LARGE-SCALE CONVEX OPTIMIZATION

Starting from where a first course in convex optimization leaves off, this text presents a unified analysis of first-order optimization methods – including parallel-distributed algorithms – through the abstraction of monotone operators. With the increased computational power and availability of big data over the past decade, applied disciplines have demanded that larger and larger optimization problems be solved. This text covers the first-order convex optimization methods that are uniquely effective at solving these large-scale optimization problems. Readers will have the opportunity to construct and analyze many well-known classical and modern algorithms using monotone operators, and walk away with a solid understanding of the diverse optimization algorithms. Graduate students and researchers in mathematical optimization, operations research, electrical engineering, statistics, and computer science will appreciate this concise introduction to the theory of convex optimization algorithms.

Ernest K. Ryu is Assistant Professor of Mathematical Sciences at Seoul National University. He previously served as Assistant Adjunct Professor with the Department of Mathematics at the University of California, Los Angeles from 2016 to 2019, before joining Seoul National University in 2020. He received a BS with distinction in physics and electrical engineering from the California Institute of Technology in 2010; and then an MS in statistics and a PhD – with the Gene Golub Best Thesis Award – in computational mathematics at Stanford University in 2016. His current research focuses on mathematical optimization and machine learning.

Wotao Yin is Director of the Decision Intelligence Lab with Alibaba Group (US), DAMO Academy, and a former Professor of Mathematics at the University of California, Los Angeles. He received his PhD in operations research from Columbia University in 2006. His numerous accolades include an NSF CAREER Award in 2008, an Alfred P. Sloan Research Fellowship in 2009, a Morningside Gold Medal in 2016, and a DAMO Award and an Egon Balas Prize in 2021. He invented fast algorithms for sparse optimization, image processing, and large-scale distributed optimization problems, and is among the top 1 percent of cited researchers by Clarivate Analytics. His research interests include computational optimization and its applications in signal processing, machine learning, and other data science problems.

"Ryu and Yin's *Large-Scale Convex Optimization* does a great job of covering a field with a long history and much current interest. The book describes dozens of algorithms, from classic ones developed in the 1970s to some very recent ones, in unified and consistent notation, all organized around the basic concept and unifying theme of a monotone operator. I strongly recommend it to any mathematician, researcher, or engineer who uses, or has an interest in, convex optimization."

– Stephen Boyd, Stanford University

"This is an absolute must-read research monograph for signal processing, communications, and networking engineers, as well as researchers who wish to choose, design, and analyze splitting-based convex optimization methods best suited for their perplexed and challenging engineering tasks."

– Georgios B. Giannakis, University of Minnesota

"This is a very timely book. Monotone operator theory is fundamental to the development of modern algorithms for large-scale convex optimization. Ryu and Yin provide optimization students and researchers with a self-contained introduction to the elegant mathematical theory of monotone operators, and take their readers on a tour of cutting-edge applications, demonstrating the power and range of these essential tools."

– Lieven Vandenberghe, University of California, Los Angeles

"First-order methods are the mainstream optimization algorithms in the era of big data. This monograph provides a unique perspective on various first-order convex optimization algorithms via the monotone operator theory, with which the seemingly different and unrelated algorithms are actually deeply connected, and many proofs can be significantly simplified. The book is a beautiful example of the power of abstraction. Those who are interested in convex optimization theory should not miss this book."

– Zhouchen Lin, Peking University

"The book covers topics from the basics of optimization to modern techniques such as operator splitting, parallel and distributed optimization, and stochastic algorithms. It is the natural next step after Boyd and Vandenberghe's *Convex Optimization* for students studying optimization and machine learning. The authors are experts in this kind of optimization. Some of my graduate students took the course based on this book when Wotao Yin was at UCLA. They liked the course and found the materials very useful in their research."

