GEORGIA INSTITUTE OF TECHNOLOGY
OFFICE OF CONTRACT ADMINISTRATION
SPONSORED PROJECT INITIATION

Date: __8/26/80__

Project Title: MATHEMATICAL PROGRAMING WITH AND WITHOUT DIFFERENTIABILITY

CO- Project No: E-24-614/G-37-624

CO- Project Director: DR. M. S. BAZARAA AND DR. J. E. SPINGARN

Sponsor: AIR FORCE OFFICE OF SCIENTIFIC RESEARCH, BOLLING AFB, D.C.   20332

Agreement Period: From __6/15/80__ Until __6/14/81__

Type Agreement: GRANT NO. AFOSR-80-0195

Amount:
$35,295 E-24-614
10,725 G-37-624
$46,020 TOTAL AFOSR
7,688 GIT (E-24-341)
$53,708 TOTAL

Reports Required: FINAL TECHNICAL REPORT

Sponsor Contact Person (s):

<table>
<tr><td>Technical Matters</td><td>Contractual Matters</td></tr>
<tr><td></td><td>(thru OCA)</td></tr>
<tr><td>DR. JOSEPH BRAM, PROGRAM MANAGER
DIRECTORATE OF MATHEMATICAL & INFORMATION
  SCIENCES
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
ATTN: AFOSR/NM
BUILDING 410
BOLLING AFB, D.C.   20332</td><td>2LT. RICKY BAILEY,BUYER
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
ATTN: AFOSR/PKZ
BUILDING 410
BOLLING AFB, D.C.   20332
(202) 767-4704</td></tr>
</table>

Defense Priority Rating: NONE

Assigned to: __ISYE/MATHEMATICS__ (School/Laboratory)

COPIES TO:

Project Director
Division Chief (EES)
School/Laboratory Director
Dean/Director—EES
Accounting Office
Procurement Office
Security Coordinator (OCA)
Reports Coordinator (OCA)

Library, Technical Reports Section
EES Information Office
EES Reports & Procedures
Project File (OCA)
Project Code (GTRI)
Other__OCA RESEARCH PROPERTY COORDINATOR__

Date: __7/8/81__

Project Title: Mathematical Programming With and
Without Differentiability

Co-Project No: E-24-614 / G-37-624

Co-Project Directors: Dr. M. S. Bazaraa & Dr. J.E. Spingarn

Sponsor: AIR FORCE OFFICE OF SCIENTIFIC RESEARCH, BOLLING AFB, D.C. 20332

Effective Termination Date: __6/14/81__

Clearance of Accounting Charges: __8/14/81 (Reporting)__

Grant/Contract Closeout Actions Remaining:

    ___ Final Invoice and Closing Documents

    _X_ Final Fiscal Report (SF 269)

    _x_ Final Report of Inventions

    _X_ Govt. Property Inventory & Related Certificate

    ___ Classified Material Certificate

    ___ Other _____

Continued by E-24-638 / G-37-633

Assigned to: __ISyE / Mathematics__ (School/̶L̶a̶b̶o̶r̶a̶t̶o̶r̶y̶)

COPIES TO:

Administrative Coordinator
Research Property Management
Accounting Office
Procurement Office
Research Security Services
✓Reports Coordinator (OCA)

Legal Services (OCA)
Library, Technical Reports
EES Research Public Relations (2)
Project File (OCA)
Other: _____

FINAL REPORT

MATHEMATICAL PROGRAMMING WITH AND WITHOUT DIFFERENTIABILITY

By

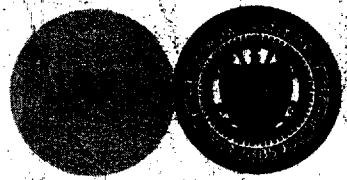M. S. Bazaraa, Principal Investigator
J. E. Spingarn

Submitted to

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Bolling Air Force Base, D. C. 20332

Under

Grant No. AFOSR–80–0195

June 1981

GEORGIA INSTITUTE OF TECHNOLOGY
SCHOOL OF INDUSTRIAL & SYSTEMS ENGINEERING
ATLANTA, GEORGIA 30332

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT. ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) MATHEMATICAL PROGRAMMING WITH AND WITHOUT DIFFERENTIABILITY		5. TYPE OF REPORT & PERIOD COVERED Final Report June 15, 1980-June 14, 1981
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) M.S. Bazaraa J.E. Spingarn		8. CONTRACT OR GRANT NUMBER(s) AFOSR-80-0195
9. PERFORMING ORGANIZATION NAME AND ADDRESS Georgia Institute of Technology Atlanta, Georgia 30332		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AirForce Office of Scientific Research Bolling AFB, D.C. 20332		12. REPORT DATE June 15, 1981
		13. NUMBER OF PAGES 22 pages
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Mathematical Programming, Quasi-Newton algorithms, Nonlinear programming, Generic optimality conditions, Nondifferentiable optimization, inexact line search, subdifferentials

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

An algorithm for solving equality constrained nonlinear programming is presented. The algorithm performs two major computations. First, the search vector is determined by solving a quadratic programming problem implicitly through a suitable projection. Second, a step size along the search vector is determined by extending the inexact line search procedure of Armijo in such a way to handle nondifferentiability of the descent function. Theorems showing global convergence of the proposed algorithm are given.

(Continued on reverse side)

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73

MATHEMATICAL PROGRAMMING

WITH AND WITHOUT DIFFERENTIABILITY

Final Report

Grant No. AFOSR-80-0195

June 1981

Submitted to

Air Force Office of Scientific Research
Bolling Air Force Base, D. C.   20332

by

GEORGIA INSTITUTE OF TECHNOLOGY
School of Industrial and Systems Engineering

Dr. M. S. Bazaraa
Principal Investigator

TABLE OF CONTENTS

I. QUASI-NEWTON ALGORITHMS FOR CONSTRAINED NONLINEAR PROGRAMMING (M. S. Bazaraa)

In this report, an algorithm for solving equality constrained nonlinear programming is presented. The algorithm performs two major computations. First, the search vector is determined by solving a quadratic programming problem implicitly through a suitable projection. Second, a step size along the search vector is determined by extending the inexact line search procedure of Armijo in such a way to handle nondifferentiability of the descent function. Theorems showing global convergence of the proposed algorithm are given without proof. The details will be provided in a forthcoming research manuscript.

