Springer Optimization and Its Applications

VOLUME 133

Aims and Scope
Optimization has been expanding in all directions at an astonishing rate during the last few decades. New algorithmic and theoretical techniques have been developed, the diffusion into other disciplines has proceeded at a rapid pace, and our knowledge of all aspects of the field has grown even more profound. At the same time, one of the most striking trends in optimization is the constantly increasing emphasis on the interdisciplinary nature of the field. Optimization has been a basic tool in all areas of applied mathematics, engineering, medicine, economics and other sciences.

The series *Springer Optimization and Its Applications* publishes undergraduate and graduate textbooks, monographs and state-of-the-art expository works that focus on algorithms for solving optimization problems and also study applications involving such problems. Some of the topics covered include nonlinear optimization (convex and nonconvex), network flow problems, stochastic optimization, optimal control, discrete optimization, multi-objective programming, description of software packages, approximation techniques and heuristic approaches.

More information about this series at http://www.springer.com/series/7393

Jan A. Snyman · Daniel N. Wilke

Practical Mathematical Optimization

Basic Optimization Theory and Gradient-Based Algorithms

Second Edition

Springer

Jan A. Snyman
Department of Mechanical
and Aeronautical Engineering
University of Pretoria
Pretoria
South Africa

Daniel N. Wilke
Department of Mechanical
and Aeronautical Engineering
University of Pretoria
Pretoria
South Africa

Additional material to this book can be downloaded from http://extras.springer.com.

ISSN 1931-6828 ISSN 1931-6836 (electronic)
Springer Optimization and Its Applications
ISBN 978-3-319-77585-2 ISBN 978-3-319-77586-9 (eBook)
https://doi.org/10.1007/978-3-319-77586-9

Library of Congress Control Number: 2018935221

Mathematics Subject Classification (2010): 65K05, 90C30, 90C26, 90C59

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To our wives and friends
Alta and Ella

Preface to the second edition

The first edition (2005) of Practical Mathematical Optimization has proved to be a rigorous yet practical introduction to the fundamental principles of mathematical optimization. As stated in the preface to the first edition also included in this new edition, the aim of the text was to equip the reader with the basic theory and algorithms to allow for the solution of practical problems with confidence and in an informed way.

However, since the publication of the first edition more than a decade ago, the complexity of and computing requirements for the solution of mathematical optimization problems have significantly increased. The accessibility and definition of computing platforms have expanded by huge proportions. The fundamental physical limitations on speeding up central processing units have spurred on the advancement of multi-core computing environments that now regularly include graphical processing units. The diversity of software platforms is ever expanding with new domain and computing specific software platforms being released weekly. This edition addresses these recent advancements together with novel ideas already touched on in the first edition that have since matured considerably. They include the handling of noise in objective functions and gradient-only optimization strategies introduced and discussed in the first edition. In order to assist in the coverage of further developments in this area, and in particular of recent work in the application of mainly gradient-only methods to piecewise smooth discontinuous objective functions, it is a pleasure to welcome as co-author for this edition the younger colleague, Daniel N. Wilke.

This second edition of Practical Mathematical Optimization now takes account of the above recent developments and aims to bring this text up to date. Thus, this book now includes a new and separate chapter dedicated to advanced gradient-only formulated solution strategies for optimizing noisy objective functions, specifically piecewise smooth discontinuous objective functions, for which solution formulations and strategies are thoroughly covered. A comprehensive set of alternative solution strategies are presented that include gradient-only line search methods and gradient-only approximations. The application of these strategies is illustrated by application to well-motivated example problems. Also new to this edition is a dedicated chapter on the construction of surrogate models using only zero-order information, zero- and first-order information, and only first-order information. The latter approach being particularly effective in constructing smooth surrogates for discontinuous functions.

A further addition is a chapter dedicated to numerical computation which informs students and practicing scientists and engineers on ways to easily setup and solve problems without delay. In particular, the scientific computing language *Python* is introduced, which is available on almost all computing platforms ranging from dedicated servers and desktops to smartphones. Thus, this book is accompanied by a *Python* module `pmo`, which makes all algorithms presented in this book easily accessible as it follows the well-known `scipy.optimize.minimize` convention. The module is designed to allow and encourage the reader to include their own optimization strategies within a simple, consistent, and systematic framework. The benefit to graduate students and researchers is evident, as various algorithms can be tested and compared with ease and convenience.

