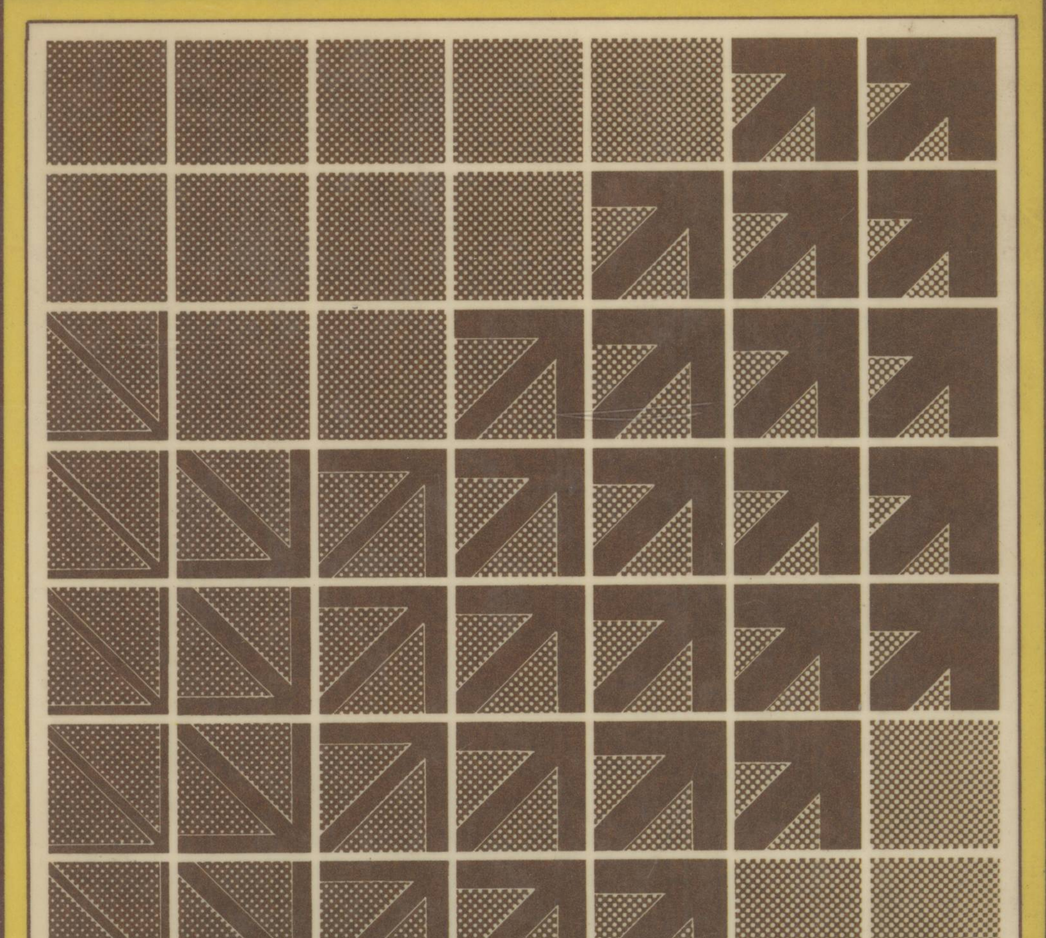
IIASA PROCEEDINGS SERIES

Nonsmooth Optimization

Proceedings of a IIASA Workshop
March 28-April 8, 1977

Claude Lemarechal and Robert Mifflin, Eds.



International Institute for Applied Systems Analysis

NONSMOOTH OPTIMIZATION

Proceedings of a IIASA Workshop,
March 28 - April 8, 1977

CLAUDE LEMARECHAL
ROBERT MIFFLIN
Editors

PERGAMON PRESS

OXFORD · NEW YORK · TORONTO · SYDNEY · PARIS · FRANKFURT

U.K.	Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England
U.S.A.	Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A.
CANADA	Pergamon of Canada, Suite 104, 150 Consumers Road, Willowdale, Ontario M2J 1P9, Canada
AUSTRALIA	Pergamon Press (Aust.) Pty. Ltd., 19a Boundary Street, Rushcutters Bay, N.S.W. 2011, Australia
FRANCE	Pergamon Press SARL, 24 rue des Ecoles, 75240 Paris, Cedex 05, France
FEDERAL REPUBLIC OF GERMANY	Pergamon Press GmbH, 6242 Kronberg-Taunus, Pferdstrasse 1, Federal Republic of Germany

