Preface

Optimization is everywhere. It is human nature to seek the best option among all that are available. Nature, too, seems to be guided by optimization—many laws of nature have a variational character. Among geometric figures in the plane with a fixed perimeter, the circle has the greatest area. Such isoperimetric problems involving geometric figures date back to ancient Greece. Fermat’s principle, discovered in 1629, stating that the tangent line is horizontal at a minimum point, seems to have influenced the development of calculus. The proofs of Rolle’s theorem and the mean value theorem in calculus use the Weierstrass theorem on the existence of maximizers and minimizers. The introduction of the brachistochrone problem in 1696 by Johann Bernoulli had a tremendous impact on the development of the calculus of variations and influenced the development of functional analysis. The variational character of laws of mechanics and optics were discovered in the seventeenth and eighteenth centuries. Euler and Lagrange forged the foundations of the calculus of variations in the eighteenth century. In the nineteenth century, Riemann used Dirichlet’s principle, which has a variational character, in his investigations in complex analysis. The simplex method for linear programming was discovered shortly after the advent of computers in the 1940s, and influenced the subsequent development of mathematical programming. The emergence of the theory of optimal control in the 1950s was in response to the need for controlling space vehicles and various industrial processes. Today, optimization is a vast subject with many subfields, and it is growing at a rapid pace. Research is proceeding in various directions—advancement of theory, development of new applications and computer codes, and establishment or renewal of ties with many fields in science and industry.

The main focus of this book is optimization in finite-dimensional spaces. In broad terms, this is the problem of optimizing a function $f$ in $n$ variables over a subset of $\mathbb{R}^n$. Thus, the decision variable $x = (x_1, \ldots, x_n)$ is finite-dimensional. A typical problem in this area is a mathematical program $(P)$, which concerns the minimization (or maximization) of a function $f(x)$ subject to finitely many functional constraints of the form $g_i(x) \leq 0$, $h_j(x) = 0$ and
a set constraint of the form $x \in C$, where $f, g_i, h_j$ are real-valued functions defined on some subsets of $\mathbb{R}^n$ and $C$ is a subset of $\mathbb{R}^n$. Any finite-dimensional vector space may be substituted for $\mathbb{R}^n$ without any loss of generality. If the domain of $f$ is an open set and there are no constraints (or more generally if the domains of $g_i, h_j$ and $C$ are open sets), then we have an unconstrained optimization problem. If all the functions are affine and $C$ is defined by linear equations and inequalities, then $(P)$ is called a linear program. If $f, g_i$ are convex functions, $h_j$ is an affine function, and $C$ is a convex set, then $(P)$ is a convex program. If the number of functional constraints is infinite, then $(P)$ is called a semi-infinite program. Mathematical programs have many real-life applications. In particular, linear programming, and more recently semidefinite programming, are enormously popular and have many industrial and scientific applications. The latter problem optimizes a linear function subject to linear equality constraints over the cone of symmetric positive semidefinite matrices.

The main goal of the theory of mathematical programming is to obtain optimality conditions (necessary and sufficient) for a local or global minimizer of $(P)$. This is an impossible task unless some kind of regularity is assumed about the data of $(P)$—the functions $f, g_i, h_j$ and the set $C$. This can be differentiability (in some form) of the functions, or the convexity of the functions as well as of the set $C$. In this book, we will assume that the functions $f, g_i, h_j$ are differentiable as many times as needed (except in cases where there is no advantage to do so), and do not develop nonsmooth analysis in any systematic way. Optimization from the viewpoint of nonsmooth analysis is competently covered in several recent books; see for example *Variational Analysis* by Rockafellar and Wets, and *Variational Analysis and Generalized Differentiation* by Mordukhovich. Another goal of the theory, important especially in convex programming, is the duality theory, whereby a second convex program $(D)$ is associated with $(P)$ such that the pair $(P)$-$(D)$ have remarkable properties which can be exploited in several useful ways. If the problem $(P)$ has a lot of structure, it may be possible to use the optimality conditions to solve analytically for the solutions to $(P)$. This desirable situation is very valuable when it is successful, but it is rare, so it becomes necessary to devise numerical optimization techniques or algorithms to search for the optimal solutions. The process of designing efficient algorithms requires a great deal of ingenuity, and the optimality conditions contribute to the process in several ways, for example by suggesting the algorithm itself, or by verifying the correctness of the numerical solutions returned by the algorithm. The role of the duality theory in designing algorithms is similar, and often more decisive.

