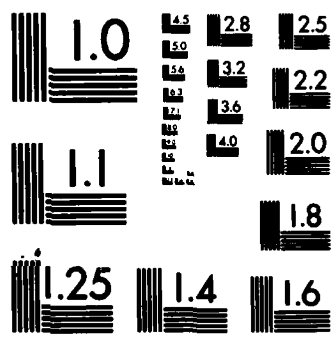
AD-A141 216 SUBGRADIENT METHODS IN DETERMINISTIC AND STOCHASTIC OPTIMZZATION(U) WASHINGTON UNIV SEATTLE DEPT OF MATHEMATICS R T ROCKAFELLAR 27 FEB 84 AFOSR-TR-84-0356 F49620-82-K-0012

1/1

UNCLASSIFIED

F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFOSR-TR- 84-0356

AD-A141 216

SUBGRADIENT METHODS IN DETERMINISTIC AND STOCHASTIC OPTIMIZATION

Final Scientific Report
Grant F4960-82-K-0012
Air Force Office of Scientific Research

R. T. Rockafellar, Principal Investigator
Department of Mathematics
University of Washington
Seattle, WA 98195

February 27, 1984

DTIC ELECTE MAY 17 1984 A

Introduction.

Research supported by this two-year grant in the period from January, 1982, through December, 1984, has resulted in a total of 11 technical articles and two doctoral theses. These range over several areas of mathematical optimization theory but share the common theme of the development and application of subgradient methods and duality to problems in mathematical programming. Fundamental advances in concept have been made, and in the case of stochastic problems, new techniques of solution have been initiated that may revolutionize the subject.

The publications are grouped under the following headings, which will be discussed individually:

1. Stochastic programming (4 papers).
2. Subgradient theory (3 papers, 1 thesis).
3. Nonlinear programming (4 papers).
4. Optimal control (1 thesis).

DTIC FILE COPY

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DTIC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12.
Distribution is unlimited.
MATTHEW J. KERPER
Chief, Technical Information Division

Approved for public release;
distribution unlimited.

84 05 15 225

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public releases; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR. 84-0356
6a. NAME OF PERFORMING ORGANIZATION University of Washington Department of Mathematics 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State and ZIP Code) Seattle, WA 98195	7b. ADDRESS (City, State and ZIP Code) Bolling AFB, DC
8a. NAME OF FUNDING/SPONSORING ORGANIZATION USAF, AFSC, AFOSR 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-82-K-0012
8c. ADDRESS (City, State and ZIP Code) Building 410 Bolling, AFB, D.C. 20332	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO.: 61102F PROJECT NO.: 2304 TASK NO.: A6 WORK UNIT NO.:
11. TITLE (Include Security Classification) See Block 19	
12. PERSONAL AUTHOR(S) Rockafellar, R.T.	
13a. TYPE OF REPORT: Final 13b. TIME COVERED: FROM Jan 82 TO Dec 83	14. DATE OF REPORT (Yr., Mo., Day): 1984 Feb 27 15. PAGE COUNT: 7
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES FIELD / GROUP / SUB. GR.	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) stochastic programming, subgradient methods, nonlinear programming, optimization 1984

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Subgradient Methods in Deterministic and Stochastic Optimization

This report surveys the eleven technical articles and two doctoral theses that were produced under this two-year grant. The work was in the following areas: (1) stochastic programming; (2) subgradient theory; (3) nonlinear programming; (4) optimal control.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Joseph Bram	22b. TELEPHONE NUMBER (Include Area Code) 22c. OFFICE SYMBOL

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

84 05 15 225

1. Stochastic Programming

Many problems of optimization require that decisions be taken before the values of certain random variables are revealed. For example, goods must be stock piled and parts must be procured before the exact demand for them is known. Little can be said in the face of total uncertainty, but in many cases there is statistical information available about the random variables in question. These are the cases to which the subject of stochastic programming is addressed.

The practice all too commonly followed of simply putting expected values in place of the random variables in a problem, and then solving the problem as it it were deterministic, has been shown to lead to poor decisions due to a lack of safety margins. The same goes for "scenario analysis" in its popular form, where several versions of what might happen are explored, but no scientific principles are invoked that take these eventualities into account in making the best compromise choices here-and-now. Stochastic programming is a relatively new discipline that helps to identify the right way to hedge against uncertainty in such situations. The theory has been under development for a number of years, but it is only now that we are reaching the stage of actually being able to solve stochastic programming problems numerically. This is chiefly due to the fact that such problems are intrinsically of very large scale (infinite-dimensional if the random variables are viewed as continuously distributed), and they can involve multiple stages in time as well. Technical progress in the design of computers has been needed in order to bring them within range of solution, but new ideas of representation and decomposition have been essential too. Some of the research under this grant has been in the forefront of these conceptual mathematical developments.

