

[Review published in *SIAM Review*, Vol. 61, Issue 2, pp. 393–394.]

First-Order Methods in Optimization. By Amir Beck. SIAM, Mathematical Optimization Society, Philadelphia, 2017. \$97.00. xii+475 pp., softcover. MOS-SIAM Series on Optimization. Vol. 25. ISBN 978-1-611974-98-0. <https://doi.org/10.1137/1.9781611974997>.

First-Order Methods in Optimization, as the title suggests, deals with (sub)gradient-based numerical algorithms for continuous optimization problems, where nonsmooth and convex problems are the center of attention. With the rise of *big data*, first-order methods in convex (and most recently non-convex) optimization have experienced a huge renaissance. Hence a book like this one, in which different theoretical tools and an abundance of numerical methods are presented in a unified and self-contained manner, was sorely needed. Given his contributions and expertise in this research area, Amir Beck is an ideal person to have written it. The book is equally valuable to both researchers and practitioners. It can also be used as a textbook; as a matter of fact I am currently using parts of it to teach a course on convex optimization. The book is subdivided into fifteen chapters: The first seven chapters provide the theoretical tools from convex analysis on which the subsequent study relies. Naturally, a special emphasis is put on subgradients and proximal operators since these are the main building blocks for the optimization methods to be studied later. Moreover, the role of L -smoothness (i.e., L -Lipschitzness of the gradients) and strong convexity are highlighted, as these are fundamental properties that yield desirable convergence rates of numerical optimization methods.

The remaining eight chapters (Chapters 8–15) are devoted to optimization methods for solving (nonsmooth) optimization problems with different underlying structures, like additive composite or (explicitly) constrained problems. Methods such as (projected) subgradient, mirror descent, proximal gradient, conditional gradient, and ADMM are studied. These chapters are largely independent of each other (and

hence can be read accordingly) and only build on the tools from the first seven chapters. The convergence results and their technique of proof are state of the art, containing (at the day of release) most current results. The book explicitly does *not* discuss termination criteria.

Many of the algorithms studied are implemented in a free MATLAB toolbox (available from www.siam.org/books/mo25) that nicely complements the book.

The style of the book is very clear and concise and, I should think, also accessible to the uninitiated reader and practitioners with working knowledge in linear algebra, multivariate differential calculus, and some basic topological notions in finite dimension. No space is wasted in the presentation of the theoretical tools, and the weighting of the different topics is clearly motivated by their importance for the subsequent study. The algorithmic part of the book definitely puts an emphasis on a unified, clean, and rigorous convergence analysis rather than on implementation details and numerical issues. Applications are usually presented after the convergence analysis has been carried out and are not used to motivate the algorithmic study but rather to illustrate the versatility of the methods. This part of the book is only slightly selective in terms of the algorithms studied, focusing on the most current ones, and not treating cutting-plane or bundle methods. It is, however, very comprehensive as to how the algorithms are studied, including dual counterparts, acceleration schemes, and specially tailored variants in the strongly convex case and stochastic versions.

I would particularly like to mention the crystal clear presentation of the accelerated proximal gradient method (FISTA), which I think is a particular gem. It is, in my opinion, better and clearer than most attempts made to explain only the special case of accelerated gradient descent. However, a reference to Nesterov's acceleration scheme (on which FISTA relies) could have been placed earlier, in Chapter 10, rather than only in the appendix.

A standing assumption almost throughout the second part of the book is that the

smooth part of the objective functions is actually L -smooth. Clearly, this is not always needed to establish mere convergence; however, for the sake of uniformity and with the focus on convergence rates, this is more than justified.

The author's choice not to deal with termination criteria for the methods presented is a deliberate choice, and even though as a reader I would have liked to see some

discussion in this regard, it is absolutely acceptable to say that this is beyond the scope and purpose of the book.

All in all, I am a big fan of the book and I strongly recommend it to everybody working in continuous optimization, machine learning, image processing, or related areas.

TIM HOHEISEL
McGill University

Copyright ©Society for Industrial and Applied Mathematics