Convex Analysis and Optimization

Preface:

This book focuses on the theory of convex sets and functions, and its connections with a number of topics that span a broad range from continuous to discrete optimization. These topics include Lagrange multiplier theory, Lagrangian and conjugate/Fenchel duality, minimax theory, and nondifferentiable optimization.

The book evolved from a set of lecture notes for a graduate course at M.I.T. It is widely recognized that, aside from being an eminently useful subject in engineering, operations research, and economics, convexity is an excellent vehicle for assimilating some of the basic concepts of real analysis within an intuitive geometrical setting. Unfortunately, the subject's coverage in academic curricula is scant and incidental. We believe that at least part of the reason is the shortage of textbooks that are suitable for classroom instruction, particularly for nonmathematics majors. We have therefore tried to make convex analysis accessible to a broader audience by emphasizing its geometrical character, while maintaining mathematical rigor. We have included as many insightful illustrations as possible, and we have used geometric visualization as a principal tool for maintaining the students' interest in mathematical proofs.

Our treatment of convexity theory is quite comprehensive, with all major aspects of the subject receiving substantial treatment. The mathematical prerequisites are a course in linear algebra and a course in real analysis in finite dimensional spaces (which is the exclusive setting of the book). A summary of this material, without proofs, is provided in Section 1.1. The coverage of the theory has been significantly extended in the exercises, which represent a major component of the book. Detailed solutions of all the exercises (nearly 200 pages) are internet-posted in the book's www page

Some of the exercises may be attempted by the reader without looking at the solutions, while others are challenging but may be solved by the advanced reader with the assistance of hints. Still other exercises represent substantial theoretical results, and in some cases include new and unpublished research. Readers and instructors should decide for themselves how to make best use of the internet-posted solutions.

An important part of our approach has been to maintain a close link between the theoretical treatment of convexity and its application to optimization. For example, in Chapter 2, after the development of some of the basic facts about convexity, we discuss some of their applications to optimization and saddle point theory; in Chapter 3, after the discussion of polyhedral convexity, we discuss its application in linear and integer programming; and in Chapter 4, after the discussion of subgradients, we discuss their use in optimality conditions. We follow this style in the remaining chapters, although having developed in Chapters 1-4 most of the needed convexity theory, the discussion in the subsequent chapters is more heavily weighted towards optimization.

The chart of the opposite page illustrates the main topics covered in the book, and their interrelations. At the top level, we have the most basic concepts of convexity theory, which are covered in Chapter 1. At the middle level, we have fundamental topics of optimization, such as existence and characterization of solutions, and minimax theory, together with some supporting convexity concepts such as hyperplane separation, polyhedral sets, and subdifferentiability

(Chapters 2-4). At the lowest level, we have the core issues of convex optimization: Lagrange multipliers, Lagrange and Fenchel duality, and numerical dual optimization (Chapters 5-8).

An instructor who wishes to teach a course from the book has a choice between several different plans. One possibility is to cover in detail just the first four chapters, perhaps augmented with some selected sections from the remainder of the book, such as the first section of Chapter 7, which deals with conjugate convex functions. The idea here is to concentrate on convex analysis and illustrate its application to minimax theory through the minimax theorems of Chapters 2 and 3, and to constrained optimization theory through the Nonlinear Farkas' Lemma of Chapter 3 and the optimality conditions of Chapter 4. An alternative plan is to cover Chapters 1-4 in less detail in order to allow some time for Lagrange multiplier theory and computational methods. Other plans may also be devised, possibly including some applications or some additional theoretical topics of the instructor's choice.

While the subject of the book is classical, the treatment of several of its important topics is new and in some cases relies on new research. In particular, our new lines of analysis include:

{(a)} A unified development of minimax theory and constrained optimization duality as special cases of the duality between two simple geometrical problems: the min common point problem and the max crossing point problem. Here, by minimax theory, we mean the analysis relating to the minimax equality

$$\inf_{x\in X}\sup_{z\in Z}\phi(x,z)=\sup_{z\in Z}\inf_{x\in X}\phi(x,z),$$

and the attainment of the "inf" and the "sup." By constrained optimization theory, we mean the analysis of problems such as

minimize $f(x)$

subject to $x\in X$, $g_j(x)\leq 0$, $j=1,...,r$,

and issues such as the existence of optimal solutions and Lagrange multipliers, and the absence of a duality gap [equality of the optimal value of the above problem and the optimal value of an associated dual problem, obtained by assigning multipliers to the inequality constraints $g_j(x)\leq 0$].

