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 Directors: 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. AFSR-80-0195

$35,295 E-24-614

10,725 G-37-624

Amount: $46,020 TOTAL AFOSR

7,688 GIT (E-24-341)

$53,708 TOTAL

Reports Required: FINAL TECHNICAL REPORT

Sponsor Contact Person(s):

Technical Matters

DR. JOSEPH BRAM, PROGRAM MANAGER
DIRECTORATE OF MATHEMATICAL & INFORMATION SCIENCES
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
ATTN: AFSR/NM
BUILDING 410
BOLLING AFB, D.C. 20332

Contractual Matters

(thru OCA)

2LT. RICKY BAILEY, BUYER
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
ATTN: AFSR/PK2
BUILDING 410
BOLLING AFB, D.C. 20332
(202) 767-4704

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)

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

OCA RESEARCH PROPERTY COORDINATOR
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
- Final Fiscal Report (SF 269)
- Final Report of Inventions
- Govt. Property Inventory & Related Certificate
- Classified Material Certificate
- Other

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

Assigned to: ISyE / Mathematics

COPIES TO:

Administrative Coordinator
Research Property Management
Accounting Office
Procurement Office
Research Security Services

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
**Title:** MATHEMATICAL PROGRAMMING WITH AND WITHOUT DIFFERENTIABILITY

**Authors:** M.S. Bazaraa, J.E. Spingarn

**Performing Organization:**
Georgia Institute of Technology, Atlanta, Georgia 30332

**Performing Organization Name and Address:**
AirForce Office of Scientific Research, Bolling AFB, D.C. 20332

**Abstract:**
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)
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

<table>
<thead>
<tr>
<th>SECTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. QUASI-NEWTON ALGORITHMS FOR CONSTRAINED NONLINEAR PROGRAMMING (M. S. Bazaraa)</td>
<td>1</td>
</tr>
<tr>
<td>I.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>I.2 Algorithm for Equality Constrained Problems</td>
<td>2</td>
</tr>
<tr>
<td>I.2.1 Description of the Algorithm</td>
<td>3</td>
</tr>
<tr>
<td>I.2.2 Computational Expedients</td>
<td>3</td>
</tr>
<tr>
<td>I.3 Global Convergence Properties</td>
<td>5</td>
</tr>
<tr>
<td>II. GENERIC OPTIMALITY CONDITIONS AND NONDIFFERENTIABLE OPTIMIZATION (J. E. Spingarn)</td>
<td>7</td>
</tr>
<tr>
<td>II.1 Abstract</td>
<td>7</td>
</tr>
<tr>
<td>II.2 Research Summary</td>
<td>8</td>
</tr>
<tr>
<td>II.3 Publications</td>
<td>15</td>
</tr>
<tr>
<td>III. ON GENERIC PROPERTIES OF VARIATIONAL PROBLEMS (J. E. Spingarn)</td>
<td>16</td>
</tr>
</tbody>
</table>
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.
1.2 Algorithm for Equality Constrained Problems

Consider the following nonlinear programming problem, where $f$ and $h_i$ are continuously differentiable functions on $\mathbb{R}^n$.

$$
P : \text{Minimize} \quad f(x)$$

$$\text{Subject to} \quad h_i(x) = 0 \quad i=1,...,\ell$$

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.

1.2.1 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 $V_h$ denotes the $nx\ell$ gradient matrix whose $i$th column is $V_{h_i}$:

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

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

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 $\gamma$ satisfying:

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

where,

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

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

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 $\phi_r(x) = f(x) + \sum_{i=1}^{\ell} |h_i(x)|$, 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 $\phi_r$ provided that $r \geq |v_{i_1}|$ for $i = 1, \ldots, \ell$, where $v_{i_1}$ is the Lagrangian multiplier associated with the $i$th 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.

