Cambridge University Press
978-1-316-51898-4 — Optimization for Data Analysis
Stephen J. Wright , Benjamin Recht
Frontmatter
More Information

Optimization for Data Analysis

Optimization techniques are at the core of data science, including data analysis and machine learning. An understanding of basic optimization techniques and their fundamental properties provides important grounding for students, researchers, and practitioners in these areas. This text covers the fundamentals of optimization algorithms in a compact, self-contained way, focusing on the techniques most relevant to data science. An introductory chapter demonstrates that many standard problems in data science can be formulated as optimization problems. Next, many fundamental methods in optimization are described and analyzed, including gradient and accelerated gradient methods for unconstrained optimization of smooth (especially convex) functions; the stochastic gradient method, a workhorse algorithm in machine learning; the coordinate descent approach; several key algorithms for constrained optimization problems; algorithms for minimizing nonsmooth functions arising in data science; foundations of the analysis of nonsmooth functions and optimization duality; and the back-propagation approach, relevant to neural networks.

STEPHEN J. WRIGHT holds the George B. Dantzig Professorship, the Sheldon Lubar Chair, and the Amar and Balinder Sohi Professorship of Computer Sciences at the University of Wisconsin–Madison. He is a Discovery Fellow in the Wisconsin Institute for Discovery and works in computational optimization and its applications to data science and many other areas of science and engineering. Wright is also a fellow of the Society for Industrial and Applied Mathematics (SIAM) and recipient of the 2014 W. R. G. Baker Award from IEEE for most outstanding paper, the 2020 Khachiyan Prize by the INFORMS Optimization Society for lifetime achievements in optimization, and the 2020 NeurIPS Test of Time award. He is the author and coauthor of widely used textbooks and reference books in optimization, including *Primal Dual Interior-Point Methods* and *Numerical Optimization*

BENJAMIN RECHT is Associate Professor in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. His research group studies how to make machine learning systems more robust to interactions with a dynamic and uncertain world by using mathematical tools from optimization, statistics, and dynamical systems. Recht is the recipient of a Presidential Early Career Award for Scientists and Engineers, an Alfred P. Sloan Research Fellowship, the 2012 SIAM/MOS Lagrange Prize in Continuous Optimization, the 2014 Jamon Prize, the 2015 William O. Baker Award for Initiatives in Research, and the 2017 and 2020 NeurIPS Test of Time awards.

Cambridge University Press
978-1-316-51898-4 — Optimization for Data Analysis
Stephen J. Wright , Benjamin Recht
Frontmatter
More Information

Optimization for Data Analysis

STEPHEN J. WRIGHT
University of Wisconsin–Madison

BENJAMIN RECHT
University of California, Berkeley

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/9781316518984
DOI: 10.1017/9781009004282

First published 2022

Printed in the United Kingdom by TJ Books Ltd, Padstow Cornwall

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

Library of Congress Cataloging-in-Publication Data
Names: Wright, Stephen J., 1960– author. | Recht, Benjamin, author.
Title: Optimization for data analysis / Stephen J. Wright and Benjamin Recht.
Description: New York : Cambridge University Press, [2021] | Includes
bibliographical references and index.
Identifiers: LCCN 2021028671 (print) | LCCN 2021028672 (ebook) |
ISBN 9781316518984 (hardback) | ISBN 9781009004282 (epub)
Subjects: LCSH: Big data. | Mathematical optimization. | Quantitative
research. | Artificial intgelligence. | BISAC: MATHEMATICS / General |
MATHEMATICS / General
Classification: LCC QA76.9.B45 W75 2021 (print) | LCC QA76.9.B45 (ebook)
| DDC 005.7–dc23
LC record available at https://lccn.loc.gov/2021028671
LC ebook record available at https://lccn.loc.gov/2021028672