– Stanley Osher, University of California, Los Angeles

LARGE-SCALE CONVEX OPTIMIZATION

Algorithms & Analyses via Monotone Operators

ERNEST K. RYU
Seoul National University

WOTAO YIN
University of California, Los Angeles
and DAMO Academy, Alibaba Group

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05-06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781009160858
DOI: 10.1017/9781009160865

First published 2023

A catalogue record for this publication is available from the British Library.

ISBN 978-1-009-16085-8 Hardback

Dedicated to our wives
Bora and Rui

Contents

Cambridge University Press & Assessment
978-1-009-16085-8 — Large-Scale Convex Optimization
Ernest K. Ryu , Wotao Yin
Frontmatter
More Information

Preface

We write this book to share an elegant perspective that provides powerful higher-level insight into first-order convex optimization methods. The study of first-order convex optimization methods, which are more effective at solving large-scale optimization problems, started in the 1960s and 1970s, but the field at the time was focused rather on second-order methods, which are more effective at solving smaller problems. It was in the 2000s that increased computation power and the availability of big data brought first-order optimization methods into the mainstream. During this modern era, the authors entered the field of optimization and discovered (but did not invent) the perspective mentioned above, and we wish to share it through this book.

> *Our goal is to present a unified analysis of convex optimization algorithms through the abstraction of monotone operators.*

The widespread modern use of first-order methods makes this perspective more relevant than ever for both researchers and users of optimization.

This book has a somewhat unconventional organization: the chapters are structured around the techniques for deriving and analyzing optimization methods, rather than around optimization methods themselves. Through this organization, we aim to provide structure to the theory and achieve intellectual economy in that we present and analyze many optimization methods with a handful of mathematical concepts. The result is, we hope, a book that serves as a concise introduction to the theory of convex optimization algorithms.

We should also explain what this book is not. This book is not a text on monotone operator theory. We use monotone operators as a means to the end of developing and analyzing optimization algorithms, but we do not focus on the study of monotone operators themselves. This book is not a comprehensive reference on the best convex optimization methods or the strongest convergence analyses. We utilize a handful of techniques to derive and analyze optimization methods, and we only present methods and results that fit this approach.

Audience

This book is meant for both mathematicians and engineers. We appeal to mathematicians by showing that the abstraction is elegant and, in some aspects, challenging

Cambridge University Press & Assessment
978-1-009-16085-8 — Large-Scale Convex Optimization
Ernest K. Ryu , Wotao Yin
Frontmatter
More Information

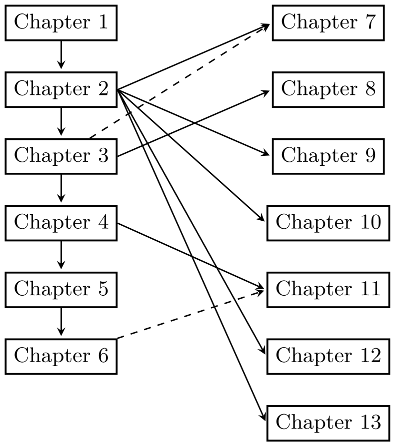

Figure 0.1 Chapter dependencies. Solid arrows denote hard dependence, while dashed arrows denote soft dependence. For example, Chapter 11 can be read after Chapter 4, but an understanding of the materials up to Chapter 6 is beneficial.

(interesting). We appeal to engineers, users of optimization, with the simplicity of the techniques and the diversity of the algorithms. In several instances, we have met engineers who know only gradient descent and ADMM, which, although powerful, are not universally feasible or best choices. This book empowers the reader to choose and even design the splitting methods best suited for any given problem.

The background required of the reader is a good knowledge of advanced calculus, linear algebra, basic probability, and basic notions of convex analysis on the topics of convex sets, convex functions, convex optimization problems, and convex duality at the level of chapters 2 through 5 of Boyd and Vandenberghe's *Convex Optimization*. Background in (mathematical) analysis and measure-theoretic probability theory is helpful but not necessary.

Informally, this book presupposes interest in convex optimization, and an appreciation of it as a useful tool. To keep the discussion concise, we focus on optimization algorithms without discussing the engineering and science origins of the optimization problems that the algorithms solve. Boyd and Vandenberghe's *Convex Optimization* is an excellent reference on the applications.

