RR-76-21

SEMISMOOTH AND SEMICONVEX FUNCTIONS IN CONSTRAINED OPTIMIZATION

ROBERT MIFFLIN

DECEMBER 1976

International Institute for Applied Systems Analysis
2361 Laxenburg, Austria

Preface

Large-scale optimization models arise in many areas of application at IIASA. For example, such models are useful for estimating the potential economic value of solar and wind energy and for determining equilibrium prices for agricultural commodities in international trade as a function of national policies. Certain methods of decomposition for solving such optimization problems require the solution of a relatively small problem whose objective function is not everywhere differentiable. This paper defines nonsmooth functions that can arise from such decomposition approaches and that can be effectively optimized by recently proposed methods for nondifferentiable optimization.

ABSTRACT

We introduce semismooth and semiconvex functions and discuss their properties with respect to nonsmooth nonconvex constrained optimization problems. These functions are locally Lipschitz, and hence have generalized gradients. The author has given an optimization algorithm that uses generalized gradients of the problem functions and converges to stationary points if the functions are semismooth. If the functions are semiconvex and a constraint qualification is satisfied, then we show that a stationary point is an optimal point.

We show that the pointwise maximum or minimum over a compact family of continuously differentiable functions is a semismooth function and that the pointwise maximum over a compact family of semiconvex functions is a semiconvex function. Furthermore, we show that a semismooth composition of semismooth functions is semismooth and gives a type of chain rule for generalized gradients.

Semismooth and Semiconvex Functions In Constrained Optimization

1. INTRODUCTION

In this paper we are interested in an inequality constrained optimization problem where the functions need not be differentiable or convex. More precisely, consider the problem of finding an $x \in R^n$ to

$$\text{minimize } f(x)$$
$$\text{subject to } h_i(x) \leq 0 \quad \text{for } i = 1,2,\ldots,m$$

where $h_1, h_2, \ldots, h_m$ and f are real-valued functions defined on R^n.

We utilize the "generalized gradient" introduced by Clarke [1,2] for "locally Lipschitz" functions. A necessary condition [2] (of the Karush [5] - John [4] type) for optimality of a point $\bar{x}$ is that the zero vector is a certain convex combination of generalized gradients of $h_1, h_2, \ldots, h_m$ and f at $\bar{x}$. In section 5 of this paper, this "stationarity" condition is concisely stated in terms of a map as given by Merrill [10] depending on the problem function generalized gradients. Our implementable algorithm for nonsmooth nonconvex optimization given in [11] uses this map and converges to such stationary points if the problem functions are "semismooth" as defined here in section 2. This algorithm can be viewed as a modification and extension of the "conjugate subgradient" type algorithms for nondifferentiable unconstrained optimization given by Lemarechal [8] and Wolfe [16] for convex functions and by Feuer [3] for min-max objectives.

Semismooth functions possess a semicontinuous relationship between their generalized gradients and directional derivatives. They are related to, but different from, the "almost differentiable" functions of Shor [13]. Notable examples of such functions are convex, concave and continuously differentiable functions.

In section 2 we also define "semiconvex" functions. These functions are "quasidifferentiable" (Pshenichnyi [12]) and essentially "semiconvexe" in the sense of Tuy [15] and, if also differentiable, are "pseudoconvex" (Mangasarian [9]). In section 5 we show that the above stationarity condition is sufficient for optimality if the problem functions are semiconvex and a constraint qualification is satisfied. This is a nondifferentiable analogue of a sufficient optimality result in [9, Theorem 10.1.1].

In sections 3 and 4, we give some important properties of semismooth and semiconvex functions. Starting from the work in [1] and [3] on min-max objectives, we show that the pointwise maximum or minimum over a compact family of continuously differentiable functions is a semismooth function. We also give an example of a semismooth function that is an extremal combination not of continuously differentiable functions, but of semismooth functions. This leads us to show that a semismooth composition of semismooth functions is semismooth and to give a type of "chain rule" for generalized gradients. Special cases of this chain rule may be found in [2].

In section 3 we also show that the pointwise maximum over a compact family of semiconvex functions is a semiconvex function. Thus, semiconvex functions behave as do convex functions with respect to the maximization operation, while pseudoconvex functions do not because of the loss of differentiability due to this nonsmooth operation.

2. DEFINITIONS AND EXAMPLES OF SEMISMOOTH AND SEMICONVEX FUNCTIONS

Let B be an open subset of R^n and $F : R^n \to R$ be *Lipschitz* on B, i.e. there exists a positive number K such that

$$|F(y) - F(z)| \leq K|y-z| \quad \text{for all } y,z \in B \quad .$$

If F is Lipschitz on each bounded subset of R^n then F is called *locally Lipschitz*.