1.2.2 Computational Expedients

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

$$d_k = -u_0 + U \zeta$$

$$U = [u_{1_1}, u_{2}, \ldots, u_{\ell}]$$

The vectors $u_0$, $u_{1_1}, \ldots, u_{\ell}$, and $\zeta$ are determined by solving the following systems
\begin{align*}
B_k u_0 &= \nabla f(x_k) \\
B_k u_i &= \nabla h_i(x_k) \quad i=1, \ldots, \ell \\
A \zeta &= \nabla h(x_k)^t u_0 - h(x_k)
\end{align*}

where,
\begin{equation}
A = \nabla h(x_k)^t U 
\end{equation}

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 backward substitution. Now, to compute $\zeta$, the matrix $A$ is determined. Here, the $ij$th 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 $LL^t$. The vector $\zeta$ is calculated by solving the system $LL^t \zeta = \nabla h(x_k)^t u_0 - h(x_k)$. Now the search direction $d_k$ is at hand.
1.3 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 1.2.1. Let $f$ and $h_i$ for $i=1,...,l$ 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 $r \geq ||\xi_k||_\infty$ for each $k$ where $\xi_k$ 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 $0 \in T(x)$ 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 Research Summary

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 $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is monotone provided that
\[ \langle x_1 - x_2, y_1 - y_2 \rangle \geq 0 \text{ whenever } y_1 \in T(x_1) \text{ and } y_2 \in T(x_2). \]

The graph of $T$ is the set $G(T) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : y \in T(x)\}$. $T$ is maximal monotone if $G(T)$ is not properly contained in the graph of another monotone mapping. If $T$ is maximal monotone, the proximal point algorithm provides a method for finding $x \in \mathbb{R}^n$ 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, then the mapping $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, defined by
\[ \partial f(x) = \{y : f(x') \geq f(x) + \langle x' - x, y \rangle \text{ } \forall x' \in \mathbb{R}^n\} \]

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 $T(x, y) = \partial_x L(x, y) - \partial_y L(x, y)$ 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 non-differentiable 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 $f: \mathbb{R}^n \to \mathbb{R}$ is defined to be the set-valued mapping $\partial f: \mathbb{R}^n \to 2^{\mathbb{R}^n}$ 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 $f: \mathbb{R}^n \to \mathbb{R}$, $f$ is convex if, and only if, $\partial f$ is monotone.

A function $f: \mathbb{R}^n \to \mathbb{R}$ is lower-C^1 if for every $x \in \mathbb{R}^n$ there is a neighborhood $U$ of $x$, a compact set $S$, and a function $g: U \times S \to \mathbb{R}$ 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 characterization 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, \( \partial f \) is "strictly submonotone". \( \partial f \) is strictly submonotone iff

\[
\lim \inf_{x_i \to \bar{x}, y_i \in \partial f(x_i)} \frac{\langle x_1 - x_2, y_1 - y_2 \rangle}{|x_1 - x_2|} > 0
\]

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 \( x \in \mathbb{R}^n \), \( y \in \mathbb{R}^n \), 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 \( \partial f \) 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 \( f: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\} \) to be lower-\( C^2 \) if \( f = g - h \) for some lower semicontinuous function \( g: \mathbb{R}^n \to \mathbb{R} \cup \{\infty\} \) and some \( C^2 \) function \( h \). T is strictly hypomonotone provided for every bounded set \( K \subseteq \mathbb{R}^n \) 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 $x \in \mathbb{R}^n$, 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 overcome 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 semidefiniteness) of that derivative. This is a very strong assumption, and perhaps one which might seem unnatural. It would, after all, be pointless to establish 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 $T: \mathbb{R}^n \to \mathbb{R}^n$ is a maximal monotone mapping and $C \subseteq \mathbb{R}^n$ is a convex set, the associated variational problem is to find $z \in \mathbb{R}^n$ 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 prove 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 $\tilde{w} \in \mathbb{R}^n$ belonging to a set of measure zero, every $\tilde{z}$ satisfying

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

also satisfies

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

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 $\tilde{w} \in \mathbb{R}^n$ in a set of measure zero, $\tilde{w} \in T(\tilde{z}) + N_C(\tilde{z})$ implies that $S^{-1} = (T+N_C)^{-1}$ is differentiable at $\tilde{w}$ and the derivative $A$ satisfies kernel $(A) = L_C(\tilde{z})$ and range $(A) = L_C(\tilde{z})$, where $L_C(x) = \{ y \in \mathbb{R}^n : y \cdot y' = 0, y' \in N(x) \}$