(b) A unification of conditions for existence of solutions of convex optimization problems, conditions for the minimax equality to hold, and conditions for the absence of a duality gap in constrained optimization. This unification is based on conditions guaranteeing that a nested family of closed convex sets has a nonempty intersection.

(c) A unification of the major constraint qualifications that guarantee the existence of Lagrange multipliers for nonconvex constrained optimization. This unification is achieved through the notion of constraint pseudonormality, which is motivated by an enhanced form of the Fritz John necessary optimality conditions.

(d) The development of incremental subgradient methods for dual optimization, and the analysis of their advantages over classical subgradient methods.

We provide some orientation by informally summarizing the main ideas of each of the above topics.

Min Common/Max Crossing Duality

In this book, duality theory is captured in two easily visualized problems: the min common point problem and the max crossing point problem, introduced in Chapter 2. Fundamentally, these problems revolve around the existence of nonvertical supporting hyperplanes to convex sets that are unbounded from above along the vertical axis. When properly specialized, this turns out to be the critical issue in constrained optimization duality and saddle point/minimax theory, under standard convexity and/or concavity assumptions.

The salient feature of the min common/max crossing framework is its simple geometry, in the context of which the fundamental constraint qualifications needed for strong duality theorems are visually apparent, and admit straightforward proofs. This allows the development of duality theory in a unified way: first within the min common/max crossing framework in Chapters 2 and 3, and then by specialization, to saddle point and minimax theory in Chapters 2 and 3, and to optimization duality in Chapter 6. All of the major duality theorems discussed in this book are derived in this way, including the principal Lagrange multiplier and Fenchel duality theorems for convex programming, and the von Neuman Theorem for zero sum games.

From an instructional point of view, it is particularly desirable to unify constrained optimization duality and saddle point/minimax theory (under convexity/concavity assumptions). Their connection is well known, but it is hard to understand beyond a superficial level, because there is not enough overlap between the two theories to develop one in terms of the other. In our approach, rather than trying to build a closer connection between constrained optimization duality and saddle point/minimax theory, we show how they both stem from a common geometrical root: the min common/max crossing duality.

We note that the constructions involved in the min common and max crossing problems arise in the theories of subgradients, conjugate convex functions, and duality. As such they are implicit in several earlier analyses; in fact they have been employed for visualization purposes in the first author's nonlinear programming textbook [Ber99]. However, the two problems have not been used as a unifying theoretical framework for constrained optimization duality, saddle point theory, or other contexts, except implicitly through the theory of conjugate convex functions, and the complicated and specialized machinery of conjugate saddle functions. Pedagogically, it appears desirable to postpone the introduction of conjugacy theory until it is needed for the limited purposes of Fenchel duality (Chapter 7), and to bypass altogether conjugate saddle function theory, which is what we have done.

Existence of Solutions and Strong Duality

We show that under convexity assumptions, several fundamental issues in optimization are intimately related. In particular, we give a unified analysis of conditions for optimal solutions to exist, for the minimax equality to hold, and for the absence of a duality gap in constrained optimization.

To provide a sense of the main idea, we note that given a constrained optimization problem, lower semicontinuity of the cost function and compactness of the constraint set guarantee the existence of an optimal solution (the Weierstrass Theorem). On the other hand, the same conditions plus convexity of the cost and constraint functions guarantee not only the existence of an optimal solution, but also the absence of a duality gap. This is not a coincidence, because as it turns out, the conditions for both cases critically rely on the same fundamental properties of compact sets, namely that the intersection of a nested family of nonempty compact sets is nonempty and compact, and that the projections of compact sets on any subspace are compact.

In our analysis, we extend this line of reasoning under a variety of assumptions relating to convexity, directions of recession, polyhedral sets, and special types of sets specified by quadratic and other types of inequalities. The assumptions are used to establish results asserting that the intersection of a nested family of closed convex sets is nonempty, and that the function $f(x)=\inf_z F(x,z)$, obtained by partial minimization of a convex function F, is lower semicontinuous. These results are translated in turn to a broad variety of conditions that guarantee the existence of optimal solutions, the minimax equality, and the absence of a duality gap.