The following research topics are currently under investigation:

1. Establishing local convergence properties of the proposed algorithm under suitable choices of the quadratic form that determines the search vector. A superlinear rate of convergence is anticipated in the case where the second order sufficiency conditions hold.

2. Generalizing both global and local convergence results from equality constrained problems to problems involving both equality and inequality constraints.

3. Computational testing of the proposed algorithms using test problems available from the literature of nonlinear programming.

I.2 Algorithm for Equality Constrained Problems

Consider the following nonlinear programming problem, where f and h_i are continuously differtiable functions on R^n.

P : Minimize   f(x)

    Subject to $h_i(x) = 0$ $i=1,....,\ell$

In this section, a description of the proposal algorithm is given. The search vector is computed by solving a quadratic programming problem implicitly in terms of a suitable projection matrix. An inexact line search is then performed to compute the step size.

I.2.1 Description of the Algorithm

Initialization

Choose r > 0, select a starting point x_1, let k=1, and go to the main step.

Main Step

1. Choose a positive definite matrix B_k and compute p_k and q_k as follows, where h denotes the constraint vector $h_1,....,h_\ell$ and ∇h denotes the $n \times \ell$ gradient matrix whose ith column is ∇h_i:

$$p_k = B_k^{-1}\{I - \nabla h(x_k)[\nabla h(x_k)^t B_k^{-1} \nabla h(x_k)]^{-1} \nabla h(x_k)^t B_k^{-1}\} \nabla h(x_k) \qquad (2.1)$$

$$q_k = B_k^{-1} \nabla h(x_k)[\nabla h(x_k)^t B_k^{-1} \nabla h(x_k)]^{-1} h(x_k) \qquad (2.2)$$

If $p_k = q_k = 0$, stop; x_k is a Kuhn-Tucker point. Otherwise let

$d_k = p_k + q_k$ and go to step 2.

2. Compute the smallest nonnegative integer γ satisfying:

$$\phi_r(x_k + (1/2)^\gamma d_k) - \phi_r(x_k) \leq (1/3)(1/2)^\gamma \grave{\phi_r}(x_k, d_k) \qquad (2.3)$$

where,

$$\phi_r(x) = f(x) + r\sum_{i=1}^{\ell} |h_i(x)| \qquad (2.4)$$

$$\grave{\phi_r}(x_k, d_k) = \nabla f(x_k)^t d_k - r\sum_{i=1}^{\ell} |h_i(x_k)| \qquad (2.5)$$

Let $x_{k+1} = x_k + (1/2)^\gamma d_k$, replace k by k+1, and go to step 1.

The algorithm can be interpreted as an exact penalty function method that attempts to minimize the single unconstrained penalty function $\phi_r(x) = f(x) + r\sum_{i=1}^{\ell} |h_i(x)|$, which under suitable conditions, results in a solution to Problem P. At a given point x_k, the search vector d_k is computed . If $d_k = 0$, the solution procedure is stopped with the conclusion that x_k is a Kuhn-Tucker point. Otherwise, d_k is a descent direction to the penalty function ϕ_r provided that $r \geq |v_i|$ for $i = 1,\ldots,\ell$, where v_i is the Lagrangian multiplier associated with the ith equality constraint. Armijo's inexact line search is then used to calculate a step size, and the process is repeated starting with the new point.

I.2.2 Computational Expedients

The calculation of the vector d_k can be simplified as follows:

$$d_k = -u_0 + U \zeta \qquad (2.6)$$

$$U = [u_1, u_2,\ldots,u_\ell] \qquad (2.7)$$

The vectors u_0, $u_1,\ldots,u_\ell$, and ζ are determined by solving the following systems

of linear equations:

$$B_k \ u_0 = \nabla f(x_k) \tag{2.8}$$

$$B_k \ u_i = \nabla h_i(x_k) \qquad i=1,\ldots,\ell \tag{2.9}$$

$$A \ \zeta = \nabla h(x_k)^t u_0 - h(x_k) \tag{2.10}$$

where,

$$A = \nabla h(x_u)^t \ U \tag{2.11}$$

In order to perform the above computations, the symmetric positive definite matrix B_k is decomposed into the form LL^t, where L is a lower triangular matrix with positive diagonal elements. Utilizing this factorization, the vectors $u_0,\ldots,u_\ell$ can be easily determined by forward and then backword substitution. Now, to compute ζ, the **matrix** A is determined. Here, the ijth element A_{ij} of A is given by $\nabla h_i(x_k)^t u_j$. Noting that A is symmetric, A_{ij} is computed for $i=1,\ldots,\ell$ and $j \geq i$. Furthermore, A is itself positive definite, then it can be factorized in the form $\hat{L}\hat{L}^t$. The vector ζ is calculated by solving the system $\hat{L}\hat{L}^t \ \zeta = \nabla h(x_k)^t u_0 - h(x_k)$. Now the search direction d_k is at hand.

I.3 Global Convergence Properties

In this section, two results involving global convergence of the algorithm are given. The first shows that each accumulation point is a Kuhn-Tucker solution to the original problem. The second result establishes convergence of the whole sequence if an accumulation point satisfies a suitable second order sufficiency condition.

Theorem 1

Consider the algorithm described in Section I.2.1. Let f and h_i for $i=1,\ldots,\ell$ be continuously differentiable. Suppose that the family of matrices $\{B_k\}$ is chosen to be uniformly positive definite. Furthermore, suppose that the sequence of generated points is contained in a compact set and that the penalty parameter r is such that $r \geq ||\zeta_k||_\infty$ for each k where ζ_k is the Lagrangian multiplier vector computed in (2.10). Then either the algorithm stops in a finite number of iterations with a Kuhn-Tucker point to problem P or else generate an infinite sequence $\{x_k\}$ of which any accumulation point is a Kuhn-Tucker point for Problem P.