Note to Instructors

The material of this book can be taught in 15 weeks of a graduate or advanced undergraduate course. We have taught this book in an undergraduate course at SNU after covering the first five chapters of Boyd and Vandenberghe's *Convex Optimization* and in a graduate-level course at UCLA. The chapters of Part I should be taught in a linear order, while the chapters of Part II can be selected independently. Figure 0.1 illustrates the chapter dependencies. While the book does not delve deeply into the analysis of any single method, it covers many methods, as listed in Table 1. In our experience, many students appreciate the variety rather than the depth of the coverage.

This book contains almost no discussion of applications. Students without prior exposure to applications may find lectures solely on algorithms dry, so an instructor using this book may need to supplement the lectures with applications of interest to the audience.

Table 1 Optimization methods covered in each chapter

Chapters	Methods
Chapter 2	Gradient descent, dual ascent, proximal point method, method of multipliers, proximal method of multipliers, forward-backward splitting, Douglas–Rachford splitting, Davis–Yin splitting, proximal gradient method, iterative soft thresholding, consensus optimization, forward-Douglas–Rachford, variable metric proximal point, variable metric forward-backward splitting, backward-backward method, averaged alternating modified reflections, PPXA
Chapter 3	ADMM, alternating minimization algorithm (Tseng), PDHG (Chambolle–Pock), Condat–Vũ, proximal method of multipliers with function linearization, PAPC/PDFP^{2}O, linearized method of multipliers, PD3O, proximal ADMM, linearized ADMM, Chen–Teboulle, DYS 3-block ADMM, doubly-linearized method of multipliers.
Chapter 5	Coordinate gradient descent block-coordinate descent, coordinate proximal-gradient descent, stochastic dual coordinate ascent, MISO/Finito, coordinate updates on conic programs.
Chapter 6	ARock, asynchronous coordinate gradient descent, asynchronous ADMM.
Chapter 7	Stochastic forward-backward method, stochastic gradient descent, stochastic proximal gradient method, stochastic proximal simultaneous gradient method, stochastic Condat–Vũ.
Chapter 8	Function-linearized proximal ADMM, golden ratio ADMM, doubly-linearized ADMM, partial linearization, near-circulant splitting, Jacobi ADMM, 2-1-2 ADMM, Trip-ADMM, split Bregman method, four-block 2-1-2-4-3-4 ADMM.
Chapter 11	Distributed ADMM, decentralized ADMM, distributed gradient descent, method of diffusion, adapt-then-combine, PG-EXTRA, NIDS.
Chapter 12	Nesterov accelerated gradient method, FISTA, accelerated proximal point method.

For example, at SNU, we discussed engineering and machine learning applications from Boyd and Vandenberghe's *Convex Optimization*.

The textbook contains adequate homework exercises with varying levels of difficulties; some basic exercises complement the main exposition, while the difficult ones are designed to challenge the mathematically gifted students. We have also made public course material, including lecture slides and videos, on the website
`https://large-scale-book.mathopt.com/`
to help prospective instructors prepare for their lectures.

Acknowledgments

This book was greatly improved by the suggestions of Pontus Giselsson, Shuvomoy Das Gupta, Howard Heaton, Jongmin Lee, Daniel McKenzie, Chanwoo Park, Jisun Park, Ruoyu Sun, Jaewook Suh, Matthew Tam, and Taeho Yoon. We also thank the students in our courses who provided us with valuable feedback.

We also acknowledge Stephen P. Boyd, the Ph.D. advisor of Ernest Ryu. Boyd has a writing style of extreme clarity, and Ernest Ryu has strived to learn from and emulate it. Those familiar with Boyd's work may recognize his influence. In particular, Chapter 2 has much overlap with the review paper "Primer on Monotone Operator Methods," written by Ernest Ryu and Stephen Boyd in 2016 [RB16].