PROBLEM COMPLEXITY AND METHOD EFFICIENCY IN OPTIMIZATION

A. S. Nemirovsky
D. B. Yudin

Wiley-Interscience Series in Discrete Mathematics

PROBLEM COMPLEXITY AND METHOD EFFICIENCY IN OPTIMIZATION

A. S. NEMIROVSKY

Senior Scientific Fellow, State University of Moscow, USSR.

D. B. YUDIN

Professor of Mathematical Methods, Faculty of Economics,
State University of Moscow, USSR

Translated by

E. R. DAWSON,

University of Dundee

A Wiley–Interscience Publication

JOHN WILEY & SONS

Chichester · New York · Brisbane · Toronto · Singapore

Translated from SLOZHNOST' ZADACH I EFFEKTIVNOST' METODOV OPTIMIZATSII by E. R. Dawson

Library of Congress Cataloging in Publication Data:

Nemirovsky, Arkadii Semenovich.
  Problem complexity and method efficiency in optimization.

  (Wiley series in discrete mathematics)
  'A Wiley–Interscience publication.'
  Translation of: Slozhnost' zadach i effektivnost'
metodov optimizatsii.
  Bibliography: p.
  Includes index.
  1. Programming (Mathematics) I. Yudin, David
Borisovich. II. Title. III. Series: Wiley–Interscience
series in discrete mathematics.
QA402.5.N45413 1983 519.7 82–11065
ISBN 0 471 10345 4

British Library Cataloguing in Publication Data:

Nemirovsky, A. S.
  Problem complexity and method efficiency in
  optimization—(Wiley series in discrete
  mathematics)
  1. Mathematical optimization
  I. Title II. Yudin, D. B. III. Slozhnost'
  zadach i effecktivnost' motodov optimizatsii.
  English
515 QA402.5

ISBN 0 471 10345 4

Typeset by Macmillan India Ltd., Bangalore

Printed in Great Britain by
Page Bros (Norwich) Ltd

PROBLEM COMPLEXITY AND METHOD EFFICIENCY IN OPTIMIZATION

WILEY-INTERSCIENCE SERIES IN DISCRETE MATHEMATICS

ADVISORY EDITORS: Ronald L. Graham
Bell Laboratories
Murray Hill, New Jersey

Dr Jan Karel Lenstra
Stichting Mathematisch Centrum
Amsterdam, The Netherlands

Graham, Rothschild and Spencer · RAMSEY THEORY

Tucker · APPLIED COMBINATORICS

Pless · INTRODUCTION TO THE THEORY OF ERROR-CORRECTING CODES

Nemirovsky and Yudin · PROBLEM COMPLEXITY AND METHOD EFFICIENCY IN OPTIMIZATION

Goulden and Jackson · COMBINATORIAL ENUMERATION

Gondran and Minoux · GRAPHS AND ALGORITHMS

Preface

There is an extensive literature devoted to numerical methods of solving extremal problems. The central position in it is occupied by studies relating to actual algorithms for optimization—their description, range of applicability, rate of convergence. There are also a number of papers on empirical comparison (based on examples) of the effectiveness of algorithms, with a view to selecting the 'best' method from the existing arsenal of such methods.

Far less attention has been paid to questions such as the following. What, in general, can be expected from numerical methods of solving problems of a given type? What are the potentially attainable limits of these methods? How complicated are problems of one sort or another, not with respect to a given concrete method, but in relation to 'all (in general)' methods of solution? It is this natural question of the 'potentially attainable effectiveness of numerical methods applied to problems of a given type' which forms the subject of investigation in this monograph.

A typical question which we shall consider is of this form. Given a family of optimization problems together with a source of information, accessible to the methods, about each solvable problem of this family, what are the potential lower bounds of laboriousness of all possible methods which solve all problems of the family with a given accuracy? Which method realizes this potential lower limit and is therefore the best one? Clearly, a precise formulation of such a question requires a formalization of the concepts of 'method', 'laboriousness of a method', etc. We fix a definite formalization of this kind (in our opinion, the most convenient formalization for studying the 'continuous' mathematical programming problems with which we shall be concerned), and we then investigate the question posed above as applied to a number of the standard non-linear programming problems: smooth multi-extremal problems, 'all convex' problems, strongly convex problems, and so on. In most cases a sufficiently conclusive answer is successfully obtained to the question in which we are interested.