Definition 1

A feasible solution $\bar{x}$ to Problem P is said to satisfy the second order sufficiency optimality conditions if $\nabla h(\bar{x})$ has full rank and if there exists a vector v such that:

1. $\nabla f(\bar{x}) + \nabla h(\bar{x})\, v = 0$

2. The Hessian of the Lagrangian function $f(x) + v^t h(x)$ with respect to x is positive definite on the tangent plane $\{y : \nabla h(\bar{x})^t y = 0\}$.

Theorem 2

Let the accumulation point in Theorem 1 satisfy the second order sufficiency optimality condition. Then the whole sequence $\{x_k\}$ converges to $\bar{x}$.

II. GENERIC OPTIMALITY CONDITIONS AND NONDIFFERENTIABLE OPTIMIZATION (J. E. Spingarn)

Monotone-type properties of the subdifferentials of nonconvex nondifferentiable functions were studied. The class of "lower-C^1" functions was characterized as that class of locally Lipschitz functions whose subdifferentials are "strictly submonotone." The proximal point algorithm for solving equations of the form $0 \varepsilon T(x)$ with T maximal monotone, was extended to "maximal strictly hypomonotone" mappings. This was shown to lead to a "proximal minimization" method for lower-C^2 functions.

Finite-dimensional variational problems were studied from a generic point of view. It was shown that for most problems in a given class, every solution is "strong" in a certain sense. The results were applied to a family of convex programming problems in order to obtain second-order conditions that are necessary for optimality for almost all problems in the family.

II.2 <u>Research Summary</u>

Our research in nondifferentiable optimization has shown that concepts related to "monotonicity" of a multifunction play a natural role in the analysis of nondifferentiable functions. Further, we have shown that the "proximal point algorithm", which makes use of monotone mappings to solve convex programming problems, can be generalized to mappings that are not monotone, and hence can be applied to nondifferentiable optimization problems.

Recall that a multifunction $T : R^n \overset{\rightarrow}{\rightarrow} R^n$ is <u>monotone</u> provided that $\langle x_1 - x_2, y_1 - y_2 \rangle \geq 0$ whenever $y_1 \in T(x_1)$ and $y_2 \in T(x_2)$. The <u>graph</u> of T is the set $G(T) = \{(x, y) \in R^n \times R^n: y \in T(x)\}$. T is <u>maximal monotone</u> if $G(T)$ is not properly contained in the graph of another monotone mapping. If T is maximal monotone, the <u>proximal point algorithm</u> provides a method for finding $x \in R^n$ with the property that $0 \in T(x)$.

The principal reason why monotonicity plays an important role in convex programming is that the subdifferential of a convex function is a maximal monotone mapping. That is if $f: R^n \rightarrow R$ is convex, then the mapping $\partial f: R^n \overset{\rightarrow}{\rightarrow} R^n$ defined by

$$\partial f(x) = \{y: f(x') \geq f(x) + \langle x' - x, y \rangle \ \forall x' \in R^n\}$$

is maximal monotone. Since a point x which satisfies $0 \in \partial f(x)$ is a global minimizer for f, the proximal point algorithm provides a method for minimizing convex functions. Maximal monotone mappings also arise in a natural way from saddle functions. A saddle function is a function $L(x, y)$ with the property that L is convex in x for fixed y and concave in y for fixed x. For such functions, the mapping $T(x, y) = \partial_x L(x, y) \times - \partial_y L(x, y)$ is maximal monotone. Since a point (x, y) with $0 \in T(x, y)$ is a saddle point for L, the proximal point algorithm thus provides a method for finding saddle points for convex-concave functions. Rockafellar has demonstrated the close tie between this

method and the important "method of multipliers" for solving constrained convex programming problems.

Using the convex (monotone) case as a model, we investigated the class of "lower-C^1" functions and characterized it in terms of properties of the subdifferential mapping.   This investigation was made possible by the recent introduction by Rockafellar and Clarke of the generalized gradient of a nondifferentiable nonconvex function.   Clarke and Rockafellar have shown that the notion of the subdifferential of a convex function has a natural extension to much broader classes of functions.   Clarke carried out this program for lower semicontinuous functions, and Rockafellar extended the theory to arbitrary extended real-valued functions.   In the locally Lipschitz case, which is the only case which concerns us here, the subdifferential of $f: R^n \to R$ is defined to be the set-valued mapping $\partial f: R^n \to R^n$ obtained by taking $\partial f(x)$ to be the convex hull of the set of all limit points of convergent sequences of the form $\nabla f(x_n)$, where $x_n \to x$ and f is differentiable at x_n.   This definition generalizes the definition given for the convex case.   It is not hard to show that for a locally Lipschitz function $f: R^n \to R$, f is convex if, and only if, ∂f is monotone.

A function $f: R^n \to R$ is lower-C^1 if for every $\bar{x} \in R^n$ there is a neighborhood U of $\bar{x}$, a compact set S, and a function $g: U \times S \to R$ such that

     (i) $f(x) = \max \{g(x, s) : s \in S\}$ for all $x \in U$

     (ii) g is continuous on $U \times S$

     (iii) g_x is continuous on $U \times S$

Because this class of functions arises through the simple operation of taking a maximum over a compact set, it is clear that this class of functions is of interest in optimization theory; it is precisely through the operation of taking maxima that nondifferentiable functions are most often encountered.

Hence, it is highly desirable to have a characteristization of such functions in terms of their subdifferential mappings. We have shown that a locally Lipschitz function f is lower-C^1 if, and only if, ∂f is "strictly submonotone". ∂f is strictly submonotone iff

$$\liminf_{\substack{x_i \to \bar{x} \\ y_i \in \partial f(x_i)}} \frac{\langle x_1 - x_2, y_1 - y_2 \rangle}{|x_1 - x_2|} \geq 0$$

for every $\bar{x}$. This result is proven in [4], where we also present several other properties of lower-C^1 functions and relate these to properties that have been studied by other researchers in the area. Another property of lower-C^1 functions, which we showed to be equivalent to f being lower C^1 when f is locally Lipschitz, is the following uniform lower differentiability property: f is lower-C^1 if, and only if, for any $x \in R^n$, $y \in R^n$, and $\varepsilon > 0$, there exists $\delta > 0$ such that

$$f(x' + ty) \geq f(x') + t \, \psi^*_{\partial f(x')} (y) - \varepsilon t$$

whenever $|x' - x| < \delta$ and $0 < t < \delta$,

where $\psi^*_{\partial f(x')} (\cdot)$ denotes the support function of $\partial f(x')$.

