Osman Güler

Foundations of Optimization

in Finite Dimensions

January 28, 2010

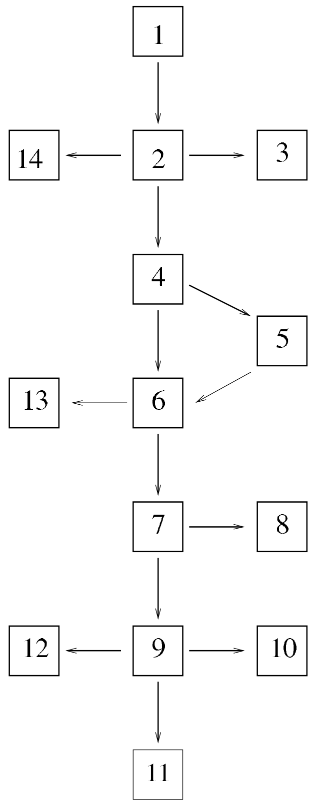
Springer
Berlin Heidelberg New York
Hong Kong London
Milan Paris Tokyo

Preface

During the past sixteen years, I taught a one-semester graduate-level course on the theory of optimization in finite dimensions on about ten occasions. This book grew out of my experience teaching these courses. Optimization in finite dimensions is a huge subject, growing at an accelerated pace. It is impossible to cover its main branches in a single course. In my view, there should be at least two basic graduate-level courses in optimization, the first dealing with the theory of optimization and the second with numerical algorithms. The two parts are highly interconnected, of course, but the theory needs to be learned first, because it is often necessary to understand and interpret algorithms.

This book is intended for a graduate-level course emphasizing the theory of optimization. Such a course should cover the necessary and sufficient conditions in unconstrained and constrained optimization problems, and the theory of duality. Convexity is important in all of these, and should be treated in sufficient detail. Also, linear programming and the theory of convex polyhedra are important special topics in optimization, so they should be covered as well. I have found Taylor's formula to be a very effective tool for computing derivatives. This alone is a compelling reason for developing a solid introduction to differential calculus as background material. We develop differentiation using coordinates and in a coordinate-free way, because we find that the latter approach is simpler and works best in most applications, and subjects such as semidefinite programming almost require it - introducing coordinates there is not very practical.

The book contains 14 chapters. The interdependency of the chapters are indicated in the Figure. The core of the book consists of the chapters in the middle, Chapters 1, 2, 4, 6, 7, 9, and 11. In my courses, I usually cover Sections 1.1-1.6, 2.1-2.6, all of Chapter 4, Sections 6.1-6.5, Sections 7.1-7.4, all of Chapter 9, and Sections 11.1-11.6. Sometimes, I also cover Chapter 8 on linear programming, Chapter 14 on algorithms, and Chapter 3 on variational principles. Several other choices exist for covering the core material. An important feature of the book is the careful selection of a couple of hundred Exercises as well as a good number of worked-out examples within the text.



The prerequisites for the course are analysis and linear algebra. The reader should be familiar with analysis in metric spaces, limits, continuity, completeness, compactness, connectedness and so on; the fundamental concepts of linear algebra, such as the notions of vector spaces, linear maps, linear independence, and matrices, are also needed. Several more advanced notions, such as Banach and topological vector spaces are used in some of the more advanced chapters, but these are not needed in the core of the course.

We hope that the book will be useful as a reference book for more experience optimizers. Many chapters contain some advanced sections, and several chapters, such as Chapters 3, 5, Sections 6.6-6.8, Chapter 12 and 13, are devoted to more advanced topics. We have tried to find the best proofs of many fundamental results. On several occasions we give proofs which may be new.

Here is an outline of the individual chapters. Chapter 1 covers 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 useful in several places in optimization, for example in semi-infinite programming. Chapter 2 develops the optimality conditions for unconstrained optimization, emphasizing the first and second order conditions for optimality. In Section 2.5 we include a proof of Sylvester's theorem on the positive definiteness of a symmetric matrix as well as Descartes's exact rule of sign, which is another tool for recognizing symmetric, positive definite as well as positive semidefinite matrices. In Section 2.6, 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 included because of the light it throws on the second order optimality conditions.

Chapter 3, devoted to Ekeland's ϵ-variational principle and its relatives, is more advanced. We use these principles 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 (finite or infinite dimensional) is likely to grow even more in the future.

The next three chapters are devoted to convexity. Chapter 4 covers 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. We have found over the years that a proper understanding of concepts like relative interior, closure, and the faces of convex sets is the key for a good understanding of separation theorems of convex sets and much else besides; moreover, that the concept of relative interior and closure can be developed, to advantage, in a purely algebraic fashion. Chapter 6 is devoted to 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 separation theorems in finite dimensions and do not depend on Chapter 5; they are sufficient for somebody who is only interested in the finite dimensional situation. Section 6.5 is devoted to the finite-dimensional version of Dubovitskii–Milyutin theorem, a convenient separation theorem, applicable to separation of several convex sets. Sections 6.6-6.8 develop separation of 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 theory 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.

Chapter 9 and 10 cover nonlinear programming, that is, the optimization of a (possibly) nonlinear function of n variable over a set of feasible region which is given by finite many (possibly) nonlinear constraints. The standard, basic theory consisting of first (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 worked-out examples of nonlinear programs. Chapter 10 gives complete solutions for seven structured optimization problems. These problems are chosen for their intrinsic importance and to show that optimization techniques can resolve important problems.

Chapter 11 deals with the duality theory of optimization. We have chosen to treat duality using Lagrangian functions. 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 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.

Semi-infinite programming is the topic of Chapter 12. This is a topic which has not been treated much in optimization textbooks, but several important problems in finite dimension require it, such as the problem of finding extremal-volume ellipsoids associated with convex bodies in $\mathbb{R}^n$. We derive Fritz John's 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 particular, 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 which 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. Development of numerical optimization algorithms 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 may be a place in a theory book for a chapter such as this, which treats in some depth a select few algorithms. This would help the reader put the theory in perspective, and accomplish at least three goals: the reader would see how theory and algorithms fit together, how they are different, and if there are differences in the thought processes that go into developing each part. It may also give extra 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 decent method for the minimization of a convex quadratic function and for minimization of a convex function with Lipschitz gradient. The converge theory of

Newton's method is treated, including the convergence theory of 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 classical proofs the open mapping theorem and Graves's theorem. In Appendix B, we give the theory of consistency of finitely many linear (both strict and weak) inequalities in arbitrary vector spaces. The algebraic proof has considerable merits, because 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.

Some central results such the basic theorems of nonlinear analysis and Motzkin's transposition theorem are given several, independent proofs. This is of course a repetition, but there is a merit to seeing a central result from several perspectives. In the end, though, it is the flexibility it gives the reader or instructor to cover the course material that decided the issue for us.

Acknowledgements

I am indebted to many people for their help at various times during the writing of this book. First of all, I would like to thank my graduate students, who during the last ten years, gave me feedback by alerting me to numerous inaccuracies, typos, ambiguities, and difficulties in the exposition in different versions of the manuscript. I would like to thank Gürkan Üstünkar and Ömer E. Kundakçıoğlu for typing my initial set of lecture notes. Prof. 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 constant inquiries about my book. Ms. Achi Dosanjh at Springer Verlag has been a constant help ever since our initial correspondence. Most of all, I would like to thank my family, my wife Colleen, and my daugher Aylin and my son Timur for their encouragements and for the many sacrifices they have made during the long time it took to write this book.

Baltimore,
August 2009

Osman Güler

Contents