All chapters except Chapter 14 are concerned with the theory of optimization. We have tried to present all the major results in the theory of finite-dimensional optimization, and strived to provide the best available proofs whenever possible. Moreover, we include several independent proofs of some of the most important results in order to give the reader and the instructor of a course using this book flexibility in learning or teaching the key subjects.
On several occasions we give proofs that may be new. Not all chapters deal exclusively with finite-dimensional spaces, however. Chapters 3, 5, 6, 14, and Appendices A and C contain, in part or fully, important results in nonlinear analysis and in the theory of convexity in infinite-dimensional settings.

Chapter 14 may be viewed as a short course on three basic optimization algorithms: the steepest descent method, Newton’s method, and the conjugate-gradient method. In particular, the conjugate-gradient method is presented in great detail. The three algorithms are chosen to be included because many computational schemes in mathematical programming have their origins in these algorithms.

**Audience and background**

The book is suitable as a textbook for a first course in the theory of optimization in finite-dimensional spaces at the graduate level. The book is also suitable for self-study or as a reference book for more advanced readers. It evolved out of my experience in teaching a graduate-level course twelve times since 1993, eleven times at the University of Maryland, Baltimore County (UMBC), and once in 2001 at Bilkent University, Ankara, Turkey. An important feature of the book is the inclusion of over two hundred carefully selected exercises as well as a fair number of completely solved examples within the text.

The prerequisites for the course are analysis and linear algebra. The reader is assumed to be familiar with the basic concepts and results of analysis in finite-dimensional vector spaces—limits, continuity, completeness, compactness, connectedness, and so on. In some of the more advanced chapters and sections, it is necessary to be familiar with the same concepts in metric and Banach spaces. The reader is also assumed to be familiar with the fundamental concepts and results of linear algebra—vector space, matrix, linear combination, span, linear independence, linear map (transformation), and so on.

**Suggestions for using this book in a course**

Ideally, a first course in finite-dimensional optimization should cover the first-order and second-order optimality conditions in unconstrained optimization, the fundamental concepts of convexity, the separation theorems involving convex sets (at least in finite-dimensional spaces), the theory of linear inequalities and convex polyhedra, the optimality conditions in nonlinear programming, and the duality theory of convex programming. These are treated in Chapters 2, 4, 6, 7, 9, and 11, respectively, and can be covered in a one-semester course. Chapter 1 on differential calculus can be covered in such a course, or referred to as needed, depending on the background of the students. In
any case, it is important to be familiar with the multivariate Taylor formulas, because they are used in deriving optimality conditions and in differentiating functions.

In my courses, I cover Chapter 1 (Sections 1.1–1.5), Chapter 2 (Sections 2.1–2.5), Chapter 4 (Sections 4.1–4.5), Chapter 6 (Sections 6.1–6.5, and assuming the results from Chapter 5 that are used in some proofs), Chapter 7 (Sections 7.1–7.4), Chapter 9 (Sections 9.1–9.2, 9.4–9.9), and Chapter 11 (Sections 11.1–11.6). This course emphasizes the use of separation theorems for convex sets for deriving the optimality conditions for nonlinear programming. This approach is both natural and widely applicable—it is possible to use the same idea to derive optimality conditions for many types of problems, from nonlinear programming to optimal control problems, as was shown by Dubovitskii and Milyutin.

Several other possibilities exist for covering most of this core material. If the goal is to cover quickly the basics of nonlinear programming but not of convexity, then one can proceed as above but skip Chapter 6 and the first two sections of Chapter 7, substitute Appendix A for Sections 7.3 and 7.4, and skip Section 9.1. In this approach, one needs to accept the truth of Theorem 11.15 without proof.

A third possibility is to follow Chapter 3 to cover the theory of linear inequalities and the basic theorems of nonlinear analysis, and then cover Chapter 9 (Sections 9.3–9.9). Still other possibilities exist for covering the core material.