Rockafellar has recently obtained some related results. He characterized the subdifferentials of "lower-C^2" functions, showing that f is lower-C^2 iff ∂f is "strictly hypomonotone".

This class of lower-C^2 functions, as well as the strict hypomontonicity property which characterizes their subdifferentials, has played a central role in our recent research. We define $f: R^n \to R \cup \{\infty\}$ to be lower-C^2 if $f = g - h$ for some lower semicontinuous function $g: R^n \to R \cup \{\infty\}$ and some C^2 function h. T is strictly hypomontone provided for every bounded set $K \subset R^n$ there is $k \geq 0$ such that $T + kI$ is monotone on K. T is maximal strictly hypomonotone pro-

vided T is strictly hypomonotone and the graph of T is not properly contained in the graph of another strictly hypomonotone mapping.

For maximal strictly hypomonotone mappings T, we have developed a locally and linearly convergent algorithm for solving equations of the form $0 \in T(x)$. These results will appear in [1]. In particular, taking $T = \partial f$, the algorithm can be used to minimize lower-C^2 functions. The algorithm is an extension of the proximal point method which solves $0 \in T(x)$ in the case where T is a maximal monotone mapping.

In the maximal monotone case, the known proximal point algorithm works because of the fact that for any $c > 0$, the proximal mapping $P(x) = (I+cT)^{-1}(x)$ is single-valued and nonexpansive. Starting from an initial point $x \in R^n$, the algorithm generates a sequence by the rule $x_{k+1} = P(x_k)$.

To extend the algorithm to cover maximal strictly hypomonotone mappings, several obstacles had to be overcome. First of all, in the hypomonotone case, the proximal mapping need no longer be single-valued. This difficulty was over-come by demonstrating the possibility of modifying the proximal mapping to make it single-valued in the vicinity of a solution, so long as the constant c is not too large. Another problem is that the basic algorithm, if defined as in the monotone case, need not converge even locally, unless a strong regularity assumption is satisfied at the solution. The regularity condition which we require to establish the local linear convergence of our algorithm asserts the differentiability of T^{-1} at 0 and the monotonicity (positive semidefin-iteness) of that derivative. This is a very strong assumption, and perhaps one which might seem unnatural. It would, after all, be pointless to esta-blish convergence of an algorithm under hypotheses that are so strong that they cannot be expected to hold. For this reason, it is fortunate that we were able to obtain results establishing the generic necessity of the required regularity condition. In other words, we showed that for "most"

problems (in a certain rigorous sense), the required regularity condition is in fact satisfied at all solutions.

The generic necessity of the regularity hypothesis was established using a result of Mignot stating that a maximal monotone mapping possesses a derivative almost everywhere. This is in contrast to our previous work on generic conditions in (differentiable) nonlinear programming [3, 5] where the principal tools for proving genericity came from differential topology; e.g., Sard's Theorem and transversity.

In addition to our work on nondifferentiable optimization, we have conducted research on the generic properties of variational problems. The results of this work have been written up in [2]. If $T: R^n \overset{\rightarrow}{\to} R^n$ is a maximal monotone mapping and $C \subset R^n$ is a convex set, the associated <u>variational problem</u> is to find $z \in R^n$ such that

$$0 \in T(z) + N_C(z),$$

where $N_C(z)$ denotes the normal cone to C at z. We showed that if T and C satisfy a certain Lipschitz condition, then generically, all solutions z are "strong" solutions in the sense that

$$0 \in T(z) + \text{relint } N_C(z).$$

The condition we imposed on C is that the normal cone mapping $N_C(\cdot)$ satisfy the following Lipschitz property:

for all $z \in C$, there exists, $\mu_z \geq 0$ and a neighborhood U_z of z such that $\text{dist}(y', N_C(z)) \leq \mu_z |z' - z| |y'|$ whenever $z' \in U_z$, $y' \in N_C(z')$.

This condition is satisfied for example, if C is polyhedral (in which case we can always take $\mu_z = 0$), or a manifold-with-corners.

We were able to proved the following generic result:

THEOREM.   If C is a closed convex set satisfying the above Lipschitz condition and T is maximal monotone and locally Lipschitz on an open set containing C, then except for $\bar{w} \in R^n$ belonging to a set of measure zero, every $\bar{z}$ satisfying

$$\bar{w} \in T(\bar{z}) + N_C(\bar{z})$$

also satisfies

$$\bar{w} \in T(\bar{z}) + \text{relint } N_C(\bar{z}).$$

We were also able to prove a further result which has consequences regarding the generic stability of solutions to a variational problems:

THEOREM.   Let C and T be as in the previous theorem. Except for $\bar{w} \in R^n$ in a set of measure zero, $\bar{w} \in T(\bar{z}) + N_C(\bar{z})$ implies that $S^{-1} = (T+N_C)^{-1}$ is differentiable at $\bar{w}$ and the derivative A satisfies kernel $(A) = L_C(\bar{z})^{\perp}$ and range $(A) = L_C(\bar{z})$, where $L_C(x) = \{y \in R^n: y \cdot y' = 0, y' \in N_C(x)\}$

These results have interesting consequences for the family

(P_{vu}) minimize $f(x) - x \cdot v$ subject to $g_i(x) \le u_i$,

   $i = 1,\ldots,m,$ and $x \in D$

of convex programming problems indexed by $(v, u) \in R^n \times R^m$. Here, the functions $f, g_1,\ldots,g_m$ are finite-valued continuously differentiable convex functions whose derivatives locally satisfy a Lipschitz property, and D is a closed convex set satisfying the Lipschitz property. If we define

$$L(x,y) = f(x) + \sum_{i=1}^{m} y_i g_i(x),$$

we prove the following:

THEOREM.   Suppose that D is a polyhedral convex set. Except for $(\bar{v},\bar{u})$ belonging to a set of measure zero, we can make the following assertion. If $\bar{x}$ is a solution to (P_{vu}), if F is the unique face of D such that $\bar{x} \in \text{relint } F$, and if $\bar{y} \in R_+^m$ satisfies the Kuhn-Tucker conditions (for almost all $(\bar{v},\bar{u})$, such $\bar{y}$ will in fact exist), then

$\{\nabla(g_i|F)(\bar{x}) : i \ \varepsilon \ I(\bar{x})\}$ is linearly independent

$H = \nabla_x^2 L(\bar{x},\bar{y})$ satisfies $\zeta'H\zeta > 0$ whenever

$0 \neq \zeta \ \varepsilon \ <F>$ and $\zeta \cdot \nabla g_i(\bar{x}) = 0 \quad i \ \varepsilon \ I(\bar{x})$

where $I(\bar{x}) = \{i : g_i(\bar{x}) = \bar{u}_i\};$ $\nabla(g_i|F)(\bar{x})$ denotes the gradient at $\bar{x}$ of the restriction of g_i to F, or equivalently, the projection of $\nabla g_i(\bar{x})$ onto $<F>$; and $\nabla_x^2 L(\bar{x},\bar{y})$ is the Hessian of L at $(\bar{x},\bar{y})$ with respect to x.

II.3 PUBLICATIONS (SPINGARN)

1. "Submonotone mappings and the proximal point algorithm," to appear in Numerical Functional Analysis and Optimization.

2. "On generic properties of variational problems," submitted to Game Theory and Mathematical Economics, Ed. O. Pallaschke.

3. "On optimality conditions for structured families of nonlinear programming problems," to appear in Mathematical Programming.

4. "Submonotone subdifferentials of Lipschitz functions," Transactions of the American Mathematical Society, 264(1981) 77-89.

5. "Second-order conditions that are necessary with probability one," Mathematical Programming with Data Perturbations, Editor: Anthony V. Viacco, Lecture Notes in Pure and Applied Mathematics, Marcel Dekker, Inc., scheduled publication date: 1981.

ON GENERIC PROPERTIES OF VARIATIONAL PROBLEMS

J. E. Spingarn

School of Mathematics
Georgia Institute of Technology
Atlanta, Georgia
U.S.A.

Finite-dimensional variational problems are studied from a
generic point of view. It is shown that for most problems
in a given class, every solution is "strong" in a certain
sense. The results are applied to a family of convex pro-
gramming problems in order to obtain second-order conditions
that are necessary for optimality for almost all problems in
the family.

INTRODUCTION

In [6], [7], and [8], we investigated optimality conditions in nonlinear pro-
gramming from the generic point of view. From this viewpoint, the important
objects for study are not individual problems, but rather families of problems;
one makes assertions about "typical" problems in a given family of problems. In
these previous investigations, our principal tool was Sard's Theorem and its
generalizations.

In this paper, we obtain similar results, but restrict ourselves to convex pro-
gramming. The convex case is simpler, but illuminating. We will show that for
convex programming problems, results similar to those of [6], [7], and [8] may be
obtained without using Sard's Theorem. Instead, we rely on a result due to
Mignot [1] concerning the differentiability of maximal monotone mappings. Using
this result, some generic assertions about variational problems are established
which are then applied to convex programming.

The relationship between variational problems and convex programming which we
exploit here is well known. For more insight into this relationship, we refer
the reader to [2] and [5].

PRELIMINARIES

A *multifunction* $T : R^n \rightarrow R^n$ is a set-valued mapping. T is *differentiable* at x
if $T(x)$ is single-valued and there is a linear mapping A such that for every
$\epsilon > 0$, $\emptyset \neq T(x+h) \subset T(x) + A(h) + \epsilon|h|B$ for all h in a neighborhood of 0,
where $B = \{z : |z| \leq 1\}$. The *inverse* of T is the multifunction defined by
$T^{-1}(y) = \{x : y \in T(x)\}$. If $C \subset R^n$ is a closed convex set, $x \in C$, we define
$$N_C(x) = \{y \in R^n : y \cdot (x'-x) \leq 0, \forall x' \in C\}$$
$$L_C(x) = \{y \in R^n : y \cdot y' = 0, \forall y' \in N_C(x)\}.$$
The *relative interior* of C, relint C, is the interior of C relative to the
smallest affine flat containing C. The *distance* from x to C is denoted by
$dist(x,C)$. A multifunction $T : R^n \rightarrow R^n$ is *monotone* if $<x-x',y-y'> \geq 0$ when-
ever $y \in T(x)$, $y' \in T(x')$. T is *maximal monotone* if T cannot be properly

extended to a monotone mapping.

VARIATIONAL PROBLEMS

The *variational problem* associated with the maximal monotone mapping $\tau : R^n \to R^n$ and the closed convex set $C \subset R^n$ is to find $z \in R^n$ such that

(1) $0 \in \tau(z) + N_C(z).$

Theorem 1 shows that if τ and C satisfy a certain Lipschitz condition then, generically, all solutions to (1) are "strong" solutions in the sense that "$N_C(z)$" can be replaced with "relint $N_C(z)$". Theorem 2 gives further first-order information about typical solutions to (1).

The condition to be imposed on C is that the normal cone mapping $N_C(\cdot)$ satisfy the following Lipschitz property:

(2)   for all $z \in C$, there exists $\mu_z \geq 0$ and a neigh-
            borhood U_z of z such that $\text{dist}(y', N_C(z)) \leq \mu_z |z'-z||y'|$
            whenever $z' \in U_z$, $y' \in N_C(z')$.

This condition is satisfied, for example, if C is polyhedral (in which case we can always take $\mu_z = 0$), a manifold-with-corners, or more generally a "cyrtohedron" (this refers to a class of piecewise smooth sets defined in Spingarn [6]).