ISBN 978-1-316-51898-4 Hardback

Contents

Preface

Optimization formulations and algorithms have long played a central role in data analysis and machine learning. Maximum likelihood concepts date to Gauss and Laplace in the late 1700s; problems of this type drove developments in unconstrained optimization in the latter half of the 20th century. Mangasarian's papers in the 1960s on pattern separation using linear programming made an explicit connection between machine learning and optimization in the early days of the former subject. During the 1990s, optimization techniques (especially quadratic programming and duality) were key to the development of support vector machines and kernel learning. The period 1997–2010 saw many synergies emerge between regularized / sparse optimization, variable selection, and compressed sensing. In the current era of deep learning, two optimization techniques—stochastic gradient and automatic differentiation (a.k.a. back-propagation)—are essential.

This book is an introduction to the basics of continuous optimization, with an emphasis on techniques that are relevant to data analysis and machine learning. We discuss basic algorithms, with analysis of their convergence and complexity properties, mostly (though not exclusively) for the case of convex problems. An introductory chapter provides an overview of the use of optimization in modern data analysis, and the final chapter on differentiation provides several perspectives on gradient calculation for functions that arise in deep learning and control. The chapters in between discuss gradient methods, including accelerated gradient and stochastic gradient; coordinate descent methods; gradient methods for problems with simple constraints; theory and algorithms for problems with convex nonsmooth terms; and duality-based methods for constrained optimization problems. The material is suitable for a one-quarter or one-semester class at advanced undergraduate or early graduate level. We and our colleagues have made extensive use of drafts of this material in the latter setting.

CAMBRIDGE
Cambridge University Press
978-1-316-51898-4 — Optimization for Data Analysis
Stephen J. Wright , Benjamin Recht
Frontmatter
More Information

This book has been a work in progress since about 2010, when we began to revamp our optimization courses, trying to balance the viewpoints of practical optimization techniques against renewed interest in non-asymptotic analyses of optimization algorithms. At that time, the flavor of analysis of optimization algorithms was shifting to include a greater emphasis on worst-case complexity. But algorithms were being judged more by their worst-case bounds rather than by their performance on practical problems in applied sciences. This book occupies a middle ground between analysis and practice.

Beginning with our courses CS726 and CS730 at University of Wisconsin, we began writing notes, problems, and drafts. After Ben moved to UC Berkeley in 2013, these notes became the core of the class EECS227C. Our material drew heavily from the evolving theoretical understanding of optimization algorithms. For instance, in several parts of the text, we have made use of the excellent slides written and refined over many years by Lieven Vandenberghe for the UCLA course ECE236C. Our presentation of accelerated methods reflects a trend in viewing optimization algorithms as dynamical systems, and was heavily influenced by collaborative work with Laurent Lessard and Andrew Packard. In choosing what material to include, we tried to not be distracted by methods that are not widely used in practice but also to highlight how theory can guide algorithm selection and design by applied researchers.

We are indebted to many other colleagues whose input shaped the material in this book. Moritz Hardt initially inspired us to try to write down our views after we presented a review of optimization algorithms at the bootcamp for the Simons Institute Program on Big Data in Fall 2013. He has subsequently provided feedback on the presentation and organization of drafts of this book. Ashia Wilson was Ben's TA in EECS227C, and her input and notes helped us to clarify our pedagogical messages in several ways. More recently, Martin Wainwright taught EECS227C and provided helpful feedback, and Jelena Diakonikolas provided corrections for the early chapters after she taught CS726. André Wibisono provided perspectives on accelerated gradient methods, and Ching-pei Lee gave useful advice on coordinate descent. We are also indebted to the many students who took CS726 and CS730 at Wisconsin and EECS227C at Berkeley who found typos and beta-tested homework problems, and who continue to make this material a joy to teach. Finally, we would like to thank the Simons Institute for supporting us on multiple occasions, including Fall 2017 when we both participated in their program on Optimization.

Madison, Wisconsin, USA
Berkeley, California, USA