First edition 1978

British Library Cataloguing in Publication Data

Nonsmooth optimization. - (International
Institute for Applied Systems Analysis.
IIASA proceedings series; Vol. 3).
1. System analysis - Congresses
2. Mathematical optimization - Congresses
I. Lemarechal, Claude II. Mifflin, Robert
III. Series
003 QA402 78-40904

ISBN 0-08-023428-3

Printed in Great Britain by
William Clowes & Sons Limited
Beccles and London

IIASA PROCEEDINGS SERIES

Volume 3

Nonsmooth Optimization

IIASA PROCEEDINGS SERIES

VOLUME 1 CARBON DIOXIDE, CLIMATE AND SOCIETY
Proceedings of a IIASA Workshop Cosponsored by WMO, UNEP, and SCOPE, February 21-24, 1978
Jill Williams, Editor

VOLUME 2 SARUM AND MRI: DESCRIPTION AND COMPARISON OF A WORLD MODEL AND A NATIONAL MODEL
Proceedings of the Fourth IIASA Symposium on Global Modelling, September 20-23, 1976
Gerhart Bruckmann, Editor

VOLUME 3 NONSMOOTH OPTIMIZATION
Proceedings of a IIASA Workshop,
March 28-April 8, 1977
Claude Lemarechal and Robert Mifflin, Editors

FORTHCOMING:

PEST MANAGEMENT
Proceedings of an International Conference,
October 25-29, 1976
G.A. Norton and C.S. Holling, Editors

METHODS AND MODELS FOR ASSESSING ENERGY RESOURCES
First IIASA Conference on Energy Resources,
May 20-21, 1975
Michel Grenon, Editor

FUTURE COAL SUPPLY FOR THE WORLD ENERGY BALANCE
Third IIASA Conference on Energy Resources,
November 28-December 2, 1977
Michel Grenon, Editor

THE SHINKANSEN HIGH-SPEED RAIL NETWORK OF JAPAN
Proceedings of a IIASA Conference,
June 27-30, 1977
A. Straszak and R. Tuch, Editors

PREFACE

Task 2 of the System and Decision Sciences Area, Optimization, is a central methodological tool of systems analysis. It is used and needed by many Tasks at IIASA, including those of the Energy Systems and the Food and Agriculture Programs. In order to deal with large-scale applications by means of decomposition techniques, it is necessary to be able to optimize functions that are not differentiable everywhere. This is the concern of the subtask Nonsmooth Optimization. Methods of nonsmooth optimization have been applied to a model for determining equilibrium prices for agricultural commodities in world trade. They are also readily applicable to some other IIASA models on allocating resources in health care systems.

This volume is the result of a workshop on Nonsmooth Optimization that met at IIASA in the Spring of 1977. It consists of papers on the techniques and theory of nonsmooth optimization, a set of numerical test problems for future experimentation, and a comprehensive research bibliography.

CONTENTS

INTRODUCTION

This volume is the result of a workshop on nonsmooth optimization held at IIASA from March 28 to April 8, 1977. The designation "First World Conference on Nonsmooth Optimization", proposed in jest by one of the participants after noting that there were only nine others in the room with him, is, however, appropriate because of the various countries represented, and because the number of scientists doing research in this field at that time was rather small.

The small number of participants, and the workshop's unusual length, made it possible to achieve a substantial exchange of information. Each morning (three working hours) was devoted to the talk of one participant, who therefore could present his work quite thoroughly. During the afternoons, discussions took place on related topics, such as: systems of inequalities, constrained problems, test problems and numerical experiments, smooth approximation of nonsmooth functions, optimization with noise, direction-finding procedures and quadratic programming, line searches, general decomposition, ... However, this workshop format would have been a failure were it not for the fact that everyone was alert and active even when not "in the spotlight". We are very grateful to all the participants, who contributed to the success of the workshop by their numerous questions and interruptions during both the formal and informal presentations.

This workshop was held under the name Nondifferentiable Optimization, but it has been recognized that this is misleading, because it suggests "optimization without derivatives". As we view it, nonsmooth optimization (NSO) is concerned with problems having functions for which gradients exist almost everywhere, but are not continuous, so that the usual gradient-based methods and results fail. The contents of these Proceedings should convince the reader of the importance of being able to compute (generalized) gradients in NSO.