THEOREM 1.   *If C is a closed convex set satisfying (2), and τ is maximal monotone and locally Lipschitz on an open set containing C, then except for $\bar{w} \in R^n$ belonging to a set of measure zero, every $\bar{z}$ satisfying*

(3a) $\bar{w} \in \tau(\bar{z}) + N_C(\bar{z})$

also satisfies

(3b) $\bar{w} \in \tau(\bar{z}) + \text{relint } N_C(\bar{z}).$

The proof of Theorem 1 hinges on

LEMMA 1.(Mignot [1]).   *A maximal monotone mapping $T : R^n \to R^n$ is either empty-valued or differentiable at all points z, except possibly for z belonging to a set of measure zero in R^n.*

Proof of Theorem 1:   By Rockafellar [4, Theorem 2], the multifunction $T = \tau + N_C$ is maximal monotone. (In general, if T_1 and T_2 are maximal monotone mappings on R^n such that relint $\text{dom}(T_1) \cap \text{relint dom}(T_2) \neq \emptyset$, then $T_1 + T_2$ is also maximal monotone). By the symmetry in the definition, it follows that T^{-1} is also maximal monotone. By the Lemma, there is a measure zero set $Q \subset R^n$ such that $w \notin Q$ implies that either $T^{-1}(w) = \emptyset$ or T^{-1} is differentiable at w.

Suppose $\bar{w} \notin Q$ and (3a) holds. In order to produce a contradiction, let us assume that (3b) does not hold. For simplicity, we assume $\bar{w} = 0$. T^{-1} is differentiable at 0 since $0 \notin Q$ ($T^{-1}(0) \neq \emptyset$ because $0 \in T(\bar{z})$). Choose $\bar{y} \in \tau(\bar{z}) + \text{relint } N_C(\bar{z})$. Then $t\bar{y} \in \tau(\bar{z}) + N_C(\bar{z})$ for $0 \leq t \leq 1$ by convexity.

If A is the derivative of T^{-1} at 0 then for any $\varepsilon > 0$,

$$(4) \qquad \emptyset \neq T^{-1}(t\bar{y}) \subset \bar{z} + tA(\bar{y}) + \varepsilon tB$$

for all t sufficiently small. But $\bar{z} \in T^{-1}(t\bar{y})$ for $0 \leq t \leq 1$, so $0 \in tA(\bar{y}) + \varepsilon tB$. Cancelling t, and letting ε approach zero, we see that $A(\bar{y}) = 0$.

Now fix $\varepsilon > 0$. For each t, choose $z_t \in T^{-1}(t\bar{y})$. By (4), we have $|z_t - \bar{z}| \leq \varepsilon |t|$ for small t. It follows by (2) that
$\operatorname{dist}(t\bar{y} - \tau(z_t), N_C(\bar{z})) \leq \mu_{\bar{z}} |z_t - \bar{z}| \, |t\bar{y} - \tau(z_t)| \leq \mu_{\bar{z}} \varepsilon |t| \, |t\bar{y} - \tau(z_t)| \leq \mu_{\bar{z}} \varepsilon M t$,
where $M \geq 0$ is such that $|t\bar{y} - \tau(z_t)| \leq M$ for t near 0. Since $\bar{y} - \tau(\bar{z}) \in \operatorname{relint} N_C(\bar{z})$ and $-\tau(\bar{z}) \in N_C(\bar{z}) \backslash \operatorname{relint} N_C(\bar{z})$, there is $\nu > 0$ such that $\operatorname{dist}(t\bar{y} - \tau(\bar{z}), N_C(\bar{z})) \geq \nu |t|$ for $t < 0$ sufficiently small. If κ is a Lipschitz constant for τ near $\bar{z}$ then
$\nu |t| \leq \operatorname{dist}(t\bar{y} - \tau(\bar{z}), N_C(\bar{z})) \leq \operatorname{dist}(t\bar{y} - \tau(z_t), N_C(\bar{z})) + |\tau(z_t) - \tau(\bar{z})| \leq$
$\mu_{\bar{z}} \varepsilon M |t| + \kappa \varepsilon |t|$ and so $\nu \leq (\mu_{\bar{z}} M + \kappa) \varepsilon$. Since ε may be chosen arbitrarily small and $\nu > 0$, this is a contradiction. $\square$

THEOREM 2. *Let C and τ be as in Theorem 1. Except for $\bar{w} \in R^n$ in a set of measure zero, (3a) implies that $T^{-1} = (\tau + N_C)^{-1}$ is differentiable at $\bar{w}$ and the derivative A satisfies*

$$\operatorname{kernel}(A) = L_C(\bar{z})^\perp \quad \text{and} \quad \operatorname{range}(A) = L_C(\bar{z}).$$

Proof: Fix $\bar{w} \notin Q$, where Q is the measure zero set defined in the proof of Theorem 1, and assume (3a). From the previous proof, we know T^{-1} is differentiable at $\bar{w}$, $\bar{z} = T^{-1}(\bar{w})$, and (3b) holds. Let A be the derivative of T^{-1} at $\bar{w}$. Again, for simplicity we take $\bar{w} = 0$.

For any $w \in L_C(\bar{z})^\perp$ sufficiently small, it follows from (3b) that $\bar{z} \in T^{-1}(w)$, which clearly implies that $L_C(\bar{z})^\perp \subset \operatorname{kernel}(A)$. Fix $u \in L_C(\bar{z})$, $u \neq 0$. We will show that $Au \neq 0$. Suppose instead that $Au = 0$. Then for any $\varepsilon > 0$, for all t sufficiently small, $\emptyset \neq T^{-1}(tu) \subset \bar{z} + \varepsilon tB$. For each such t, choose $z_t \in T^{-1}(tu)$. Also, choose $y_t \in N_C(z_t)$ such that $tu = \tau(z_t) + y_t$. By (2), we have $\operatorname{dist}(y_t, N_C(\bar{z})) \leq \mu_{\bar{z}} |z_t - \bar{z}| \, |y_t| \leq \mu_{\bar{z}} \varepsilon t |y_t|$ for t sufficiently small.