Pseudonormality and Lagrange Multipliers

In Chapter 5, we discuss Lagrange multiplier theory in the context of optimization of a smooth cost function, subject to smooth equality and inequality constraints, as well as an additional set constraint. Our treatment of Lagrange multipliers is new, and aims to generalize, unify, and streamline the theory of constraint qualifications.

The starting point for our development is an enhanced set of necessary conditions of the Fritz John type, that are sharper than the classical Karush-Kuhn-Tucker conditions (they include extra conditions, which may narrow down the field of candidate local minima). They are also more general in that they apply when there is an abstract (possibly nonconvex) set constraint, in addition to the equality and inequality constraints. To achieve this level of generality, we bring to bear notions of nonsmooth analysis, and we find that the notion of regularity of the abstract constraint set provides the critical distinction between problems that do and do not admit a satisfactory theory.

Fundamentally, Lagrange multiplier theory should aim to identify the essential constraint structure that guarantees the existence of Lagrange multipliers. For smooth problems with equality and inequality constraints, but no abstract set constraint, this essential structure is captured by the classical notion of quasiregularity (the tangent cone at a given feasible point is equal to the cone of first order feasible variations). However, in the presence of an additional set constraint, the notion of quasiregularity breaks down as a viable unification vehicle. Our development introduces the notion of pseudonormality as a substitute for quasiregularity for the case of an abstract set constraint. Pseudonormality unifies and expands the major constraint qualifications, and simplifies the proofs of Lagrange multiplier theorems. In the case of equality constraints only, pseudonormality is implied by either one of two alternative constraint qualifications: the linear independence of the constraint gradients and the linearity of the constraint functions. In fact, in this case, pseudonormality is not much different than the union of these two constraint qualifications. However, pseudonormality is a meaningful unifying property even in the case of an additional set constraint, where the classical proof arguments based on quasiregularity fail. Pseudonormality also provides the connecting link between constraint qualifications and the theory of exact penalty functions.

An interesting byproduct of our analysis is a taxonomy of different types of Lagrange multipliers for problems with nonunique Lagrange multipliers. Under some convexity assumptions, we show that if there exists at least one Lagrange multiplier vector, there exists at least one of a special type, called informative, which has nice sensitivity properties. The nonzero components of such a multiplier vector identify the constraints that need to be violated in order to improve the optimal cost function value. Furthermore, a particular informative Lagrange multiplier vector characterizes the direction of steepest rate of improvement of the cost function for a given level of the norm of the constraint violation. Along that direction, the equality and inequality constraints are violated consistently with the signs of the corresponding multipliers.

The theory of enhanced Fritz John conditions and pseudonormality are extended in Chapter 6 to the case of a convex programming problem, without assuming the existence of an optimal solution or the absence of a duality gap. They form the basis for a new line of analysis for asserting the existence of informative multipliers under the standard constraint qualifications.

Incremental Subgradient Methods

In Chapter 8, we discuss one of the most important uses of duality: the numerical solution of dual problems, often in the context of discrete optimization and the method of branch-and-bound. These dual problems are often nondifferentiable and have special structure. Subgradient methods have been among the most popular for the solution of these problems, but they often suffer from slow convergence.

We introduce incremental subgradient methods, which aim to accelerate the convergence by exploiting the additive structure that a dual problem often inherits from properties of its primal problem, such as separability. In particular, for the common case where the dual function is the sum of a large number of component functions, incremental methods consist of a sequence of incremental steps, each involving a single component of the dual function, rather than the sum of all components.

Our analysis aims to identify effective variants of incremental methods, and to quantify their advantages over the standard subgradient methods. An important question is the selection of the order in which the components are selected for iteration. A particularly interesting variant uses randomization of the order to resolve a worst-case complexity bottleneck associated with the natural deterministic order. According to both analysis and experiment, this randomized variant performs substantially better than the standard subgradient methods for large scale problems that typically arise in the context of duality. The randomized variant is also particularly well-suited for parallel, possibly asynchronous, implementation, and is the only available method, to our knowledge, that can be used efficiently within this context.