We do not claim that the formalization we have adopted of the problem of selecting the best possible numerical method of optimization is entirely adequate to represent the full content of the original formulation. There is no need to dwell now on this aspect of the matter, since a full motivation of the approach we have adopted and a discussion of its merits and limitations are given in Section 1.1. We remark only that, in our opinion, the formalization we have adopted enables rather useful, if rough, information (orders of magnitude only) to be obtained about the potential possibilities of numerical methods of solving extremal problems of the standard types. It is up to the reader to share or repudiate this opinion.

Another question closely related to that of finding potential lower bounds for the laboriousness of methods which solve problems of a given class with a given error, i.e. with the problem (as we have defined it) of computing the 'complexity' of a class, is the question of estimating the effectiveness of one or other method of solving problems of this class. It is natural to define this effectiveness as the inverse of the ratio of the laboriousness of the method in question to a 'standard laboriousness', i.e. to the complexity of the class. The effectiveness of a method shows to what extent it can be improved as regards laboriousness, i.e. to what extent it is non-optimal.

We touch on only a small part of the problem of estimating the effectiveness of traditional numerical methods of optimization. This is understandable; to solve this problem it is not sufficient merely to have available a standard of effectiveness (basically our efforts are concentrated precisely on obtaining such a standard); it is further necessary to have estimates of the laboriousness and error of the method in which we are interested on the class of problems under consideration. The profusion of standard numerical methods obviously precludes the possibility of the authors' being able to estimate these characteristics for some arbitrary representative of the group of methods.

In this book the effectiveness of some of the most natural and simple methods is evaluated. For reasons which will become clear later, all these methods are the methods of convex programming. Of the methods of non-smooth convex programming, we evaluate the gradient method and the Kelly method (these might be said to exhaust the list of traditional algorithms for non-smooth convex optimization). The extensive field of algorithms for the minimization of smooth and (strongly) convex problems is examined to a far less extent; here we restrict our attention to the gradient method with minimization in the anti-gradient direction and to some simple versions of the method of conjugate directions. The methods considered for strongly convex programming problems turn out to be inefficient; they are unnecessarily sensitive to the degree of conditionality of the problem under consideration, and their effectiveness tends to zero as the conditionality deteriorates. We remark that negative results of this kind also enable certain conclusions to be drawn regarding the effectiveness of a number of traditional methods which are not explicitly considered in this book.

Let us give an example. An extensive family of methods of feasible directions for solving constrained convex problems is known. The rate of convergence of these methods is generally estimated under the hypothesis of strong convexity of the problem under solution. Thus it is natural to study these methods on the class of strongly convex problems. In Chapter 7 it is shown that the complexity of this class is determined essentially only by the required accuracy, the conditionality of the problem, and its dimension, but not by the number of constraints. On the other hand, if there are no constraints, then most versions of the method of feasible directions turn into the method of gradient descent with minimization in the direction of descent. So the effectiveness of the methods of feasible directions cannot be essentially greater than that of the gradient method, and therefore it too tends to zero as the conditionality of the problem deteriorates.

The limited size of the book does not allow us to dwell on consequences of this sort; they will certainly be self-evident to the reader. We remark further that the judgement expressed previously about the method of feasible directions being ineffective (like similar statements in the main text) is a judgement made on the basis of the definition of laboriousness which has been adopted and which turned out to be not quite adequate for the intuitively understood computational complexity of a method. It should not therefore be interpreted as a call for unconditional discrimination against the corresponding methods; categorical verdicts of this kind are scarcely admissible generally.

We mention some differences of the approaches adopted in this monograph, and of the results obtained, from the traditional treatments in optimization theory.

The traditional approach to estimates of the rate of convergence of numerical methods of optimization is usually of an asymptotic character; the type of asymptotic behaviour of the laboriousness of a method for a required accuracy is elucidated. The question of when the 'exit' on to this asymptotic behaviour takes place is investigated comparatively rarely, as is, incidentally, the important question of the effect of other parameters, apart from accuracy, of the class of problems (parameters such as, for example, the dimension of the problems under solution).

In this monograph, on the other hand, all the estimates given for the laboriousness of numerical methods of optimization (as also, incidentally, for most of the estimates of complexity) are of a non-asymptotic character. We write down in explicit form the upper bounds for the methods under examination, as a function of the required accuracy (measured in a sensible way) and of the parameters which distinguish the class of problems to be solved (such as the dimensionality of the problem, the number of constraints, the characteristics of the geometry of the domain of the problem, etc.).