Paper [2], "Deterministic and stochastic optimization problems of Bolya type in discrete time," deals with multistage problems in stochastic programming. In such problems there is a discrete time variable $t = 0, 1, \ldots, T$. At time t, the values of certain of the random variables are revealed or at least narrowed down, and a decision vector x_t is chosen subject to certain constraints and costs that may depend in part on the preceding decisions $x_0, \ldots, x_{t-1}$. The choice of x_t in turn may affect future constraints and costs. Altogether the situation may be very complicated. The problem is to make the decisions in such a way that total expected cost is minimized.

Paper [2] seeks to identify structure of a special nature that lends hope in being able to solve such a difficult problem. The emphasis here is on being able to understand what goes on when there is significant number of time periods. Uniformity of some sort from one time period to the next is needed to keep the situation in hand. Convexity assumptions are needed to simplify matters further. This paper develops a discrete-time analogue of the Hamiltonian differential equation in the calculus of variations that serves to characterize optimality in this problem. The dual variables in this characterization are certain conditional expectations of prices. These prices serve to decompose the problem into a separate convex programming problem at each time t. The results that are obtained constitute a fundamental advance in the way that multistage problems have been handled. There is little doubt that these results will have an important role in the design of computational methods eventually, but problems with more than two time periods are still some stages away from numerical feasibility.

Paper [3], "On the interchange of subdifferentiation and conditional expectation for convex functionals," pins down a technical point that enters into the developments in [2].

In paper [7], "A dual solution procedure for quadratic stochastic programs with simple recourse," the aim is to provide a viable method of computation for a class of problems that is much narrower but nonetheless of considerable importance in applications. The problems have only two stages ("here-and-now" and "recourse") and are linear-quadratic in structure. Furthermore, the first stage decisions, while they do affect the costs and constraints in the second stage, do not have the power to make the second stage infeasible. The second stage decision process is of a particularly simple character.

The approach to such problems in [7] is to introduce an appropriate dual problem involving Lagrange multipliers that are random variables with unknown distributions. The interesting thing about this dual problem is that although it is very large in dimension it can nevertheless be used effectively as a means of solving the primal problem.

The secret is the following. The dual problem consists of maximizing a certain quadratic functional over a convex set. This cannot be tackled directly, because the quadratic functional has an inconveniently complicated expression, and the convex set is described by too many constraints. What we do is to solve a sequence of subproblems in which we maximize the functional not over the whole set, but over a polytope generated as the convex hull of a relatively small number of elements of the set, i.e. dual feasible solutions. This is possible because of the quadratic form of the functional: each subproblem can be expressed in terms of the parameters used in the convex hull representation, and the coefficients that one gets in this way are certain expectations that are readily computed! The solution to the construction of the polytope used in the next subproblem in the sequence.

This method has been programmed and has already led to solutions to problems that no one previously has been able to handle.

The ideas are developed much further in article [13], "A Lagrangian finite generation technique for solving linear-quadratic problems in stochastic programming." This paper, which is not yet finished, extends the method to a vastly larger class of problems and investigates properties of convergence. The main result is surprisingly powerful. It says that for strictly quadratic problems, the number of dual feasible solutions used in generating the polytopal representation does not have to escalate -- it can be kept fixed and one will still achieve a linear rate of convergence to the optimal solutions to the primal and dual problems. This is important because the number in question determines the dimension of the quadratic programming subproblem that must be solved in each iteration. If this number were to increase without bound, as happens in typical cutting-plane algorithms, for instance, we would soon be unable to continue.

For problems that are not already strictly quadratic, [13] provides a technique for introducing the strictly quadratic terms iteratively and still maintaining a linear rate of convergence.

A-1

2. Subgradient Theory

Applications of optimization in many areas lead to the consideration of functions which are not everywhere smooth (continuously differentiable). This is not because the data and parameters in the problems in such areas behave nonsmoothly in some pathological way. Rather it is a consequence of the very nature of optimization and the techniques that can be used in decomposing large-scale problems into smaller ones.

The basic difficulty is this. The property of smoothness is preserved under classical operations like addition, multiplication and composition of functions, but it is not preserved under operations like taking the pointwise maximum or minimum of a collection of functions, or optimizing the value of a function with respect to some of its arguments while the other arguments are still treated as variables. Additional insight into the difficulty is obtained from the geometry of constraints. In classical problems of physics and engineering, the constraints are typically in the form of systems of equations. These serve to focus our attention on a certain curve, surface, or higher dimensional smooth manifold embedded in the state space at large. If inequality constraints are present at all, they are few in number and interact in simple ways. For example, one may have a ball, cube, or some other region whose boundary is easily describable as composed of smooth pieces that joint together in regular ways. In most of the modern applications of optimization, however, the number of inequality constraints can be enormous. The characterization of the boundary of the feasible region may be very complicated. There may be no immediate way to identify just which constraints are active or inactive at a given point. It may be easier then to think of the boundary as a nonsmooth "surface", perhaps represented by the graphs of one or more nonsmooth functions.

