Preface

This book, as the title suggests, is about first-order methods, namely, methods that exploit information on values and gradients/subgradients (but not Hessians) of the functions comprising the model under consideration. First-order methods go back to 1847 with the work of Cauchy on the steepest descent method. With the increase in the amount of applications that can be modeled as large- or even huge-scale optimization problems, there has been a revived interest in using simple methods that require low iteration cost as well as low memory storage.

The primary goal of the book is to provide in a self-contained manner a comprehensive study of the main first-order methods that are frequently used in solving large-scale problems. This is done by gathering and reorganizing in a unified manner many results that are currently scattered throughout the literature. Special emphasis is placed on rates of convergence and complexity analysis. Although the name of the book is "first-order methods in optimization," two disclaimers are in order. First, we will actually also consider methods that exploit additional operations at each iteration such as prox evaluations, linear oracles, exact minimization w.r.t. blocks of variables, and more, so perhaps a more suitable name would have been "simple methods in optimization." Second, in order to be truly self-contained, the first part of the book (Chapters 1–7) is actually purely theoretical and contains essential topics that are crucial for the developments in the algorithmic part (Chapters 8–15).

The book is intended for students and researchers with a background in advanced calculus and linear algebra, as well as prior knowledge in the fundamentals of optimization (some convex analysis, optimality conditions, and duality). A MATLAB toolbox implementing many of the algorithms described in the book was developed by the author and Nili Guttmann-Beck and can be found at `www.siam.org/books/mo25`.

The outline of the book is as follows. Chapter 1 reviews important facts about vector spaces. Although the material is quite fundamental, it is advisable not to skip this chapter since many of the conventions regarding the underlying spaces used in the book are explained. Chapter 2 focuses on extended real-valued functions with a special emphasis on properties such as convexity, closedness, and continuity. Chapter 3 covers the topic of subgradients starting from basic definitions, continuing with directional derivatives, differentiability, and subdifferentiability and ending with calculus rules. Optimality conditions are derived for convex problems (Fermat's optimality condition), but also for the nonconvex composite model, which will be discussed extensively throughout the book. Conjugate functions are the subject of Chapter 4, which covers several issues, such as Fenchel's inequality, the biconjugate, calculus rules, conjugate subgradient theorem, relations with the infimal convolution, and Fenchel's duality theorem. Chapter 5 covers two different but closely related subjects: smoothness and strong convexity—several characterizations of each of these concepts are given, and their relation via the conjugate correspondence theorem is established. The proximal operator is discussed in Chapter 6, which includes a large amount of prox computations as well as calculus rules. The basic properties of the proximal mapping (first and second prox theorems and Moreau decomposition) are proved, and the Moreau envelope concludes the theoretical part of the chapter. The first part of the book ends with Chapter 7, which contains a study of symmetric spectral functions. The second, algorithmic part of the book starts with Chapter 8 with primal and dual projected subgradient methods. Several stepsize rules are discussed, and complexity results for both the convex and the strongly convex cases are established. The chapter also includes discussions on the stochastic as well as the incremental projected subgradient methods. The non-Euclidean version of the projected subgradient method, a.k.a. the mirror descent method, is discussed in Chapter 9. Chapter 10 is concerned with the proximal gradient method as well as its many variants and extensions. The chapter also studies several theoretical results concerning the so-called gradient mapping, which plays an important part in the convergence analysis of proximal gradient–based methods. The extension of the proximal gradient method to the *block* proximal gradient method is discussed in Chapter 11, while Chapter 12 considers the dual proximal gradient method and contains a result on a primal-dual relation that allows one to transfer rate of convergence results from the dual problem to the primal problem. The generalized conditional gradient method is the topic of Chapter 13, which contains the basic rate of convergence results of the method, as well as its block version, and discusses the effect of strong convexity assumptions on the model. The alternating minimization method is the subject of Chapter 14, where its convergence (as well as divergence) in many settings is established and illustrated. The book concludes with a discussion on the ADMM method in Chapter 15.

My deepest thanks to Marc Teboulle, whose fundamental works in first-order methods form the basis of many of the results in the book. Marc introduced me to the world of optimization, and he is a constant source and inspiration and admiration. I would like to thank Luba Tetruashvili for reading the book and for her helpful remarks. It has been a pleasure to work with the extremely devoted and efficient SIAM staff. Finally, I would like to acknowledge the support of the Israel Science Foundation for supporting me while writing this book.