In the literature on methods of optimization the rate of convergence (the rate, not the fact of convergence itself) is established generally only as applied to classes of sufficiently 'good' problems. In the present monograph optimization methods are studied as a rule on wider classes of problems. In most parts of the book nothing more is required of the problems to be solved other than their convexity, unless perhaps the boundedness of the domain of the problem and the continuity of the functionals appearing in it (sometimes these functionals have to satisfy a Lipschitz condition). Even the classes of strongly convex problems considered in Chapter 7 are still somewhat wider than classes of 'good' functions on which it is traditionally accepted that optimization methods should be evaluated. Of course, evaluating methods on wider families of problems than usual leads to a worsening of the potentially attainable guarantees of their work. It turns out, however, that in many cases this worsening is not too considerable, and it represents an acceptable payment for the extension of the store of problems over which the new guarantees are extended.

Having sketched in general terms the purpose of the monograph, we shall briefly characterize the contents of the work by sections (a more detailed survey of the results is given in Section 1.1). The first chapter is of an introductory character: here we form the language in which later we formulate and solve the problems concerning the potential effectiveness of numerical methods. The separation of the later material into chapters is dictated by the necessity to examine separately the various classes of extremal problems. In the last section of Chapter 1, smooth (but not necessarily convex) problems are examined. The results in this section, some well-known, some new, are of a negative character (it turns out that this class of problem is, in the general case, hopelessly too complicated for solution). Convex programming, to which the main attention is paid in the book, presents a much more optimistic picture. In Chapters 2, 3, and 4 we examine classes of 'all (in general)' convex problems (including non-smooth ones) which can be solved by first-order methods when there is exact calculation of the values and derivatives of the components in the problem. In Chapters 5 and 6 we study convex problems in which the values and derivatives of the components are observed mixed with noise, i.e. problems of convex stochastic programming. In Chapter 7 the classes of smooth convex problems and strongly convex problems are considered, and in Chapter 9 we deal with problems solvable by zeroth-order methods, i.e. methods working with the values but not with the derivatives of the components in the problem. Chapter 8 stands somewhat by itself; it deals with the estimation of the laboriousness of some popular methods of solving strongly convex problems. The appendix contains a résumé of a number of classical mathematical concepts and theorems which are used in the book and which are not always familiar to applied mathematicians. In general it should be mentioned that the treatment in this book is such that it should be accessible to a reader having the normal training of a numerical analyst interested in optimization theory. More complicated mathematical apparatus is brought in only in the 'formal niceties' (of secondary importance) in some of the proofs.

The exposition of material relating to concrete classes of extremal problems usually comprises the following main steps:

(1) a description of the class of problems in question;
(2) a description of some methods of solving problems of the given class;
(3) a lower bound (over all conceivable methods of solving the problems of a given class) for the potentially attainable laboriousness of these methods for a given error.

As a rule the choice of actual methods in (2) leads to estimates for the laboriousness which essentially are the same as the potential lower bounds in (3). As a result we obtain, on the one hand, a sufficiently complete idea of the 'objective complexity' of the given class of problems, and on the other hand, a basis for the theoretical recommendations on the use of the methods in (2) which cannot essentially be improved as regards laboriousness. We mention that the 'sub-optimal' methods which we adduce are, in a number of cases, in substantial measure new.

We point out that a reader can, if he wishes, restrict his attention to the sections which deal with some particular class of problems.

As regards the nature of the exposition it is worth mentioning the following. We have tried to distinguish as clearly as possible the ideas which form the basis of the constructions here presented, and to describe the numerical methods precisely. The formal proofs are kept separate; some of them are given in separate sections. In a first reading these proofs could, if so desired, be omitted, and this would not hinder the reader from using the methods described, although it would make detailed understanding of their mechanism more difficult.

Many of the results are formulated as exercises, inviting the reader to prove some proposition which has only been formulated. If some not entirely trivial fact is concerned, then often a proof is given (enclosed in angular brackets $\langle \ldots \rangle$). We stress that the reader should acquaint himself with the propositions enunciated in the exercises, whether or not he actually carries out the exercises themselves.

