

---

# **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.

© 2005 Springer Science+Business Media, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, Inc., 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

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

<b>PREFACE</b>	<b>xv</b>
<b>TABLE OF NOTATION</b>	<b>xix</b>
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 What is mathematical optimization? . . . . .	1
1.2 Objective and constraint functions . . . . .	4
1.3 Basic optimization concepts . . . . .	6
1.3.1 Simplest class of problems: Unconstrained one-dimensional minimization . . . . .	6
1.3.2 Contour representation of a function of two variables ( $n = 2$ ) . . . . .	7
1.3.3 Contour representation of constraint functions . . . . .	10
1.3.4 Contour representations of constrained optimization problems . . . . .	11
1.3.5 Simple example illustrating the formulation and solution of an optimization problem . . . . .	12
1.3.6 Maximization . . . . .	14
1.3.7 The special case of Linear Programming . . . . .	14

1.3.8	Scaling of design variables . . . . .	15
1.4	Further mathematical prerequisites . . . . .	16
1.4.1	Convexity . . . . .	16
1.4.2	Gradient vector of $f(\mathbf{x})$ . . . . .	18
1.4.3	Hessian matrix of $f(\mathbf{x})$ . . . . .	20
1.4.4	The quadratic function in $\mathbb{R}^n$ . . . . .	20
1.4.5	The directional derivative of $f(\mathbf{x})$ in the direction $\mathbf{u}$ . . . . .	21
1.5	Unconstrained minimization . . . . .	21
1.5.1	Global and local minima; saddle points . . . . .	23
1.5.2	Local characterization of the behaviour of a multi- variable function . . . . .	24
1.5.3	Necessary and sufficient conditions for a strong local minimum at $\mathbf{x}^*$ . . . . .	26
1.5.4	General indirect method for computing $\mathbf{x}^*$ . . . . .	28
1.6	Exercises . . . . .	31
<b>2</b>	<b>LINE SEARCH DESCENT METHODS FOR UNCON- STRAINED MINIMIZATION</b>	<b>33</b>
2.1	General line search descent algorithm for unconstrained minimization . . . . .	33
2.1.1	General structure of a line search descent method . . . . .	34
2.2	One-dimensional line search . . . . .	35
2.2.1	Golden section method . . . . .	36
2.2.2	Powell's quadratic interpolation algorithm . . . . .	38
2.2.3	Exercises . . . . .	39

2.3	First order line search descent methods . . . . .	40
2.3.1	The method of steepest descent . . . . .	40
2.3.2	Conjugate gradient methods . . . . .	43
2.4	Second order line search descent methods . . . . .	49
2.4.1	Modified Newton's method . . . . .	50
2.4.2	Quasi-Newton methods . . . . .	51
2.5	Zero order methods and computer optimization subroutines . . . . .	52
2.6	Test functions . . . . .	53
<b>3</b>	<b>STANDARD METHODS FOR CONSTRAINED OPTI- MIZATION</b>	<b>57</b>
3.1	Penalty function methods for constrained minimization . .	57
3.1.1	The penalty function formulation . . . . .	58
3.1.2	Illustrative examples . . . . .	58
3.1.3	Sequential unconstrained minimization technique (SUMT) . . . . .	59
3.1.4	Simple example . . . . .	60
3.2	Classical methods for constrained optimization problems . . . . .	62
3.2.1	Equality constrained problems and the La- grangian function . . . . .	62
3.2.2	Classical approach to optimization with inequality constraints: the KKT conditions . . . . .	70
3.3	Saddle point theory and duality . . . . .	73
3.3.1	Saddle point theorem . . . . .	73



3.3.2	Duality . . . . .	74
3.3.3	Duality theorem . . . . .	74
3.4	Quadratic programming . . . . .	77
3.4.1	Active set of constraints . . . . .	78
3.4.2	The method of Theil and Van de Panne . . . . .	78
3.5	Modern methods for constrained optimization . . . . .	81
3.5.1	The gradient projection method . . . . .	81
3.5.2	Multiplier methods . . . . .	89
3.5.3	Sequential quadratic programming (SQP) . . . . .	93
<b>4</b>	<b>NEW GRADIENT-BASED TRAJECTORY AND AP- PROXIMATION METHODS</b>	<b>97</b>
4.1	Introduction . . . . .	97
4.1.1	Why new algorithms? . . . . .	97
4.1.2	Research at the University of Pretoria . . . . .	98
4.2	The dynamic trajectory optimization method . . . . .	100
4.2.1	Basic dynamic model . . . . .	100
4.2.2	Basic algorithm for unconstrained problems (LFOP)	101
4.2.3	Modification for constrained problems (LFOPC)	101
4.3	The spherical quadratic steepest descent method . . . . .	105
4.3.1	Introduction . . . . .	105
4.3.2	Classical steepest descent method revisited . . . . .	106
4.3.3	The SQSD algorithm . . . . .	107

4.3.4	Convergence of the SQSD method . . . . .	108
4.3.5	Numerical results and conclusion . . . . .	113
4.3.6	Test functions used for SQSD . . . . .	117
4.4	The Dynamic-Q optimization algorithm . . . . .	119
4.4.1	Introduction . . . . .	119
4.4.2	The Dynamic-Q method . . . . .	119
4.4.3	Numerical results and conclusion . . . . .	123
4.5	A gradient-only line search method for conjugate gradient methods . . . . .	126
4.5.1	Introduction . . . . .	126
4.5.2	Formulation of optimization problem . . . . .	127
4.5.3	Gradient-only line search . . . . .	128
4.5.4	Conjugate gradient search directions and SUMT .	133
4.5.5	Numerical results . . . . .	135
4.5.6	Conclusion . . . . .	139
4.6	Global optimization using dynamic search trajectories . .	139
4.6.1	Introduction . . . . .	139
4.6.2	The Snyman-Fatti trajectory method . . . . .	141
4.6.3	The modified bouncing ball trajectory method . .	143
4.6.4	Global stopping criterion . . . . .	145
4.6.5	Numerical results . . . . .	148
<b>5</b>	<b>EXAMPLE PROBLEMS</b>	<b>151</b>
5.1	Introductory examples . . . . .	151