To logically accommodate the new material, this edition has been restructured into two parts. The basic optimization theory that covers introductory optimization concepts and definitions, search techniques for unconstrained minimization, and standard methods for constrained optimization is covered in the first five chapters to form Part I. This part contains a chapter of detailed worked-out example problems, while other chapters in Part I are supplemented by example problems and exercises that can be done by hand using only pen, paper, and a calculator. In Part II, the focus shifts to computer applications of relatively new and mainly gradient-based numerical strategies and algorithms that are covered over four chapters. A dedicated computing chapter using *Python* is included as the final chapter of Part II, and the reader is encouraged to consult this chapter as required to complete the exercises in the preceding three chapters. The chapters in Part II are also supplemented by numerical exercises that are specifically designed so as to encourage the students to plan, execute, and reflect on numerical investigations. In summary, the twofold purpose of these questions is to allow the reader, in the first place, to gain a deeper understanding of the conceptual material presented and, secondly, to assist in developing systematic and scientific numerical investigative skills that are so crucial for the modern-day researcher, scientist, and engineer.

Jan Snyman and Nico Wilke
Pretoria
30 January 2018

Preface to the first edition

It is intended that this book is used in senior- to graduate-level semester courses in optimization, as offered in mathematics, engineering, computer science, and operations research departments. Hopefully, this book will also be useful to practicing professionals in the workplace.

The contents of this book represent the fundamental optimization material collected and used by the author, over a period of more than twenty years, in teaching Practical Mathematical Optimization to undergraduate as well as graduate engineering and science students at the University of Pretoria. The principal motivation for writing this work has not been the teaching of mathematics per se, but to equip students with the necessary fundamental optimization theory and algorithms, so as to enable them to solve practical problems in their own particular principal fields of interest, be it physics, chemistry, engineering design, or business economics. The particular approach adopted here follows from the author's own personal experiences in doing research in solid-state physics and in mechanical engineering design, where he was constantly confronted by problems that can most easily and directly be solved via the judicious use of mathematical optimization techniques. This book is, however, not a collection of case studies restricted to the above-mentioned specialized research areas, but is intended to convey the basic optimization principles and algorithms to a general audience in such a way that, hopefully, the application to their own practical areas of interest will be relatively simple and straightforward.

Many excellent and more comprehensive texts on practical mathematical optimization have of course been written in the past, and I am much indebted to many of these authors for the direct and indirect influence their work has had in the writing of this monograph. In the text, I have tried as far as possible to give due recognition to their contributions. Here, however, I wish to single out the excellent and possibly underrated book of D. A. Wismer and R. Chattergy (1978), which served to introduce the topic of nonlinear optimization to me many years ago, and which has more than casually influenced this work.

With so many excellent texts on the topic of mathematical optimization available, the question can justifiably be posed: Why another book and what is different here? Here, I believe, for the first time in a relatively brief and introductory work, due attention is paid to certain inhibiting difficulties that can occur when fundamental and classical gradient-based algorithms are applied to real-world problems. Often students, after having mastered the basic theory and algorithms, are disappointed to find that due to real-world complications (such as the presence of noise and discontinuities in the functions, the expense of function evaluations, and an excessive large number of variables), the basic algorithms they have been taught are of little value. They then discard, for example, gradient-based algorithms and resort to alternative non-fundamental methods. Here, in Chapter 4 (now Chapter 6) on new gradient-based methods, developed by the author and his co-workers, the above-mentioned inhibiting real-world difficulties are discussed, and it is shown how these optimization difficulties may be overcome without totally discarding the fundamental gradient-based approach.

The reader may also find the organization of the material in this book somewhat novel. The first three chapters present the basic theory, and classical unconstrained and constrained algorithms, in a straightforward manner with almost no formal statement of theorems and presentation of proofs. Theorems are of course of importance, not only for the more mathematically inclined students, but also for practical people interested in constructing and developing new algorithms. Therefore, some of the more important fundamental theorems and proofs are presented separately in Chapter 6 (now Chapter 5). Where relevant, these theorems are referred to in the first three chapters. Also, in order to prevent cluttering, the presentation of the basic material in Chapters 1 to 3 is interspersed with very few worked-out examples. Instead, a generous number of worked-out example problems are presented separately in Chapter 5 (now Chapter 4), in more or less the same order as the presentation of the corresponding theory given in Chapters 1 to 3. The separate presentation of the example problems may also be convenient for students who have to prepare for the inevitable tests and examinations. The instructor may also use these examples as models to easily formulate similar problems as additional exercises for the students, and for test purposes.

Although the emphasis of this work is intentionally almost exclusively on gradient-based methods for nonlinear problems, this book will not be complete if only casual reference is made to the simplex method for solving linear programming (LP) problems (where of course use is also made of gradient information in the manipulation of the gradient vector $\mathbf{c}$ of the objective function, and the gradient vectors of the constraint functions contained in the matrix $\mathbf{A}$). It was therefore decided to include, as Appendix A, a short introduction to the simplex method for LP problems. This appendix introduces the simplex method along the lines given by Chvatel (1983) in his excellent treatment of the subject.

