i) + \sum_{i=1}^m c_i^2 \\
&= \|b\|^2 - \sum_{i=1}^m (b, e_i)^2 + \sum_{i=1}^m (c_i - (b, e_i))^2,
\end{aligned}$$

where we have used the orthonormality of the $e_i$ to simplify the expression. As is readily seen, the norm of the error is a minimum when