For such reasons, the development of tools of mathematical analysis that replace classical differential calculus in certain situations has long played an important part in optimization theory. Thus even in linear programming, it has been necessary to introduce concepts of one-sided directional derivatives and subgradients of piecewise linear functions in order to understand the shadow price interpretation of dual optimal solution vectors and its implications for sensitivity analysis. Subgradients and subderivatives were first introduced by this writer. The original domain of research was convex programming and its applications and extensions in optimal control and mathematical economics. In the mid 1970's, the writer's student F.H. Clarke found the right way to extend the subgradient concept from convex functions to a far larger class of functions. This opened up all of nonlinear programming and variational theory to new methods of analysis, and today efforts are being made far and wide in using these methods to achieve a better understanding of optimization problems and their modes of representation.

Article [6], "Generalized subgradients in mathematical programming," is a survey of the theory and its main results. It was put together for a special "state-of-the-art" volume that was published in connection with the 1982 mathematical programming symposium in Bonn. This was the eleventh in the series of international meetings in mathematical programming, held every third year. At this meeting the writer was awarded the George Dantzig Prize for the contributions he has made to mathematical programming through his work on subgradients and duality.

J.S. Treiman, a Ph.D. student supported by this grant as a research assistant, has made further contributions in this area. His paper [8], "Characterization of Clarke's tangent and normal cones in finite and infinite dimensions," provides new theoretical insights. His thesis [11], "A new characterization of Clarke's tangent cone and its applications to subgradient analysis and optimization," is a very substantial piece of work indeed. It fills a major gap that has been an obstacle to progress with infinite-dimensional problems like those in optimal control and stochastic programming.

For finite-dimensional problems we have for some time been able to take advantage of two complementary approaches to the notion of "subgradient". There has been a direct approach in terms of convex hulls of limits of gradients or special sorts of subgradients taken at neighboring points, as well as in indirect approach in terms of certain directional derivatives and duality. For infinite-dimensional problems, however, only the second approach has been available. Treiman's thesis [11] provides the remedy by developing the correct extension of the first approach for a large class of infinite-dimensional spaces. This was no easy achievement and required deep understanding of Banach space geometry. Some time will be required in digesting such a fundamental theoretical advance, but it should have many long-range effects.

In the recently completed paper [12], "Extensions of subgradient calculus with applications to optimization," the writer has made numerous sharp improvements to one of the principal branches of subgradient theory, namely the formulas that can be used for calculating the subgradients of a given function from the known subgradients of other functions out of which it has been constructed. Such formulas are essential, for instance, in deriving necessary conditions for optimality in optimization problems of practically every kind. Even problems that are stated in terms of smooth functions benefit from the results, which lead ot expressions of marginal values and characterizations of stability under perturbation.

This long paper [12] was several years in the making and is the culmination of much research. Although it deals with finite-dimensional situations only, the results of Treiman mentioned earlier hold the promise of supporting a number of extensions to infinite dimensions. Preliminary work of Treiman and the writer in this direction is well under way.

3. Nonlinear Programming

One of the areas of nonlinear programming that has been improved radically by the new subgradient methods is the study of "marginal values" and perturbations. Suppose that the objective and constraint functions in a typical nonlinear programming problem depend on various parameters. Lump these together into a parameter vector v and then think of the optimal value in the problem as depending on v. Even if the objective and constraint functions are smoothly behaved, this optimal value function may be far from smooth. Here indeed lies one of the major motivations of subgradient theory: the desire to understand better how the optimal value does change with the parameter vector v and in particular to derive bounds or estimates for generalized rates of change, or so-called marginal values.

In paper [9], "Directional differentiability of the optimal value function in a mathematical programming problem," definitive results are obtained in identifying the circumstances under which directional derivatives exist in the ordinary sense. The results go far beyond what was known previously, which applied only to special perturbations in convex programming or cases in nonlinear programming that are so ideal that the optimal value function turns out to be smooth. It is interesting that to meet the challenge even of problems whose constraint and objective functions are twice continuously differentiable, all the tools of nonsmooth analysis must be brought to bear. Furthermore, a new and more complete form of second-order optimality conditions is required.

Such conditions have been developed in paper [1], "Marginal values and second-order conditions for optimality." The latter was completed under this grant, but much of the research that went into it was performed under the predecessor grant, AF-AFOSR-77-3204.

Other marginal value results are presented in paper [9], "Differentiability properties of the minimum value in an optimization problem depending on parameters." These too are based on subgradient analysis.