Regarding terminology: apart from the standard terminology, which we use without special explanation, we have to use extensively a number of specific concepts and special notation. On first encountering non-standard notation and terms appearing without commentary, the reader should consult the list of notation at the end of the book, or the subject index, where he will find a reference to the section in which the corresponding object is first defined. An exception occurs with certain secondary concepts and notation used only in some one chapter. Accordingly, to understand a particular section of a chapter, the reader will generally need to be acquainted with all the preceding sections of that chapter.

A few words about the bibliography and references. The list includes only those works which are directly referred to in the main text; the list is short and makes no claim to cite all the works which deal with the theme of the monograph. When using the results of others, the authors have *tried* to mention the fact by pointing out the source of their information (without setting themselves the task of identifying the original source without fail). In speaking of results which have become part of the chrestomathy, so to say, we have replaced direct citation by phrases such as 'it is well known that . . .'. It is quite possible that some of the results regarded by the authors as original may in actual fact be re-discoveries of already known facts (that is why we used the word 'tried' in a previous sentence); in that case we beg in advance the pardon of the first discoverers.

In conclusion we regard it as our pleasant duty to thank E. G. Goldshtein and B. T. Polyak for stimulating discussions of the results of the work.

A. S. NEMIROVSKY, D. B. YUDIN

June 1978

Preface to the English Edition

This book is one of the series 'Theory and methods of systems analysis' published under the guidance of an editorial board of economists and cyberneticists headed by D. M. Gvishiani, a Soviet philosopher and son-in-law of the former Soviet prime minister, Alexei N. Kosygin. Of the authors of the present work, Professor D. B. Yudin holds the chair of mathematical methods in the faculty of economics in Moscow university, and Dr. A. S. Nemirovsky, a senior scientific fellow at the same university, is a disciple of the late distinguished mathematician G. E. Shilov. The book is based on, and is an extension of, a series of papers by these authors published mainly in the journal *Economics and Mathematical Methods*.

The authors set up their own mathematical model in order to investigate questions concerning the complexity of optimization problems and efficiency of methods of solving them. They obtain bounds for the potential efficiency of methods of solving standard classes of optimization problems, and propose new methods which largely realize these potential bounds. They apply their apparatus to draw perhaps surprising conclusions about a number of popular methods of optimization. But, as with all mathematics, the reader must remember that the technical terms have precisely the meaning assigned to them in the definition of the concepts; this is particularly necessary when every-day words such as 'method' and 'complexity' are being used as technical terms. In particular, as the authors themselves point out, their analysis does not deal at all with such practically important aspects of methods as the simplicity of their computational organization and computational stability.

When not being strictly formal, as in the definitions and statements and proofs of theorems, the authors write in a lively and informal style, which may at times even be humorous. They have a habit of frequently putting words into quotation marks, presumably to point up and lend immediacy to their exposition; this practice has been followed in the translation.

F. R. Dawson
Department of Mathematical Sciences, University of Dundee

Contents

1

Introduction

1.1 FORMULATION OF THE PROBLEM OF ESTIMATING THE COMPLEXITY OF OPTIMIZATION PROBLEMS, AND THE MAIN RESULTS OF THE WORK. AN INFORMAL DESCRIPTION

In this section we describe informally the set of problems we shall be concerned with and the direction of our investigations. Our aim is to show that the approach adopted in this work for evaluating the complexity of problems and the effectiveness of methods is a natural one.

1.1.1

We shall study the potentialities of numerical methods in solving mathematical programming problems. We need hardly mention how important such methods are in the application of mathematics to practical problems. The widening field of applications and the power of computers is leading firstly to a sharp growth in the complexity of the optimization problems which have to be 'worked out to the answer', and secondly to a continuous reinforcement of the arsenal of methods used for this purpose. In this situation there is naturally a growing tendency to take a hard look at the theory of these methods themselves. By contemporary standards, its mere convergence gives no method the right to exist; 'decency' requires us also to estimate its laboriousness.

The next stage is that one wants to find the potentially attainable lower limits for the amount of labour needed to solve a given type of problem, and to construct methods which attain these limits, i.e. methods—in some sense optimal—which ensure solutions of the required quality for all the problems in question, with the least possible amount of labour. These are precisely the problems to which the present work is devoted. Our target is a theoretical analysis of the potentialities of numerical methods.

A strict formulation of the problems arising in this connection requires a formalization of ideas such as 'a class of problems of a given type', 'the