If more time is available, an instructor may choose to cover Chapter 14 on algorithms, Chapter 8 on linear programming, Chapter 10 on nonlinear programming, or Chapter 12 on semi-infinite programming. In a course oriented more toward convexity, the instructor may cover Chapter 5, 6, or 13 for a more in-depth study of convexity. In particular, Chapters 5 and 6 contain very detailed, advanced results on convexity.

Chapters 4–8, 11, and 13 can be used for a stand-alone one-semester course on the theory of convexity. If desired, one may supplement the course by presenting the theory of Fenchel duality using, for example, Chapters 1–3 and 6 of the book Convex Analysis and Variational Problems by Ekeland and Temam. The theory of convexity has an important place in optimization. We already mentioned the role of the separation theorems for convex sets in deriving optimality conditions in mathematical programming. The theory of duality is a powerful tool with many uses, both in theory of optimization and in the design of numerical optimization algorithms. The role of convexity in the complexity theory of optimization is even more central; since the work of Nemirovskii and Yudin in the 1970s on the ellipsoid method, we know that convex programming (and some close relatives) is the only known class of problems that are computationally tractable, that is, for which polynomial-time methods can be developed.
The major pathways through the book are indicated in the following diagram.

Comments on the contents of individual chapters

Chapter 1 includes background material on differential calculus. Two novel features of the chapter are the converse of Taylor’s formula and Danskin’s theorem. The first result validates the role of Taylor’s formula for computing derivatives, and Danskin’s formula is a useful tool in optimization.

Chapter 2 develops the first-order and second-order optimality conditions in unconstrained optimization. Section 2.4 deals with quadratic forms and symmetric matrices. We recall the spectral decomposition of a symmetric matrix, give the eigenvalue characterizations of definite and semidefinite matrices, state Descartes’s exact rule of sign (whose proof is given in Appendix B), and use it as tool for recognizing definite and semidefinite matrices. We also include a proof of Sylvester’s theorem on the positive definiteness of a symmetric matrix. (An elegant optimization-based proof is given in an exercise at the end of the chapter.) In Section 2.5, we give the proofs of the inverse and implicit function theorems and Lyusternik’s theorem using an optimization-based approach going back at least to Carathéodory. A proof of Morse’s lemma is given in Section 2.6 because of the light it throws on the second-order optimality conditions.

Chapter 3 is devoted to Ekeland’s \( \epsilon \)-variational principle (and its relatives) and its applications. We use it to prove the central result on linear inequalities (Motzkin’s transposition theorem), and the basic theorems of nonlinear analysis in a general setting. Variational principles are fascinating, and their importance in optimization is likely to grow even more in the future.
The next three chapters are devoted to convexity. Chapter 4 treats the fundamentals of convex analysis. We include Section 4.1 on affine geometry because of its intrinsic importance, and because it helps make certain results in convexity more transparent.

Chapter 5 delves into the structure of convex sets. A proper understanding of concepts such as the relative interior, closure, and the faces of convex sets is essential for proving separation theorems involving convex sets and much else. The concept of the relative interior is developed in both algebraic and topological settings.

Chapter 6 is devoted to the separation of convex sets, the essential source of duality, at least in convex programming. The chapter is divided into two parts. Sections 6.1–6.5 deal with the separation theorems in finite dimensions and do not depend heavily on Chapter 5. They are sufficient for somebody who is interested in only the finite-dimensional situation. Section 6.5 is devoted to the finite-dimensional version of the Dubovitskii–Milyutin theorem, a convenient separation theorem, applicable to the separation of several convex sets. Sections 6.6–6.8 treat the separation theorems involving two or several convex sets in a very general setting. Chapter 5 is a prerequisite for these sections, which are intended for more advanced readers.

Chapters 7 and 8 treat the theories of convex polyhedra and linear programming, respectively. Two sections, Section 7.5 on Tucker’s complementarity theorem and Section 8.5 on the existence of strictly complementary solutions in linear programming, are important in interior-point methods.

Chapters 9 and 10 treat nonlinear programming. The standard, basic theory consisting of first-order (Fritz John and KKT) and second-order conditions for optimality is given in Chapter 9. A novel feature of the chapter is the inclusion of a first-order sufficient optimality condition that goes back to Fritz John, and several completely solved examples of nonlinear programs. Chapter 10 gives complete solutions for seven structured optimization problems. These problems are chosen for their intrinsic importance and to demonstrate that optimization techniques can resolve important problems.