Let π denote orthogonal projection onto $L_C(\bar{z})$. Then
$$
\begin{aligned}
|\tau(z_t) - \tau(\bar{z})| &= |tu - y_t - \tau(\bar{z})| \\
&\geq |\pi(tu - y_t - \tau(\bar{z}))| \\
&= |tu - \pi(y_t)| \\
&\geq t|u| - |\pi(y_t)|.
\end{aligned}
$$

Now,
$$
\begin{aligned}
|\pi(y_t)| &\leq \operatorname{dist}(y_t, N_C(\bar{z})) \leq \mu_{\bar{z}} \varepsilon t |y_t| \\
&\leq \mu_{\bar{z}} \varepsilon t (|tu| + |\tau(z_t)|) \leq \mu_{\bar{z}} \varepsilon t M
\end{aligned}
$$
where M is some bound for the quantity in parentheses, for t in some neighborhood of 0. Hence, $|\tau(z_t) - \tau(\bar{z})| \geq t(|u| - \mu_{\bar{z}} \varepsilon M)$. This shows that
for any $\varepsilon > 0$

$$\frac{|\tau(z_t) - \tau(\bar{z})|}{|z_t - \bar{z}|} \geq \frac{|u| - \mu_{\bar{z}}\epsilon M}{\epsilon}$$

holds for all t sufficiently near zero. Since ϵ can be chosen arbitrarily small, this contradicts the Lipschitz property of τ. Thus no such u exists, and we must have $L_C(\bar{z})^\perp = \mathrm{kernel}(A)$.

Suppose next that $Au \notin L_C(\bar{z})$ for some u. Then $v'Au < -u'Au$ for some $0 \neq v \in L_C(\bar{z})^\perp$, and hence $(v+u)'A(v+u) = (v+u)'Au = v'Au+u'Au < 0$. However, A, being the derivative at $\bar{w}$ of the monotone mapping T^{-1} must be positive semidefinite. That is, we must have $(v+u)'A(v+u) \geq 0$. This contradiction shows that $\mathrm{range}(A) \subset L_C(\bar{z})$. By a dimension argument, it is clear that we must actually have $\mathrm{range}(A) = L_C(\bar{z})$. $\square$

CONSEQUENCES IN CONVEX PROGRAMMING

Consider the parametrized family

(P_{vu}) minimize $f(x) - x \cdot v$ subject to $g_i(x) \leq u_i$,

                     $i = 1,\ldots,m$, and $x \in D$

of convex programming problems indexed by $(v,u) \in R^n \times R^m$. Here, the functions $f, g_1,\ldots,g_m$ are finite-valued continuously differentiable convex functions whose derivatives locally satisfy a Lipschitz property, and D is a closed convex set satisfying (2). Affine equality constraints could also be handled, but are omitted for brevity.

Let us define

$$L(x,y) = f(x) + \sum_{i=1}^{m} y_i g_i(x),$$
$$C = D \times R_+^m$$
$$\tau(x,y) = \left(\nabla f(x) + \sum_{i=1}^{m} y_i \nabla g_i(x), \ldots,-g_i(x),\ldots\right).$$

Since D satisfies (2), so does C. Also, τ is maximal monotone and locally Lipschitz, so that the assumptions for Theorem 1 are satisfied. Hence, except for $w = (v,u) \in R^n \times R^m$ belonging to a set of measure zero, every $z = (x,y)$ satisfying (3a) also satisfies (3b). The significance of this fact for the family (P_{vu}) can be seen from the following

PROPOSITION 1.   *(3a) holds iff* $z = (x,y)$ *satisfies the Kuhn-Tucker conditions for* (P_{vu}), *namely*

(5a) $\nabla f(x) - v + \sum_{i=1}^{m} y_i \nabla g_i(x) \in N_D(x)$,

           $g_i(x) \leq u_i$, $y_i \geq 0$, $y_i g_i(x) = 0$, $i=1,\ldots,m$,

           and $x \in D$

while (3b) holds iff

(5b) $\nabla f(x) - v + \sum_{i=1}^{m} y_i \nabla g_i(x) \in \mathrm{relint}\, N_D(x)$,

           $g_i(x) \leq u_i$, $y_i \geq 0$, $i=1,\ldots,m$,

           $y_i > 0$ iff $g_i(x) < u_i$, $i=1,\ldots,m$, and $x \in D$.

Proof: Similar to the proof of ([6], Lemma 3.1). □

Thus Theorem 1 implies that for "most" problems (P_{vu}): if x is a minimizer, and *if* there exists some $y \geq 0$ such that (x,y) satisfies the Kuhn-Tucker conditions (5a), then (x,y) also satisfies the strengthened conditions (5b). The following shows that the existence of such a y is also guaranteed for most problems

PROPOSITION 2.   *Except for (v,u) belonging to a set of measure zero, (P_{vu}) has the property that if x is a solution then there exists y such that (x,y) satisfies the Kuhn-Tucker conditions (5a).*

Proof: Let K denote the set of $u \in R^m$ such that $\{x \in D: g_1(x) \leq u_1, \ldots, g_m(x) \leq u_m\}$ is nonempty. For any $u \in \text{int } K$, the Slater condition holds for (P_{vu}). By ([3], Theorem 28.2), this implies that a Kuhn-Tucker vector exists for (P_{vu}) so long as the infimum in (P_{vu}) is not $-\infty$. Thus if $u \in \text{int } K$, $v \in R^n$, and x solves (P_{vu}), then (x,y) satisfies (5a). Since the boundary of K is of measure zero, the proof is complete. □

According to Theorem 2, the mapping $T^{-1} = (\tau + N_C)^{-1}$ is differentiable at almost all $w = (v,u) \in R^n \times R^m$. By Proposition 1, $T^{-1}(v,u)$ is the set of all pairs (x,y) satisfying the Kuhn-Tucker conditions (5) for the problem (F_{vu}). The differentiability of T^{-1} at (v,u) can thus be translated into an assertion about the sensitivity of (x,y) to changes in (v,u).