Quite a different area of nonlinear programming is the topic of [5], "Automatic step sizes for the descent algorithms in monotropic programming." The problems in question are linearly constrained but have objective functions that can be expressed as a sum of linear functions composed with convex functions of a single real variable. Piecewise linear or quadratic programming meets this prescription for instance. For problems of such type there are primal and dual methods of solution in which a direction of descent is determined by some pivoting routine and a line search is then carried out. In the case of the dual methods there is the complication that we would like to be able to follow the procedure in terms of the data as it is represented in primal form, but this is hard to do for the line search because of the number of function evaluations that may be involved. This article demonstrates that a certain automatic step size rule can be used in such cases to avoid line search entirely.

4. Optimal Control

Problems of optimal control have long been of interest to the writer, and they have provided much motivation for theoretical developments. They have also been the beneficiaries of those developments. The work on optimal control has not, however, conformed to the standard framework of the subject, which was put together with problems of mechanical engineering in mind. Rather this work has been aimed at problems of an economic character such as inventory control or the exploitation of natural resources.

A notable characteristic of such problems is the dependence of the control set at any given time on the state of the system at that time. The celebrated maximum principle of Pontriagin makes no allowance for such a possibility at all! Methods of convex analysis have previously been used by the writer to get around this lack, at least for problems of convex type, and F.H. Clarke has made progress with nonconvex problems.

An important question which arises in this context is that of properly extending the formulation of optimal control problems to allow for impulse controls. This is a question of merit on its own, but it also derives much weight from the duality between impulse controls and constraints on the states of a system. The multipliers for state constraints in the primal problem correspond to impulse controls in the dual problem, and vice versa.

In J. Murray's thesis [10], "On the proper extension of optimal control problems to admit impulses," the challenge is taken up in the light of existence theory. The point of view is the following. Impulse controls should make sense as idealized limits of ordinary controls. As such they should be obtainable from techniques of compactification that are designed to supply "solutions" to classes of problems that do not enjoy growth properties adequate to secure the existence of solutions (optimal trojectories) in the ordinary sense.

Murray succeeds in finding by a limit process the natural extension of an optimal control problem to the larger control space in which impulses can occur. He uncovers at the same time the fact that impulses can be not only in the simple form of jumps but also "distributed continuously in singular time." The possibility of the latter phenomena seems to have been overlooked by all those who worked previously on impulse controls, an observation which calls much of the existing literature into question.

It is hoped that the understanding provided by Murray's results will eventually make possible the incorporation of stochastic elements into control problems with state-dependent controls.

Papers published or written under this grant

1. R.T. Rockafellar, "Marginal values and second-order conditions for optimality, *Math. Programming* 26 (1983), 245-286.

2. R.T. Rockafellar and R.J.B. Wets, "Deterministic and stochastic optimization problems of Bolya type in discrete time," *Stochastics* 10 (1983), 273-312.

3. R.T. Rockafellar and R.J.B. Wets, "On the interchange of subdifferentiation and conditional expectation for convex functionals," *Stochastics* 7 (1982), 173-182.

4. R.T. Rockafellar, "Directional differentiability of the optimal value function in a nonlinear programming problem," *Math. Programming Studies* 21 (1984).

5. R.T. Rockafellar, "Automatic step sizes for the descent algorithms in monotropic programming," in *Mathematical Programming* (R. Cottle, M. Kelmanson, and B. Korte, eds.), Elsevier, 1984.

6. R.T. Rockafellar, "Generalized subgradients in mathematical programming," in *Mathematical Programming Bonn 1982: The State of the Art* (A. Bachem, M. Groetschel, and B. Korte, eds.), Springer-Verlag, 1983.

7. R.T. Rockafellar and R.J.B. Wets, "A dual solution procedure for quadratic stochastic programs with simple recourse," in *Numerical Methods* (V. Pereyra, A. Reinoza, eds.), Springer-Verlag Lecture Notes in Math. No. 1005, 1983, 252-265.

8. J. Treiman, "Characterization of Clarke's tangent and normal cones in finite and infinite dimensions," *Nonlinear Analysis, Theory, Methods, Appl.* 7 (1983), 771-783.

9. R.T. Rockafellar, "Differentiability properties of the minimum value in an optimization problem depending on parameters," *Proceedings of the International Congress of Mathematicians, Warsaw, 1983,* not yet in print.

10. J. Murray, "On the proper extension of optimal control problems to admit impulses," Ph.D. Thesis, University of Washington, 1983.

11. J. Treiman, "A new characterization of Clark's tangent cone and its applications to subgradient analysis and optimization," Ph.D. thesis, University of Washington, 1983.

12. R.T. Rockafellar, "Extensions of subgradient calculus with applications to optimization," *Nonlinear Analysis Theory, Methods, Appl,* submitted.

13 R.T. Rockafellar and R.J.B. Wets, "A Lagrangian finite generation technique for solving linear-quadratic problems in stochastic programming," in preparation.

END

FILMED

6-84

DTIC