Chapter 11 deals with duality theory. We have chosen to treat duality using the Lagrangian function. This approach is completely general for convex programming, because it is equivalent to the approach by Fenchel duality in that context, and more general because it is sometimes applicable beyond convex programming. We establish the general correspondence between saddle point and duality in Section 11.2 and apply it to nonlinear programming in Section 11.3. The most important result of the chapter is the strong duality theorem for convex programming given in Section 11.4, under very weak conditions. It is necessary to use sophisticated separation theorems to achieve this result. After treating several examples of duality in Section 11.5, we turn to the duality theory of conic programming in Section 11.6. As a novel application, we give a proof of Hoffman’s lemma using duality in Section 11.8.

Semi-infinite programming is the topic of Chapter 12. This subject is not commonly included in most optimization textbooks, but many impor-
tant problems in finite dimensions require it, such as the problem of finding the extremal-volume ellipsoids associated with a convex body in $\mathbb{R}^n$. We derive the Fritz John optimality conditions for these problems using Danskin’s theorem when the set indexing the constraints is compact. In the rest of the chapter we solve several specific, important semi-infinite programming problem rather than giving a systematic theory. Another method to treat convex semi-infinite programs, using Helly’s theorem, is given in Section 13.2.

Chapter 13 is devoted to several special topics in convexity that we deem interesting: the combinatorial theory of convex sets, homogeneous convex functions, decomposition of convex cones, and norms of polynomials. The last topic finds an interesting application to self-concordant functions in interior-point methods.

The focus of Chapter 14 is on algorithms. The development of numerical algorithms for optimization problems is a highly intricate art and science, and anything close to a proper treatment would require several volumes. This chapter is included in our book out of the conviction that there should be a place in a book on theory for a chapter such as this, which treats in some depth a few select algorithms. This should help the reader put the theory in perspective, and accomplish at least three goals: the reader should see how theory and algorithms fit together, how they are different, and whether there are differences in the thought processes that go into developing each part. It should also give the reader additional incentive to learn more about algorithms.

We choose to treat three fundamental optimization algorithms: the steepest-descent (and gradient projection), Newton’s, and conjugate-gradient methods. We develop each in some depth and provide convergence rate estimates where possible. For example, we provide the convergence rate for the steepest-descent method for the minimization of a convex quadratic function, and for the minimization of a convex function with Lipschitz gradient. The convergence theory of Newton’s method is treated, including the convergence theory due to Kantorovich. Finally, we give a very extensive treatment of the conjugate-gradient method. We prove its remarkable convergence properties and show its connection with orthogonal polynomials.

In Appendix A, we give the theory for the consistency of a system of finitely many linear (both strict and weak) inequalities in arbitrary vector spaces. The algebraic proof has considerable merits: it is very general, does not need any prerequisites, and does not use the completeness of the field over which the vector space is defined. Consequently, it is applicable to linear inequalities with rational coefficients.

In Appendix B, we give a short proof of Descartes’s exact rule of sign, and in Appendix C, the classical proofs of the open mapping theorem and Graves’s theorem.
Acknowledgements

I am indebted to many people for their help at various stages of writing this book. First of all, I would like to thank my graduate students, who during the last twelve years have given me valuable feedback by alerting me to numerous inaccuracies, typos, ambiguities, and difficulties in the exposition in the different versions of the manuscript. I would like to thank Gürkan Üstünkar and Ömer E. Kundakçioğlu for typing my initial set of lecture notes. My colleague Jinglai Shen taught twice using my manuscript, and gave me a number of excellent suggestions. My brother-in-law Wilson Renne kept me on track with his inquiries about the status of my book. Achi Dosanjh at Springer Verlag has been a constant help ever since our initial correspondence. Most of all, I would like to thank my wife, Colleen, my daughter, Aylin, and my son, Timur, for their encouragement, and for the many sacrifices they have made during the long time it took to write this book. Finally, I am grateful to the National Science Foundation for summer support during the writing of this book.

Ellicott City, Maryland,  
February 2010

Osman Güler
Foundations of Optimization
Güler, O.
2010, XVIII, 442 p., Hardcover