

# Preface to the Second Edition

If this book were my child, I would say that it has grown up after passing through a painful adolescence. It has rebelled at times over my indecision. While it has aspired to greatness, I have aspired to closure. Although the second edition of *Optimization* occasionally demanded more exertion than I could muster, writing it has broadened my intellectual horizons. I hope my own struggles to reach clarity will translate into an easier path for readers.

The book's stress on mathematical fundamentals continues. Indeed, I was tempted to re-title it *Mathematical Analysis and Optimization*. I resisted this temptation because my ultimate goal is still to teach optimization theory. Nonetheless, there is a new chapter on the gauge integral and expanded treatments of differentiation and convexity. The focus remains on finite-dimensional optimization. The sole exception to this rule occurs in the new chapter on the calculus of variations. In my view functional analysis is just too high a rung on the ladder of mathematical abstraction.

Covering all of optimization theory is simply out of the question. Even though the second edition is more than double the length of the first, many important topics are omitted. The most grievous omissions are the simplex algorithm of linear programming and modern interior point methods. Fortunately, there are many admirable books devoted to these subjects. My development of adaptive barrier methods and exact penalty methods also partially compensates.

In addition to the two chapters on integration and the calculus of variations, four new chapters treat block relaxation (block descent and block ascent) and various advanced topics in the convex calculus, including the

Fenchel conjugate, subdifferentials, duality, feasibility, alternating projections, projected gradient methods, exact penalty methods, and Bregman iteration. My own interests in data mining and biological applications have dictated the nature of these chapters. High-dimensional problems are driving the discipline of optimization. These are qualitatively different from traditional problems, and standard algorithms such as Newton's method are often impractical. Penalization, model sparsity, and the MM algorithm now assume dominant roles. Fortunately, many of the challenging modern problems can also be phrased as convex programs.

In the first edition I eschewed the convention of setting vectors and matrices in boldface type. In the second edition I embrace it. Although this decision improves readability, it carries with it some residual ambiguity. The main difficulty lies in distinguishing constant vectors and matrices from vector and matrix-valued functions. In general, I have elected to set functions in ordinary type even when they are vector or matrix valued. The exceptions occur in the calculus of variations, where functions are considered vectors in infinite-dimensional spaces. Thus, a function appears in ordinary type when its argument is displayed and in boldface type when its argument is omitted.

Many people have helped me prepare this second edition. Hua Zhou and Tongtong Wu, my former postdoctoral fellows, and Eric Chi, my current postdoctoral fellow, deserve special credit. Without their assistance, the book would have been intellectually duller and graphically drearier. I would also like to thank my former doctoral students David Alexander, David Hunter, Mary Sehl, and Jinjin Zhou for proofreading and critiquing the new material. The students in my optimization class checked most of the exercises. I am indebted to Forrest Crawford, Gabriela Cybis, Gary Evans, Mitchell Johnson, Wesley Kerr, Kevin Keys, Omid Kohannim, Lewis Lee, Matthew Levinson, Lae Un Kim, John Ranola, and Moses Wilkes for their help.

Finally, let me report on my daughters Maggie and Jane, to whom this book is dedicated. Maggie is now embarked on a postdoctoral fellowship in medical ethics at Macquarie University in Sydney, Australia. Jane is completing her dissertation in biostatistics at the University of Washington. Assimilating their scholarship will keep me young for many years to come.

# Preface to the First Edition

This foreword, like many forewords, was written afterwards. That is just as well because the plot of the book changed during its creation. It is painful to recall how many times classroom realities forced me to shred sections and start anew. Perhaps such adjustments are inevitable. Certainly I gained a better perspective on the subject over time. I also set out to teach optimization theory and wound up teaching mathematical analysis. The students in my classes are no less bright and eager to learn about optimization than they were a generation ago, but they tend to be less prepared mathematically. So what you see before you is a compromise between a broad survey of optimization theory and a textbook of analysis. In retrospect, this compromise is not so bad. It compelled me to revisit the foundations of analysis, particularly differentiation, and to get right to the point in optimization theory.

The content of courses on optimization theory varies tremendously. Some courses are devoted to linear programming, some to nonlinear programming, some to algorithms, some to computational statistics, and some to mathematical topics such as convexity. In contrast to their gaps in mathematics, most students now come well trained in computing. For this reason, there is less need to emphasize the translation of algorithms into computer code. This does not diminish the importance of algorithms, but it does suggest putting more stress on their motivation and theoretical properties. Fortunately, the dichotomy between linear and nonlinear programming is fading. It makes better sense pedagogically to view linear programming as a special case of nonlinear programming. This is the attitude taken in

the current book, which makes little mention of the simplex method and develops interior point methods instead. The real bridge between linear and nonlinear programming is convexity. I stress not only the theoretical side of convexity but also its applications in the design of algorithms for problems with either large numbers of parameters or nonlinear constraints.