We have adopted the following topical classification for the papers: subgradient optimization (three papers), descent methods (four papers), and field of applicability (one paper).

The first paper, by B.T. Poljak, exhaustively surveys the Soviet work on subgradient optimization done since 1962. For this method he gives the most important results obtained and the various extensions that have been developed.

J.L. Goffin studies rates of convergence in subgradient optimization. He shows under which conditions linear convergence can be obtained and provides bounds on the best possible rate of convergence. These bounds are given in terms of condition numbers that do not depend on derivative continuity.

The paper by R. Chaney and A.A. Goldstein addresses the question: What is the most general framework for the method of subgradient optimization to be applicable and convergent? Hence, they present the method in an abstract setting and study the minimal hypotheses required to ensure convergence.

One of the important conclusions of this workshop has been that nonsmooth optimization and nonlinear programming (NLP) are, in fact, equivalent fields. It was known that NLP is contained in NSO via exact penalty function methods, but B.N. Pshenichnyi's paper demonstrates the reverse containment via feasible direction methods.

In his paper, C. Lemarechal describes, in a unified setting, descent methods developed recently in Western countries. He also provides ideas for improvement of these methods.

Many methods for solving constrained optimization problems require the repeated solution of constrained least squares problems for search direction determination. An efficient and reliable algorithm for solving such subproblems is given in the paper by R. Mifflin.

The paper by P. Wolfe is concerned with line searches. He gives an APL algorithm that effectively deals with the issues of when to stop a line search with a satisfactory step size and how to determine the next trial step size when the stopping criterion is not met.

The last paper, by J. Gauvin, studies the differential properties of extremal value functions. This is important for the application of various decomposition schemes for solving large-scale optimization problems, because these approaches require the solution of nonsmooth problems involving extremal-value functions, and in order to guarantee convergence we need to know whether certain "semismoothness" conditions (such as Lipschitz continuity) are satisfied.

We then give four nonsmooth optimization test problems. They were selected because they are easy to work with and because they are representative both of the field of applicability and of the range of difficulty of

NSO. Problems 1 and 3 are examples of minimax problems and are not very difficult. Problem 2 is a nonconvex problem coming from a well-known NLP test problem, and problem 4 involves a piecewise-linear function. The last two are sufficiently difficult to slow down considerably the speed of convergence of any of the NSO methods we know of.

We conclude this volume with a large NSO bibliography. It was compiled by the participants and is an update of the bibliography given in *Mathematical Programming Study 3*. We wish to thank D.P. Bertsekas, V.F. Demjanov, M.L. Fisher, and E.A. Nurminskii for the items they communicated to us.

On behalf of all the participants we would like to acknowledge IIASA's generous support and to thank I. Beckey, L. Berg, A. Fildes, and G. Lindelof for their optimal organizational contributions, which led to a smooth-running workshop.

We are especially indebted to M.L. Balinski who was instrumental in establishing a Nonsmooth Optimization group at IIASA and who spent much of his time and energy to secure a truly international participation at this workshop.

C. Lemarechal
R. Mifflin

SUBGRADIENT METHODS: A SURVEY OF SOVIET RESEARCH

B. T. Poljak

This paper reviews research efforts by Soviet authors concerning the subgradient technique of nondifferentiable minimization and its extensions. It does not cover the works based on the concept of steepest descent (by V.F. Demjanov, B.N. Pshenichnyi, E.S. Levitin, and others) or on the use of a specific form of the minimizing function (for example minimax techniques). The paper essentially uses the review by N.Z. Shor [1]. The theorems given below are mostly simplified versions of results shown in the original papers.

1. THE SUBGRADIENT METHOD

Let $f(x)$ be a convex continuous function in the space R^n. A vector $\partial f(x) \in R^n$ is called its subgradient at the point x, if it satisfies the condition

$$f(x+y) \geq f(x) + (\partial f(x), y) \quad , \quad \forall\, y \in R^n \quad . \tag{1}$$

A subgradient exists (although, generally speaking, it may be not unique) for all $x \in R^n$. If $f(x)$ is differentiable, the subgradient is unique and coincides with the gradient $\partial f(x) = \nabla f(x)$. The rules of subgradient calculation for various types of functions are well known [2,3]. In particular, with $f(x) = \max_{1 \leq i \leq m} f_i(x)$ where $f_i(x)$ are convex differentiable functions, it is true that

$$\partial f(x) = \sum_{i \in I(x)} \alpha_i \nabla f_i(x) \quad , \qquad \alpha_i \geq 0 \quad ,$$

$$\sum_{i \in I(x)} \alpha_i = 1 \quad , \qquad I(x) = \{i : f_i(x) = f(x)\}$$

(for instance one may take $\partial f(x) = \nabla f_i(x)$ where $i \in I(x)$ is arbitrary).

