PRACTICAL MATHEMATICAL OPTIMIZATION

An Introduction to Basic Optimization Theory and
Classical and New Gradient-Based Algorithms

Applied Optimization

VOLUME 97

Series Editors:

Panos M. Pardalos
University of Florida, U.S.A.

Donald W. Hearn
University of Florida, U.S.A.

PRACTICAL MATHEMATICAL OPTIMIZATION

An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms

By

JAN A. SNYMAN
University of Pretoria, Pretoria, South Africa

Springer

Library of Congress Cataloging-in-Publication Data

A C.I.P. record for this book is available from the Library of Congress.

AMS Subject Classifications: 65K05, 90C30, 90-00

ISBN 0-387-24348-8 e-ISBN 0-387-24349-6 Printed on acid-free paper.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 11161813

springeronline.com

To
Alta
my wife and friend

Contents

Preface

It is intended that this book be 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 practising professionals in the workplace.

The contents of the 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 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 organisation 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. 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, 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 non-linear problems, the 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, 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

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
$f(\mathbf{x}^*)$	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,...
$\{\mathbf{x} \| \ldots\}$	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
$\boldsymbol{\nabla} f(\mathbf{x}) = \mathbf{g}(\mathbf{x})$	gradient vector $= \left[\dfrac{\partial f}{\partial x_1}(\mathbf{x}), \dfrac{\partial f}{\partial x_2}(\mathbf{x}), \ldots, \dfrac{\partial f}{\partial x_n}(\mathbf{x})\right]^T$ (here $\mathbf{g}$ not to be confused with the inequality constraint function vector)

$\boldsymbol{\nabla}^2$	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 $= [\boldsymbol{\nabla} h_1, \boldsymbol{\nabla} h_2, \ldots, \boldsymbol{\nabla} h_r]$
$\frac{\partial \mathbf{g}}{\partial \mathbf{x}}$	$n \times m$ Jacobian matrix $= [\boldsymbol{\nabla} g_1, \boldsymbol{\nabla} g_2, \ldots, \boldsymbol{\nabla} g_m]$
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 $\mathbf{b}$
$\lim\limits_{i \to \infty}$	limit as i tends to infinity