The assertions in Theorem 2 concerning the properties of the derivative of T^{-1} are more difficult to relate to the family (P_{vu}). We will do this only for the case where D is polyhedral convex. In this case, it is clear that $C = D \times R^m_+$ is also polyhedral convex. Suppose that $\bar{w} = (\bar{v},\bar{u})$ and $\bar{z} = (\bar{x},\bar{y})$ satisfy the conclusions of Theorem 2. In other words, assume that $T^{-1} = (\tau + N_C)^{-1}$ is differentiable at $\bar{w}$, $\bar{z} = T^{-1}(\bar{w})$, and the derivative A satisfies

(6a) $\text{kernel}(A) = L_C(\bar{z})^\perp$

(6b) $\text{range}(A) = L_C(\bar{z})$

and also (by the proof of Theorem 1),

(6c) $\bar{w} \in \tau(\bar{z}) + \text{relint } N_C(\bar{z})$.

The point z lies in the relative interior of a unique face G of C. Since for any $z \in \text{relint } G$, $L_C(z)$ is simply the subspace of $R^n \times R^m$ parallel to G, it makes sense to introduce the notation $L_C(z) = \langle G \rangle$. For $z \in \text{relint } G$, $N_C(z)$ is also independent of z, so it makes sense to introduce the notation $N_C(z) = N_C(G)$. Note that $\text{relint } N_C(G)$ is the interior of $N_C(G)$ relative to $\langle G \rangle^\perp$. Let π denote orthogonal projection onto $\langle G \rangle$.

We will show next that (6) implies that the function $\pi \circ \tau : G \to \langle G \rangle$ has a nonsingular derivative at $\bar{z}$. For simplicity, consider only the case $\bar{w} = 0$. Since $\pi(\bar{w}) = \pi(\tau(\bar{z}))$, we have $\pi(\tau(\bar{z})) - \tau(\bar{z}) \in \text{relint } N_C(G)$. For $z \in G$ sufficiently close to $\bar{z}$, since τ is continuous and by (6c), we have $\pi(\tau(z)) - \tau(z) \in N_C(G)$. Equivalently, $z \in T^{-1}(\pi(\tau(z)))$ for all $z \in G$ suffi-

ciently near $\bar{z}$. On the other hand, if $z \in T^{-1}(u)$ with $u \in \langle G \rangle$ and $z \in G$ near $\bar{z}$, then clearly $u = \pi(\tau(z))$. This demonstrates that for $u \in \langle G \rangle$ and $z \in G$ in a neighborhood of $\bar{z}$, $z \in T^{-1}(u)$ if and only if $z \in (\pi \circ \tau)^{-1}(u)$. But we know that T^{-1} is differentiable at $\bar{w} = 0$, with the restriction of the derivative to $\langle G \rangle$ being a nonsingular mapping : $\langle G \rangle \to \langle G \rangle$ by (6a) and (6b). Hence $(\pi \circ \tau)^{-1}$ considered as a map : $\langle G \rangle \to G$ is differentiable at 0, with nonsingular derivative. It is an easy consequence of this that $\pi \circ \tau$: $G \to \langle G \rangle$ is also differentiable at $\bar{z}$, with a nonsingular derivative. It is now not difficult to prove the following consequence of Theorem 2:

THEOREM 3. *Suppose that* C *is a polyhedral convex set. Except for* $(\bar{v}, \bar{u})$ *belonging to a set of measure zero, we can make the following assertion. If* $\bar{x}$ *is a solution to* (P_{vu}), *if* F *is the unique face of* D *such that* $\bar{x} \in \mathrm{relint}\, F$, *and if* $\bar{y} \in R_+^m$ *satisfies (5b) (for almost all* $(\bar{v}, \bar{u})$, *such* $\bar{y}$ *will in fact exist by Propositions 1 and 2 and Theorem 1), then*

(7a) $\{\nabla(g_i|F)(\bar{x}) : i \in I(\bar{x})\}$ is linearly independent

(7b) $H = \nabla_x^2 L(\bar{x}, \bar{y})$ satisfies $\zeta' H \zeta > 0$ whenever

$0 \neq \zeta \in \langle F \rangle$ and $\zeta \cdot \nabla g_i(\bar{x}) = 0 \ \forall i \in I(\bar{x})$

where $I(\bar{x}) = \{i : g_i(\bar{x}) = \bar{u}_i\}$; $\nabla(g_i|F)(\bar{x})$ *denotes the gradient at* $\bar{x}$ *of the restriction of* g_i *to* F, *or equivalently, the projection of* $\nabla g_i(\bar{x})$ *onto* $\langle F \rangle$; *and* $\nabla_x^2 L(\bar{x}, \bar{y})$ *is the Hessian of* L *at* $(\bar{x}, \bar{y})$ *with respect to* x.

Proof: If we let $E = \{y \in R_+^m : y_i > 0$ iff $i \in I(\bar{x})\}$, then $G = F \times E$ is the unique face of $C = D \times R_+^m$ that contains $\bar{z} = (\bar{x}, \bar{y})$. We have already seen that $\pi \circ \tau$, restricted to G, has a nonsingular derivative at $\bar{z}$. The fact that (7a) and (7b) hold is a direct consequence of the nonsingularity of this derivative. The argument, being identical to one found in the proof of ([8], Theorem 2), is omitted. $\square$

REFERENCES

[1]   Mignot, F., Un théorème de differentiabilité; application aux projections et au controle dans les inequations, Seminaire sur les Equations aux Derivées Partielles, 1973-1974, Université-Paris VI.

[2]   Robinson, S. M., Generalized equations and their solutions, part I: basic theory, Mathematical Programming Study 10 (1979) 128-141.

[3]   Rockafellar, R. T., Convex Analysis (Princeton University Press, Princeton, N.J., 1970).

[4]   _____, On the maximality of sums of nonlinear monotone operators, Trans. Amer. Math. Soc., 149 (1970) 75-88.

[5]   _____, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, Math. of O. R., 1 (1976) 97-118.

[6]   Spingarn, J. E., Fixed and variable constraints in sensitivity analysis, SIAM J. Control and Optimization 18 (1980) 297-310.

[7] _____, On optimality conditions for structured families of nonlinear programming problems, Mathematical Programming, to appear.

[8] _____ and Rockafellar, R. T., The generic nature of optimality conditions in nonlinear programming 4(1979) 425-430.