The subgradient minimization method for $f(x)$ on R^n is an iterative process of the form

$$x_{k+1} = x_k - \gamma_k \partial f(x_k) / \| \partial f(x_k) \| \qquad (2)$$

where $\gamma_k \geq 0$ is a step size. For differentiable functions this method coincides with the gradient one. The major difference between the gradient and the subgradient methods is that, generally speaking, the direction $-\partial f(x_k)$ is not a descent direction at the point x_k; i.e., the values of $f(x_k)$ for nondifferentiable functions do not decrease monotonically in the method (2).

The subgradient method was developed in 1962 by N.Z. Shor and used by him for solving large-scale transportation problems of linear programming [4]. Although published in a low-circulation publication, this pioneering work became widely known to experts in the optimization area in the USSR. Also of great importance for the propagation of nondifferentiable optimization concepts were the reports by the same author presented in a number of conferences in 1962-1966.

Publication of papers [5,6,7] giving a precise statement of the method and its convergence theorems may be regarded as the culmination of the first stage in developing subgradient techniques.

Let us get down to describing the basic results concerning the subgradient method. As is known, the gradient method for minimization of smooth functions employs the following ways to regulate the step-size:

$$\gamma_k = \alpha \| \partial f(x_k) \| \ ,$$

i.e.

$$x_{k+1} = x_k - \alpha \nabla f(x_k) \ , \qquad 0 < \alpha < \bar{\alpha}$$

(the ordinary gradient method);

$$\gamma_k = \arg\min_{\gamma} f(x_k - \gamma \partial f(x_k)/\| \partial f(x_k) \|)^{\dagger}$$

(the steepest descent method).

Simple examples may be constructed to show that neither of these methods converges in nondifferentiable minimization; this necessitates the construction of new principles of selecting the step size. Consider the major ones. Hereinafter we shall assume $f(x)$ to be convex and continuous and denote $f^* = \inf f(x)$ and $X^* = \operatorname{Arg\,min} f(x)$.

(a) $\gamma_k = \gamma > 0$. This constant-step method was suggested in [4]. The simplest example, $f(x) = |x|$, $x \in R^1$, explicitly proves that this method does not converge. One may show, however, that it gives a solution "with an accuracy of γ".

Theorem 1 [4]

Let X^* be nonempty. Then for any $\delta > 0$ there exists $\bar{\gamma} > 0$ such that in the method (2) with $\gamma_k = \gamma$, $0 < \gamma < \bar{\gamma}$ we have $\liminf f(x_k) < f^* + \delta$.

Reference [4] has described the following way of step-size regulation resting upon this result, although it has not been entirely formalized. A certain $\gamma > 0$ is chosen and the computation is made with $\gamma_k = \gamma$ until the values of $f(x_k)$ start to oscillate about a certain limit. After this γ is halved and the process is repeated.

(b) The sequence γ_k is chosen a priori regardless of the computation results and satisfies the condition

$$\sum_{k=0}^{\infty} \gamma_k = \infty \quad , \qquad \gamma_k \to 0 \quad . \tag{3}$$

This way of choosing the step-size has been suggested in [5] and [6] independently.

---

†Hereafter $\arg\min_{\gamma} \rho(\gamma)$ will mean an arbitrary minimum point of the function $\rho(\gamma)$, $\operatorname{Arg\,min}_{\gamma} \rho(\gamma)$ is the set of all minimum points.