The author gratefully acknowledges the input and constructive comments of the following colleagues to different parts of this work: Nielen Stander, Albert Groenwold, Ken Craig, and Danie de Kock. A special word of thanks goes to Alex Hay. Not only did he significantly contribute to the contents of Chapter 4 (now Chapter 6), but he also helped with the production of most of the figures and in the final editing of the manuscript. Thanks also to Craig Long who assisted with final corrections and to Alna van der Merwe who typed the first LaTeX draft.

Jan Snyman
Pretoria
31 May 2004

Contents

PART II: GRADIENT-BASED ALGORITHMS

Table of notation

$\mathbb{R}^n$	n-dimensional Euclidean (real) space
^{T}	(superscript only) transpose of a vector or matrix
$\mathbf{x}$	column vector of variables, a point in $\mathbb{R}^n$ $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$
$\in$	element in the set
$f(\mathbf{x}), f$	objective function
$\mathbf{x}^*$	local optimizer
$\mathbf{x}_g^*$	non-negative associated gradient projection point
$f(\mathbf{x}^*)$	optimum function value
$g_j(\mathbf{x}), g_j$	j^{th} inequality constraint function
$\mathbf{g}(\mathbf{x})$	vector of inequality constraint functions
$h_j(\mathbf{x}), h_j$	j^{th} equality constraint function
$\mathbf{h}(\mathbf{x})$	vector of equality constraint functions
C^1	set of continuous differentiable functions
C^2	set of continuous and twice continuous differentiable functions
$\min, \min\limits_{\mathbf{x}}$	minimize w.r.t. $\mathbf{x}$
$\mathbf{x}^0, \mathbf{x}^1, \ldots$	vectors corresponding to points 0,1,...
$\{\mathbf{x} \vert \ldots\}$	set of elements $\mathbf{x}$ such that ...
$\dfrac{\partial f}{\partial x_i}$	first partial derivative w.r.t. x_i
$\dfrac{\partial \mathbf{h}}{\partial x_i}$	$= [\dfrac{\partial h_1}{\partial x_i}, \dfrac{\partial h_2}{\partial x_i}, \ldots, \dfrac{\partial h_r}{\partial x_i}]^T$
$\dfrac{\partial \mathbf{g}}{\partial x_i}$	$= [\dfrac{\partial g_1}{\partial x_i}, \dfrac{\partial g_2}{\partial x_i}, \ldots, \dfrac{\partial g_m}{\partial x_i}]^T$
$\boldsymbol{\nabla}$	first derivative operator
$\boldsymbol{\nabla}_A$	first associated derivative operator

$\boldsymbol{\nabla} f(\mathbf{x})$	gradient vector $= \left[\frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{x})\right]^T$
$\boldsymbol{\nabla}^2$	second derivative operator (elements $\frac{\partial^2}{\partial x_i \partial x_j}$)
$\mathbf{H}(\mathbf{x}) = \boldsymbol{\nabla}^2 f(\mathbf{x})$	Hessian matrix (second derivative matrix)
$\left.\frac{df(\mathbf{x})}{d\lambda}\right\vert_{\mathbf{u}}$	directional derivative at $\mathbf{x}$ in the direction $\mathbf{u}$
$\subset, \subseteq$	subset of
$\vert \cdot \vert$	absolute value
$\vert\vert \cdot \vert\vert$	Euclidean norm of vector
$\cong$	approximately equal
$F(\ )$	line search function
$F[,]$	first order divided difference
$F[,,]$	second order divided difference
$(\mathbf{a}, \mathbf{b})$	scalar product of vector $\mathbf{a}$ and vector $\mathbf{b}$
$\mathbf{I}$	identity matrix
θ_j	j^{th} auxiliary variable
L	Lagrangian function
λ_j	j^{th} Lagrange multiplier
$\boldsymbol{\lambda}$	vector of Lagrange multipliers
$\exists$	exists
$\Rightarrow$	implies
$\{\cdots\}$	set
$V[\mathbf{x}]$	set of constraints violated at $\mathbf{x}$
ϕ	empty set
$\mathcal{L}$	augmented Lagrange function
$\langle a \rangle$	maximum of a and zero
$\frac{\partial \mathbf{h}}{\partial \mathbf{x}}$	$n \times r$ Jacobian matrix $= [\boldsymbol{\nabla} h_1, \boldsymbol{\nabla} h_2, \dots, \boldsymbol{\nabla} h_r]$
$\frac{\partial \mathbf{g}}{\partial \mathbf{x}}$	$n \times m$ Jacobian matrix $= [\boldsymbol{\nabla} g_1, \boldsymbol{\nabla} g_2, \dots, \boldsymbol{\nabla} g_m]$
s_i	slack variable
$\mathbf{s}$	vector of slack variables
D	determinant of matrix $\mathbf{A}$ of interest in $\mathbf{Ax} = \mathbf{b}$
D_j	determinant of matrix $\mathbf{A}$ with j^{th} column replaced by $\mathbf{b}$
$\lim_{i \to \infty}$	limit as i tends to infinity