Simplicial Algorithms for Minimizing Polyhedral Functions

Polyhedral functions provide a model for an important class of problems that includes both linear programming and important applications in data analysis. General methods for minimizing such functions using the polyhedral geometry explicitly are developed. Such methods approach a minimum by moving from extreme point to extreme point along descending edges and are described generically as simplicial. The best known member of this class is the simplex method of linear programming, but simplicial methods have found important applications in discrete approximation and statistics. In particular, the general approach considered here has permitted the development of finite algorithms for the rank regression problem, a fascinating technique of statistical estimation. The key ideas are those of developing a general format for specifying the polyhedral function and the application of this to derive multiplier conditions to characterize optimality. Also considered is the application of the general approach to the development of active set algorithms for polyhedral function constrained problems and associated Lagrangian forms. Such methods are finding current application in statistics and data mining.

M. R. Osborne has held positions in Australia, Britain, and the United States during his long and distinguished academic career and is the author or editor of six books. He is currently Professor Emeritus in the Centre for Mathematics and its Applications at Australian National University in Canberra, and a Fellow of the Australian Academy of Science.

Simplicial Algorithms for Minimizing Polyhedral Functions

M. R. OSBORNE
Australian National University

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2001

Printed in the United States of America

Typeface Times Roman 10/13 pt. *System* LaTeX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data
Osborne, M. R. (Michael Robert)
Simplicial algorithms for minimizing polyhedral functions / M. R. Osborne.
p. cm.
ISBN 0-521-79133-2
1. Algorithms. 2. Polyhedral functions. I. Title.
QA9.58.O73 2000
511.8 – dc21 00-036287

ISBN 0 521 79133 2 hardback

Contents

Preface

Preliminary Remarks

Polyhedral functions provide a model for an important class of problems that includes not only the linear programming problem but also important applications in data analysis. The data analysis problems can involve nonconvex as well as convex objective functions. This book presents methods for minimizing such functions that use the polyhedral geometry explicitly. Such methods approach a minimum by descending from vertex to vertex along descending edges and are described generically as *simplicial*. The best known member of this class is the simplex method for solving linear programming problems, but the simplicial approach has found important applications in discrete approximation where it is applied to solve various formulations of l_1 and l_∞ problems. Polyhedral functions provide a convenient framework and make it possible to attempt a broadly based approach to the general class of problems that can be solved by simplicial methods. These problems have been described in fair detail in my book *Finite Algorithms in Optimization and Data Analysis* [45]. Here much of the same ground is covered with an emphasis on more recent developments. Thus the basic convex analysis that was previously treated in some detail is summarized in a preliminary chapter; there is less attention given to linear programming and l_1 fitting problems, and considerably more given to the general class of separable piecewise linear functions, and to minimizing the rank regression functional. The derivation of finite algorithms for this latter problem appears to require this general approach and provides one of the more significant justifications for it. A new departure is the development of active set methods for polyhedral constrained problems and their associated Lagrangians. Problems of this kind are occurring in significant new applications in statistics and the emerging area of data mining.

The main tools are much the same as those developed in the earlier book. Differences or extensions include:

- Some properties of generalized derivatives in the sense of [11] have been included as a preliminary to considering some aspects of the nonconvex problems.
- The structure functional formalism introduced in [45] is developed to give a general basis for representing polyhedral functions by means of their differential structure. Structure functionals play a role in simplicial methods analogous to that of constraints in linear programming. Here properties of completeness and redundancy are introduced to relate them directly to the geometry of the epigraph of the polyhedral convex function. Extreme points are defined by the vanishing of suitable sets of structure functionals, and edges correspond to exactly one structure functional in such a set relaxing from zero while the remaining members remain active.
- The treatment of degeneracy in linear programming in [45] has been widely appreciated. Here it has been expanded to reflect practical experience gained subsequently.
- Homotopy methods for postoptimality studies in linear programming and quantile regression problems are considered. It is shown that the homotopy paths are piecewise linear in the dual variables and piecewise constant in the state variables. The successive piecewise constant states are linked by directions of nonuniqueness, and these transitions occur at the changes in slope of the dual variables. The corresponding calculation has the character of a pivoting step in a simplicial algorithm.
- An extended discussion of the problems of minimizing piecewise linear, separable functions is presented. Here the simplicial algorithms make a linesearch in the descent edge direction, and it is shown that it is often possible to do considerably better in the sense of taking fewer iterations to locate the minimum than is possible with the sorting based algorithms currently in favor. The first numerical treatment of a nonconvex problem in this class is probably due to Womersley [68]. This work is discussed in some detail here. Of interest is the apparent inherent ability of the simplicial methods to avoid aspects of the nonconvex structure.
- The rank regression problem is treated in considerable detail. Derivation of the conjugate function and the Fenchel dual problem are considered (comparison with the l_1 dual is interesting). A reduced gradient algorithm is developed, and the fascinating asymptotic linearity properties are used both to provide a least squares method for computing a consistent initial point and to justify the use of a secant based linesearch. Programming details are discussed in some

depth to make the point that this is a case where object oriented language facilities serve the algorithm implementer well. Nonconvex problems appear to have a natural place in this class of robust estimation procedures, and it is of interest that the difficulty avoiding property observed for piecewise linear separable problems appears to hold in this case also.

- Interest in variable selection problems in statistics and in the analysis and summarizing of very large data sets in data mining has led to consideration of optimization problems subject to a polyhedral constraint and to related unconstrained Lagrangian problems that incorporate polyhedral functions in the objective. The interest here is in the development of descent algorithms that permit compact treatment of the polyhedral component.

The attention given to interior point methods, and the excitement generated by them, make the utility of simplicial methods a fair question. It seems that even in the linear programming case, where it is clear that interior point methods have an important role to play, the answer is not completely clear. Certainly it seems that simplicial methods will retain their importance for postoptimality problems, where it is necessary to move from the solution of a given problem to one that is a close neighbor in an appropriate sense, even if the first solution is obtained by an interior point method. However, there are examples where the simplicial algorithms are known to be effective. The simplicial ascent (exchange) algorithm for the l_∞ problem is discussed in [45], Appendix 2, where it is shown to live close to a second-order convergent algorithm for continuous approximation problems and so must be expected to be competitive in most circumstances. A rigorous treatment is given in [57]. Both the linear programming and l_∞ problems can be characterized as having simple vertex structures in the nondegenerate case. Here the main interest is in what will be shown to be more complex cases. The l_1 problem is perhaps the simplest of these, and an argument is developed in [45], Appendix 2, that simplicial methods can be expected to be efficient for the l_1 problem in a somewhat similar sense to that in the l_∞ exchange algorithm. The l_1 problem has a Fenchel dual, which is the well-known interval linear program, and it is argued that one can expect to solve this dual problem efficiently by interior point methods also [55]. There may not be much difference between the two approaches in this case provided the linesearch is performed adequately (some new comparisons are presented here). The situation seems less clear for piecewise linear problems with more complex structure. In the rank regression case the Fenchel dual is derived in Chapter 2 and bears a family resemblance to the l_1 form. However, the constraint set has enormously greater complexity in general. How to begin to implement an interior point method becomes a question of interest. Here the property of asymptotic linearity can be exploited to provide a consistent starting point for the simplicial algorithm, and the resulting hybrid algorithm appears very effective.

Acknowledgments

There is a very special role reserved for Karen George in the genesis of this work. Her significant contributions are obvious enough from the bibliography, but this tells little of the real story. Only problems of ill health have prevented her from participating very much as an equal partner in the putting together of this material. Chapter 5 is my account of aspects of her doctoral research, and her contributions are strongly evident also both in the treatment of degeneracy and in the discussion of the linesearch computations. In a very real sense she is responsible for stimulating much of the work that goes beyond the "Finite Algorithms" book [45].