These results have interesting consequences for the family

$$(P_{vu}) \quad \text{minimize } f(x) - x \cdot v \text{ subject to } g_i(x) \leq u_i, \quad i = 1, \ldots, m, \text{ and } x \in D$$

of convex programming problems indexed by $(v, u) \in \mathbb{R}^n \times \mathbb{R}^m$. 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 $(\tilde{v},\tilde{u})$ belonging to a set of measure zero, we can make the following assertion. If $\tilde{x}$ is a solution to $(P_{vu})$, if $F$ is the unique face of $D$ such that $\tilde{x} \in \text{relint } F$, and if $\tilde{y} \in \mathbb{R}^m_+$ satisfies the Kuhn-Tucker conditions (for almost all $(\tilde{v},\tilde{u})$, such $\tilde{y}$ will in fact exist), then
\( \{ \nabla (g_i | F)(\tilde{x}) : i \in I(\tilde{x}) \} \) is linearly independent

\( H = \nabla^2_{x} L(\tilde{x}, \tilde{y}) \) satisfies \( \zeta' H \zeta > 0 \) whenever

\[ 0 \neq \zeta \in \langle F \rangle \text{ and } \zeta \cdot V g_i (\tilde{x}) = 0 \text{ } i \in I(\tilde{x}) \]

where \( I(\tilde{x}) = \{ i : g_i (\tilde{x}) = \bar{u}_i \} \); \( \nabla (g_i | F)(\tilde{x}) \) denotes the gradient at \( \tilde{x} \) of the restriction of \( g_i \) to \( F \), or equivalently, the projection of \( V g_i (\tilde{x}) \) onto \( \langle F \rangle \); and \( \nabla^2_{x} L(\tilde{x}, \tilde{y}) \) is the Hessian of \( L \) at \( (\tilde{x}, \tilde{y}) \) with respect to \( x \).
II.3 PUBLICATIONS (SPINGARN)


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 \( T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \) 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
\( \varepsilon > 0 \), \( \emptyset \neq T(x+h) \subset T(x) + A(h) + \varepsilon |h| B \) 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 \( C \subset \mathbb{R}^n \) is a closed convex set, \( x \in C \), we define
\( N_C(x) = \{ y \in \mathbb{R}^n : y \cdot (x' - x) \leq 0, \forall x' \in C \} \)
\( L_C(x) = \{ y \in \mathbb{R}^n : y \cdot y' = 0, \forall y' \in N_C(x) \} \).
The relative interior of \( C \), \( \text{relint} C \), is the interior of \( C \) relative to the
smallest affine flat containing \( C \). The distance from \( x \) to \( C \) is denoted by
\( \text{dist}(x,C) \). A multifunction \( T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \) is monotone if \( \langle x-x',y-y' \rangle \geq 0 \) whenever
\( y \in T(x) \), \( y' \in T(x') \). \( T \) is maximal monotone if \( T \) cannot be properly
extended to a monotone mapping.

VARIATIONAL PROBLEMS

The variational problem associated with the maximal monotone mapping \( \tau : \mathbb{R}^n \to \mathbb{R}^n \) and the closed convex set \( C \subseteq \mathbb{R}^n \) is to find \( z \in \mathbb{R}^n \) such that

\[
0 \in \tau(z) + N_C(z).
\]

Theorem 1 shows that if \( \tau \) 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:

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

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 "cyrto-hedron" (this refers to a class of piecewise smooth sets defined in Spin"eurn [6]).

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

\[
\text{(3a) } \tilde{w} \in \tau(\tilde{z}) + N_C(\tilde{z})
\]

also satisfies

\[
\text{(3b) } \tilde{w} \in \tau(\tilde{z}) + \text{relint } N_C(\tilde{z}).
\]

The proof of Theorem 1 hinges on

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

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 \( \mathbb{R}^n \) such that \( \text{relint dom}(T_1) \cap \text{relint dom}(T_2) \neq \emptyset \), 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 \( Q \subseteq \mathbb{R}^n \) such that \( w \notin Q \) implies that either \( T^{-1}(w) = \emptyset \) or \( T^{-1} \) is differentiable at \( w \).