Theorem 2 [5,6]

In the method (2),(3) $\liminf f(x_k) = f^*$. If X^* is nonempty and bounded then $\rho(x_k, X^*) \to 0$, where

$$\rho(x, X^*) = \min_{x^* \in X^*} \| x - x^* \| \quad .$$

It is clear that in the general case the method (2),(3) cannot converge faster than γ_k tends to zero. In particular, this method never converges at the rate of geometrical progression or at the rate

$$0(k^{-s}), \quad s > 1 \quad .$$

(c) In certain cases the value of f^* is known. For instance, if

$$f(x) = \sum_{i=1}^{m} f_i(x)_+ \quad ,$$

where $f_i(x)$ are convex functions,

$$f_i(x)_t = \max \{0, f_i(x)\} \quad ,$$

and the system of inequalities $f_i(x) \leq 0$ $i = 1,\ldots,m$ is solvable, then X^* is the set of solutions of this system and $f^* = 0$. Then one may take

$$\gamma_k = \lambda \frac{(f(x_k) - f^*)}{\| \partial f(x_k) \|} \quad , \quad 0 < \lambda < 2 \quad . \tag{4}$$

In solving systems of inequalities the method (3),(4) coincides with the known relaxation method of Kaczmarz, Agmon, Motzkin, Schoenberg, and Eremin [8]. The method for general problems of nonsmooth function minimization has in essence been suggested by I.I. Eremin [9] and systematically developed in [10].

Theorem 3 [9,10]

Let x^* be the unique minimum point for $f(x)$. Then in the method (2),(4) $x_k \to x^*$. If the condition

$$f(x) - f^* \geq \ell \| x - x^* \| \quad , \quad \ell > 0 \tag{5}$$

holds, the method converges with the rate of a geometrical progression.

The advantages of the method (2),(4) are the simplicity of selecting the step size (since no auxiliary problems should be solved and no characteristics of $f(x)$ other than f^* should be known) and its applicability, since for a smooth strongly convex $f(x)$ the method also converges with the rate of a geometrical progression [10]. Reference [10] has shown a way to modify the technique when f^* is unknown.

(d) N.Z. Shor [11] has suggested an essentially different method for choosing γ_k:

$$\gamma_k = \gamma_0 q^k \quad , \quad 0 < q < 1 \quad . \tag{6}$$

Note that the condition (3) is not satisfied for (6).

Theorem 4 [11,12,13]

Let the condition

$$(\partial f(x), x - x^*) \geq \ell \| \partial f(x) \| \, \| x - x^* \| \quad , \quad \ell > 0 \tag{7}$$

hold. Then there exists a pair $\bar{q}$ (which depends on ℓ) and $\bar{\gamma}$ (which depends on $\| x_0 - x^* \|, \ell$) such that with $1 > q \geq \bar{q}$, $\gamma_0 \geq \bar{\gamma}$ in the method (2),(6) we have

$$\| x_k - x^* \| \leq C(q, \gamma_0) q^k \quad .$$

The relationship of $\bar{q}(\ell)$ and $\bar{\gamma}(\| x_0 - x^* \|, \ell)$ may be expressed explicitly. However, practical implementation of the method (2),(6) faces difficulties because generally the values of ℓ and $\| x_0 - x^* \|$ are unknown.

The above results prove that the convergence rate for any of the step-size regulating rules is linear at best. The denominator of the geometrical progression for the ill-conditioned problems (i.e. for functions with greatly extended level sets) is near unity. Thus the convergence rate of all the versions of the subgradient method may be rather poor.

2. ACCELERATING CONVERGENCE OF THE SUBGRADIENT METHOD

One of the reasons why the subgradient method converges so slowly lies in its Markov nature. The subsequent iteration makes no use of the information obtained at the previous steps. The major concept of all techniques for accelerating convergence is the use of this information (i.e. the values $f(x_i)$, $\partial f(x_i)$, $i=0,\ldots,k-1$).

The first methods of this type were those developed by Kelley and by Cheney and Goldstein [14,15], based on piecewise-linear approximation of the function. An original technique suggested in [16] and [17] independently made use only of the values $\partial f(x_i)$. Let M_k be a polyhedron in R^n in which the minimum point is localized after k iterations. Then for an arbitrary $x_{k+1} \in M_k$ one may take

$$M_{k+1} = M_k \cap \{x:(\partial f(x_{k+1}), x - x_{k+1}) \leq 0\} \quad .$$

If the center of gravity M_k is taken as the point x_{k+1} one may show [17] that for the volume V_k of the polyhedron M_k the following expression holds:

$$V_{k+1} \leq [1-(1 - \frac{1}{N-1})^N]V_k \quad ,$$

where N is the dimension of the space. Thus for problems of any dimension

$$V_k \leq V_0 q^k \quad , \quad q = 1 - e^{-1} \quad .$$