Much of the special character of this work derives from efforts to develop an accessible simplicial algorithm for the rank regression problem. I am grateful to Allan Miller for drawing my attention to the significance of this problem.

Other collaborators who have had a major influence on the development of the presented material include David Clark, Stephen Pruess, David Ryan, Berwin Turlach, Alistair Watson, and Rob Womersley. The references summarize their contributions but do not do justice to their impact. Alistair Watson and Rob Womersley have helped significantly by making detailed comments on the text. The nonconvex objective function figures in Chapter 5 were prepared by Alistair Watson and are a result of computations by Alistair, Karen George, and Rob Womersley. Chapter 6 is a direct result of a visit to the ANU by Alistair Watson in which recent work on the variable selection problem ([49], [50]) was extended to the Lagrangian formulation.

Notation

$\{x; P(x)\}$	set of elements with property P
$\{x\}$	set consisting of a single element
$\mathcal{X}$	point set
$\mathcal{X}^o$	interior of $\mathcal{X}$
Co	set complement
ν	index set $\{i, j, k, \ldots\}$ or $\{\nu(1), \nu(2), \ldots\}$
$\lvert\nu\rvert$	number of elements in ν
R	$\{x; -\infty < x < \infty\}$
R^p	$R \times R \times R \cdots \times R$ (p terms)
$H(\mathbf{u}, \nu)$	hyperplane (set $\{\mathbf{x}; \mathbf{u}^T\mathbf{x} = \nu\}$)
$H^+(\mathbf{u}, \nu)$	$\{\mathbf{x}; \mathbf{u}^T\mathbf{x} > \nu\}$
$H^-(\mathbf{u}, \nu)$	$\{\mathbf{x}; \mathbf{u}^T\mathbf{x} \leq \nu\}$
$\mathcal{C}$	cone, $\mathbf{x} \in \mathcal{C} \Rightarrow \lambda\mathbf{x} \in \mathcal{C}\, \forall\lambda \geq 0$
$\mathcal{C}^*$	polar cone, $\{\mathbf{x}; \mathbf{x}^T\mathbf{y} \leq 0,\ \forall\mathbf{y} \in \mathcal{C}\}$
$\mathcal{L}^\perp$	orthogonal complement of linear space $\mathcal{L}$
∂	subdifferential (of convex function)
∇	gradient operator
$f'(\mathbf{x} : \mathbf{t})$	directional derivative in direction $\mathbf{t}$
$f^o(\mathbf{x} : \mathbf{t})$	generalized directional derivative
$\mathcal{T}(\mathcal{S}, \mathbf{x})$	tangent cone to $\mathcal{S}$ at $\mathbf{x}$
$\mathbf{L}(\mathbf{x}, \mathbf{u})$	Lagrangian
$\delta(\mathbf{x} \mid \mathcal{S})$	indicator function of $\mathcal{S}$
$\delta^*(\mathbf{x} \mid \mathcal{S})$	support function of $\mathcal{S}$
A	matrix (dimensions from context)
A^T	matrix transpose
A_ν	submatrix, rows pointed to by ν
cl	closure
A_{k*}	kth row of A

A_{*j}	jth column of A
$\mathbf{x}$	vector (dimension from context)
$\mathbf{x}_\sigma$	subvector pointed to by σ
$\mathbf{e}_i$	ith coordinate vector
$\mathbf{e}$	$\mathbf{e} \in R^p = \sum_{i=1}^{p} \mathbf{e}_i$ (dimension from context)
I_p	unit matrix of dimension p
$\|\cdot\|$	norm
$\|\cdot\|_\alpha$	$\alpha = 1$, l_1 norm; $\alpha = 2$, l_2 norm; $\alpha = \infty$, l_∞ norm
aff	affine hull
arg	argument
cond	condition number
conv	convex hull
det	determinant
dim	dimension
eff	effective domain
epi	epigraph
lim	limit
rank	rank (of matrix)
ri	relative interior
sgn	sign
span	linear hull
trace	trace of matrix