Suppose \( \tilde{w} \notin Q \) and (3a) holds. In order to produce a contradiction, let us assume that (3b) does not hold. For simplicity, we assume \( \tilde{w} = 0 \). \( T^{-1} \) is differentiable at 0 since \( 0 \notin Q \) \( (T^{-1}(0) \neq \emptyset \text{ because } 0 \in \tau(\tilde{z})) \). Choose \( \tilde{y} \in \tau(\tilde{z}) + \text{relint } N_C(\tilde{z}) \). Then \( t\tilde{y} \in \tau(\tilde{z}) + N_C(\tilde{z}) \) for \( 0 \leq t \leq 1 \) by convexity.
If $A$ is the derivative of $T^{-1}$ at 0 then for any $\varepsilon > 0$,  
\begin{equation}
0 \neq T^{-1}(ty) < z + tA(y) + \varepsilon tB
\end{equation}
for all $t$ sufficiently small. But $z \in T^{-1}(ty)$ for $0 \leq t \leq 1$, so $0 \in tA(y) + \varepsilon tB$. Cancelling $t$, and letting $\varepsilon$ approach zero, we see that $A(y) = 0$.

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

**THEOREM 2.** Let $C$ and $\tau$ be as in Theorem 1. Except for $\tilde{w} \in \mathbb{R}^n$ in a set of measure zero, (3a) implies that $T^{-1} = (\tau + N_C)^{-1}$ is differentiable at $\tilde{w}$ and the derivative $A$ satisfies
\[\text{kernel}(A) = L_C(\tilde{z})^\perp \text{ and range}(A) = L_C(\tilde{z}).\]

**Proof:** Fix $\tilde{w} \in Q$, 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 $\tilde{w}$, $\tilde{z} = T^{-1}(\tilde{w})$, and (3b) holds. Let $A$ be the derivative of $T^{-1}$ at $\tilde{w}$. Again, for simplicity we take $\tilde{w} = 0$.

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

Let $\pi$ denote orthogonal projection onto $L_C(\tilde{z})$. Then
\[|\tau(z_t) - \tau(\tilde{z})| = |tu - y_t - \tau(\tilde{z})|
\geq |\pi(tu - y_t - \tau(\tilde{z}))|
= |tu - \pi(y_t)|
\geq t|u| - |\pi(y_t)|.
\]

Now,
\[|\pi(y_t)| \leq \text{dist}(y_t, N_C(\tilde{z})) \leq \mu_z \varepsilon t |y_t|
\leq \mu_z \varepsilon t (|tu| + |\tau(z_t)|) \leq \mu_z \varepsilon t M
\]
where $M$ is some bound for the quantity in parentheses, for $t$ in some neighborhood of 0. Hence, $|\tau(z_t) - \tau(\tilde{z})| \geq t(|u| - \mu_z \varepsilon M)$. This shows that for any $\varepsilon > 0$. \(\square\)
holds for all $t$ sufficiently near zero. Since $c$ can be chosen arbitrarily small, this contradicts the Lipschitz property of $\tau$. Thus no such $u$ exists, and we must have $L_C(\tilde{z})^\perp = \text{kernel}(A)$.

Suppose next that $Au \notin L_C(\tilde{z})$ for some $u$. Then $v'Au < -u'Au$ for some $0 \neq v \in L_C(\tilde{z})^\perp$, and hence $(v+u)'A(v+u) = (v+u)'Au = v'Au + u'Au < 0$. However, $A$, being the derivative at $\tilde{w}$ 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 $\text{range}(A) \subseteq L_C(\tilde{z})$. By a dimension argument, it is clear that we must actually have $\text{range}(A) = L_C(\tilde{z})$. 