This graduate-level textbook presupposes knowledge of calculus and linear algebra. I develop quite a bit of mathematical analysis from scratch and feature a variety of examples from linear algebra, differential equations, and convexity theory. Of course, the greater the prior exposure of students to this background material, the more quickly the beginning chapters can be covered. If the need arises, I recommend the texts [82, 134, 135, 188, 222, 223] for supplementary reading. There is ample material here for a fast-paced, semester-long course. Instructors should exercise their own discretion in skipping sections or chapters. For example, Chap. 10 on the EM algorithm primarily serves the needs of students in biostatistics and statistics. Overall, my intended audience includes graduate students in applied mathematics, biostatistics, computational biology, computer science, economics, physics, and statistics. To this list I would like to add upper-division majors in mathematics who want to see some rigorous mathematics with real applications. My own background in computational biology and statistics has obviously dictated many of the examples in the book.

Chapter 1 starts with a review of exact methods for solving optimization problems. These are methods that many students will have seen in calculus, but repeating classical techniques with fresh examples tends simultaneously to entertain, instruct, and persuade. Some of the exact solutions also appear later in the book as parts of more complicated algorithms.

Chapters 2 through 4 review undergraduate mathematical analysis. Although much of this material is standard, the examples may keep the interest of even the best students. Instructors should note that Carathéodory's definition rather than Fréchet's definition of differentiability is adopted. This choice eases the proof of many results. The gauge integral, another good addition to the calculus curriculum, is mentioned briefly.

Chapter 5 gets down to the serious business of optimization theory. McShane's clever proof of the necessity of the Karush–Kuhn–Tucker conditions avoids the complicated machinery of manifold theory and convex cones. It makes immediate use of the Mangasarian–Fromovitz constraint qualification. To derive sufficient conditions for optimality, I introduce second differentials by extending Carathéodory's definition of first differentials. To my knowledge, this approach to second differentials is new. Because it melds so effectively with second-order Taylor expansions, it renders critical proofs more transparent.

Chapter 6 treats convex sets, convex functions, and the relationship between convexity and the multiplier rule. The chapter concludes with the derivation of some of the classical inequalities of probability theory. Prior exposure to probability theory will obviously be an asset for readers here.

Chapters 8 and 9 introduce the MM and EM algorithms. These exploit convexity and the notion of majorization in transferring minimization of the objective function to a surrogate function. Minimizing the surrogate function drives the objective function downhill. The EM algorithm, which is a special case of the MM algorithm, arose in statistics. It is a slight misnomer to call these algorithms. They are really prescriptions for constructing algorithms. It takes experience and skill to wield these tools effectively, so careful attention to the examples is imperative.

Chapter 10 covers Newton's method and its statistical variants, scoring and the Gauss-Newton algorithm. To make this material less dependent on statistical knowledge, I have tried to motivate several algorithms from the perspective of positive definite approximation of the second differential of the objective function. Chapter 11 covers the conjugate gradient algorithm, quasi-Newton algorithms, and the method of trust regions. These classical subjects are in danger of being dropped from the curriculum of nonlinear programming. In my view, this would be a mistake.

Chapter 12 is devoted to convergence questions, both local and global. This material beautifully illustrates the virtues of soft analysis. Instructors wanting to emphasize practical matters may be tempted to sacrifice Chap. 12, but the constant interplay between theory and practice in designing new algorithms argues for its inclusion.

Chapter 13 on convex programming ends the book where more advanced treatises would start. I discuss adaptive barrier methods as a novel application of the MM algorithm, Dykstra's algorithm for finding feasible points in convex programming, and the rudiments of duality theory. These topics belong to the promised land. All you get here is a glimpse from the mountaintop looking out across the river.

Let me add a few words about notation. Lower-division undergraduate texts carefully distinguish between scalars and vectors by setting vectors in boldface type. This convention is considered cumbersome in higher mathematics and is dropped. However, mathematical analysis is plagued by a proliferation of superscripts and subscripts. I prefer to avoid superscripts because of the possible confusion with powers. This decision makes it difficult to distinguish an element of a vector sequence from a component of a vector. My compromise is to represent the  $m$ th entry of a vector sequence as  $x_{(m)}$  and the  $n$ th component of that sequence element as  $x_{mn}$ . Similar conventions hold for matrices. Thus,  $M_{jkl}$  is the entry in row  $k$  and column  $l$  of the  $j$ th matrix  $M_{(j)}$  of a sequence of matrices. Elements of scalar sequences are subscripted in the usual fashion without the enclosing parentheses.

I would like to thank my UCLA students for their help and patience in debugging this text. If it is readable, it is because their questions cut through the confusion. In retrospect, there were more contributing students than I can credit. Let me single out Jason Aten, Lara Bauman, Brian Dolan,

Wei-Hsun Liao, Andrew Nevai-Tucker, Robert Rovetti, and Andy Yip. Paul Maranian kindly prepared the index and proofread my last draft. Finally, I thank my ever helpful and considerate editor, John Kimmel.

I dedicate this book to my daughters, Jane and Maggie. It has been a privilege to be your father. Now that you are adults, I hope you can find the same pleasure in pursuing ideas that I have found in my professional life.



<http://www.springer.com/978-1-4614-5837-1>

Optimization

Lange, K.

2013, XVII, 529 p., Hardcover

ISBN: 978-1-4614-5837-1