5.2	Line search descent methods . . . . .	157
5.3	Standard methods for constrained optimization . . . . .	170
5.3.1	Penalty function problems . . . . .	170
5.3.2	The Lagrangian method applied to equality constrained problems . . . . .	172
5.3.3	Solution of inequality constrained problems via auxiliary variables . . . . .	181
5.3.4	Solution of inequality constrained problems via the Karush-Kuhn-Tucker conditions . . . . .	184
5.3.5	Solution of constrained problems via the dual problem formulation . . . . .	191
5.3.6	Quadratic programming problems . . . . .	194
5.3.7	Application of the gradient projection method . . . . .	199
5.3.8	Application of the augmented Lagrangian method . . . . .	202
5.3.9	Application of the sequential quadratic program- ming method . . . . .	203
<b>6</b>	<b>SOME THEOREMS</b>	<b>207</b>
6.1	Characterization of functions and minima . . . . .	207
6.2	Equality constrained problem . . . . .	211
6.3	Karush-Kuhn-Tucker theory . . . . .	215
6.4	Saddle point conditions . . . . .	218
6.5	Conjugate gradient methods . . . . .	223
6.6	DFP method . . . . .	227

<b>A THE SIMPLEX METHOD FOR LINEAR PROGRAMMING PROBLEMS</b>	<b>233</b>
A.1 Introduction . . . . .	233
A.2 Pivoting for increase in objective function . . . . .	235
A.3 Example . . . . .	236
A.4 The auxiliary problem for problem with infeasible origin .	238
A.5 Example of auxiliary problem solution . . . . .	239
A.6 Degeneracy . . . . .	241
A.7 The revised simplex method . . . . .	242
A.8 An iteration of the RSM . . . . .	244

# 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 Chvatal (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  $\text{\LaTeX}$  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 $\mathbb{R}^n$ $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$
$\in$	element in the set
$f(\mathbf{x}), f$	objective function
$\mathbf{x}^*$	local optimizer
$f(\mathbf{x}^*)$	optimum function value
$g_j(\mathbf{x}), g_j$	$j^{\text{th}}$ inequality constraint function
$\mathbf{g}(\mathbf{x})$	vector of inequality constraint functions
$h_j(\mathbf{x}), h_j$	$j^{\text{th}}$ 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
$\min, \min$	minimize w.r.t. $\mathbf{x}$
$\mathbf{x}^0, \mathbf{x}^1, \dots$	vectors corresponding to points 0,1,...
$\{\mathbf{x} \mid \dots\}$	set of elements $\mathbf{x}$ such that ...
$\frac{\partial f}{\partial x_i}$	first partial derivative w.r.t. $x_i$
$\frac{\partial \mathbf{h}}{\partial x_i}$	$= \left[ \frac{\partial h_1}{\partial x_i}, \frac{\partial h_2}{\partial x_i}, \dots, \frac{\partial h_r}{\partial x_i} \right]^T$
$\frac{\partial \mathbf{g}}{\partial x_i}$	$= \left[ \frac{\partial g_1}{\partial x_i}, \frac{\partial g_2}{\partial x_i}, \dots, \frac{\partial g_m}{\partial x_i} \right]^T$
$\nabla$	first derivative operator
$\nabla f(\mathbf{x}) = \mathbf{g}(\mathbf{x})$	gradient vector $= \left[ \frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right]^T$ (here $\mathbf{g}$ not to be confused with the inequality constraint function vector)



$\nabla^2$	second derivative operator (elements $\frac{\partial^2}{\partial x_i \partial x_j}$ )
$\mathbf{H}(\mathbf{x}) = \nabla^2 f(\mathbf{x})$	Hessian matrix (second derivative matrix)
$\left. \frac{df(\mathbf{x})}{d\lambda} \right _{\mathbf{u}}$	directional derivative at $\mathbf{x}$ in the direction $\mathbf{u}$
$\subset, \subseteq$	subset of
$ \cdot $	absolute value
$\ \cdot\ $	Euclidean norm of vector
$\cong$	approximately equal
$F(\cdot)$	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
$\theta_j$	$j^{\text{th}}$ auxiliary variable
$L$	Lagrangian function
$\lambda_j$	$j^{\text{th}}$ Lagrange multiplier
$\boldsymbol{\lambda}$	vector of Lagrange multipliers
$\exists$	exists
$\Rightarrow$	implies
$\{\dots\}$	set
$V[\mathbf{x}]$	set of constraints violated at $\mathbf{x}$
$\phi$	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 = $[\nabla h_1, \nabla h_2, \dots, \nabla h_r]$
$\frac{\partial \mathbf{g}}{\partial \mathbf{x}}$	$n \times m$ Jacobian matrix = $[\nabla g_1, \nabla g_2, \dots, \nabla g_m]$
$s_i$	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 $j^{\text{th}}$ column replaced by $\mathbf{b}$
$\lim_{i \rightarrow \infty}$	limit as $i$ tends to infinity