CONSEQUENCES IN CONVEX PROGRAMMING

Consider the parametrized family

$$(P_{vu}) \quad \text{minimize } f(x) - x \cdot v \text{ subject to } g_i(x) \leq u_i, \quad i = 1, \ldots, m, \text{ and } x \in D$$

of convex programming problems indexed by $(v,u) \in \mathbb{R}^n \times \mathbb{R}^m$. Here, the functions $f, g_1, \ldots, g_m$ are finite-valued continuous 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

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

$$C = D \times \mathbb{R}^m_+$$

$$\tau(x,y) = (\nabla f(x) + \sum_{i=1}^m y_i \nabla g_i(x), \ldots, -g_i(x), \ldots).$$

Since $D$ satisfies (2), so does $C$. Also, $\tau$ is maximal monotone and locally Lipschitz, so that the assumptions for Theorem 1 are satisfied. Hence, except for $w = (v,u) \in \mathbb{R}^n \times \mathbb{R}^m$ 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

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

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

and $x \in D$

while (3b) holds iff

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

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

$$y_i > 0 \text{ iff } g_i(x) < u_i, \quad i = 1, \ldots, m, \text{ and } x \in D.$$
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 \(u \in \mathbb{R}^m\) 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 \(u \in \text{int} K\), \(v \in \mathbb{R}^n\), and \(x\) solves \((P_{vu})\), then \((x,y)\) satisfies (5a). Since the boundary of \(K\) is of measure zero, the proof is complete. \(\square\)

According to Theorem 2, the mapping \(T^{-1} = (\tau + N_C)^{-1}\) is differentiable at almost all \(w = (v,u) \in \mathbb{R}^n \times \mathbb{R}^m\). 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 \(C = D \times \mathbb{R}^m\) is also polyhedral convex. Suppose that \(\tilde{w} = (\tilde{v}, \tilde{u})\) and \(\tilde{z} = (\tilde{x}, \tilde{y})\) satisfy the conclusions of Theorem 2. In other words, assume that \(T^{-1} = (\tau + N_C)^{-1}\) is differentiable at \(\tilde{w}\), \(\tilde{z} = T^{-1}(\tilde{w})\), and the derivative \(A\) satisfies

\[
\begin{align*}
\text{(6a)} & \quad \text{kernel}(A) = L_C(\tilde{z})^\perp \\
\text{(6b)} & \quad \text{range}(A) = L_C(\tilde{z}) \\
\text{(6c)} & \quad \tilde{w} \in \tau(\tilde{z}) + \text{relint } N_C(\tilde{z}).
\end{align*}
\]

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 \(\mathbb{R}^n \times \mathbb{R}^m\) 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 \(\pi\) 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 \(z\). For simplicity, consider only the case \(\tilde{w} = 0\). Since \(\pi(\tilde{w}) = \pi(\tau(\tilde{z}))\), we have \(\pi(\tau(\tilde{z})) - \tau(\tilde{z}) \in \text{relint } N_C(G)\). For \(z \in G\) sufficiently close to \(\tilde{z}\), since \(\tau\) 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 \(z \in G\) sufficiently close to \(\tilde{z}\), which implies that \(\tau\) is continuous at \(\tilde{z}\).
ciently near \( \bar{z} \). On the other hand, if \( z \in T^{-1}(u) \) with \( u \in <G> \) and \( z \in G \) near \( \bar{z} \), then clearly \( u = \pi(\tau(z)) \). This demonstrates that for \( u \in <G> \) 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 \( <G> \) being a nonsingular mapping : \( <G> \to <G> \) by (6a) and (6b).

Hence \((\pi \circ \tau)^{-1}\) considered as a map : \( <G> \to G \) is differentiable at 0, with nonsingular derivative. It is an easy consequence of this that \( \pi \circ \tau : G \to <G> \) 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 \( X \) is a solution to \((P_{\bar{v}, \bar{u}})\), if \( F \) is the unique face of \( D \) such that \( \bar{x} \in \text{relint} F \), 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

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

\[
(7b) \quad H = \nabla^2_{\bar{x}} L(\bar{x}, \bar{y}) \text{ satisfies } \zeta^t H \zeta > 0 \text{ whenever } 0 \neq \zeta \in <F> \text{ and } \zeta \cdot \nabla g_i(\bar{x}) = 0 \forall i \in I(\bar{x})
\]

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 \( <F> \); and \( \nabla^2_{\bar{x}} L(\bar{x}, \bar{y}) \) is the Hessian of \( L \) at \((\bar{x}, \bar{y})\) with respect to \( x \).

**Proof:** If we let \( E = \{y \in R^m_+ : y_i > 0 \text{ iff } i \in I(\bar{x})\} \), 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**