Let $x \in B$ and $d \in R^n$. As in Clarke [2], let

$$F^0(x;d) = \limsup_{\substack{h\to 0\\ t\downarrow 0}} [F(x+h+td) - F(x+h)]/t$$

and let $\partial F(x)$ denote the *generalized gradient* of F at x defined by

$$\partial F(x) = \{g \in R^n : \langle g,d\rangle \leqq F^0(x;d) \text{ for all } d \in R^n\} \quad .$$

The following two propositions collect together useful properties of F^0 and ∂F from Clarke [1,2] and Lebourg [7], respectively.

Proposition 1.

(a) $\partial F(x)$ is a nonempty convex compact subset of R^n.

(b) $F^0(x;d) = \max\, [\langle g,d\rangle : g \in \partial F(x)]$

(c) F is differentiable almost everywhere in B and $\partial F(x)$ is the convex hull of all the points g of the form

$$g = \lim_{k\to\infty} \nabla F(x_k)$$

where $\{x_k\} \to x$ and F has a gradient ∇F at each $x_k \in B$.

(d) If $\{x_k\} \subset B$ converges to x and $g_k \in \partial F(x_k)$ for each k then $|g_k| \leqq K$ and each accumulation point g of $\{g_k\}$ satisfies $g \in \partial F(x)$, i.e. ∂F is bounded on bounded subsets of B and ∂F is *uppersemicontinuous* on B.

Proposition 2.

Let y and z be in a convex subset of B. Then there exists $\lambda \in (0,1)$ and $g \in \partial F(y+\lambda(z-y))$ such that

$$F(z) - F(y) = \langle g,z-y\rangle \quad ,$$

i.e. a mean value result holds.

By combining part (d) of Proposition 1 with Proposition 2 one may easily establish the following useful result:

Lemma 1. Let $\{t_k\} \downarrow 0$, $\{h_k\} \to 0 \in R^n$ and F^* be any accumulation point of

$$\{[F(x+h_k+t_kd) - F(x+h_k)]/t_k\} \quad .$$

Then there exists $g \in \partial F(x)$ such that

$$F^* = \langle g,d\rangle \quad .$$

If $\lim_{t\downarrow 0} [F(x+td) - F(x)]/t$ exists it is denoted by $F'(x;d)$ and called the *directional derivative* of F at x in the direction d. If $F'(x;d)$ exists and equals $F^0(x;d)$ for each $d \in R^n$ the F is said to be *quasidifferentiable* at x (Pshenichnyi [12]). Note that if $F'(x;d)$ exists then, by Lemma 1, there exists $g \in \partial F(x)$ such that

$$F'(x;d) = \langle g,d\rangle$$

and, if, in addition, F is quasidifferentiable at x, then, by parts (a) and (b) of Proposition 1, g is a maximizer of $\langle\cdot,d\rangle$ over $\partial F(x)$.

Definition 1. $F : R^n \to R$ is *semismooth* at $x \in R^n$ if

(a) F is Lipschitz on a ball about x

and

(b) for each $d \in R^n$ and for any sequences $\{t_k\} \subset R_+$, $\{\theta_k\} \subset R^n$ and $\{g_k\} \subset R^n$ such that

$$\{t_k\} \downarrow 0,\ \{\theta_k/t_k\} \to 0 \in R^n \text{ and } g_k \in \partial F(x+t_kd+\theta_k) \quad ,$$

the sequence $\{\langle g_k,d\rangle\}$ has exactly one accumulation point.

Lemma 2. If F is semismooth at x then for each $d \in R^n$, $F'(x;d)$ exists and equals $\lim_{k\to\infty} \langle g_k,d\rangle$ where $\{g_k\}$ is any sequence as in Definition 1.

Proof: Suppose $\{\tau_k\} \downarrow 0$. By Proposition 2, there exist $t_k \in (0,\tau_k)$ and $g_k \in \partial F(x+t_kd)$ such that

$$F(x+\tau_kd) - F(x) = \tau_k\langle g_k,d\rangle \quad .$$

Then, by Definition 1 with $\theta_k = 0 \in R^n$, since $\{t_k\} \downarrow 0$,

$$\lim_{k\to\infty} [F(x+\tau_k d) - F(x)]/\tau_k = \lim_{k\to\infty} \langle g_k, d\rangle \quad .$$

Since $\{\tau_k\}$ is an arbitrary positive sequence converging to zero, $F'(x;d)$ exists and equals the desired limit.□

Definition 2. Let X be a subset of R^n. $F : R^n \to R$ is *semiconvex* at $x \in X$ (with respect to X) if

(a) F is Lipschitz on a ball about x

(b) F is quasidifferentiable at x

and

(c) $x + d \in X$ and $F'(x;d) \geqq 0$ imply $F(x+d) \geqq F(x)$.

Tuy's [15] earlier concept of semiconvexity does not include quasidifferentiability, but we include it in order to obtain Theorems 8 and 9 given below. A semiconvex function that is also differentiable is called "pseudoconvex" (Mangasarian [9, Chapter 9]).

We say that F is semismooth (quasidifferentiable, semiconvex) on $X \subset R^n$ if F is semismooth (quasidifferentiable, semiconvex) at each $x \in X$. We denote the convex hull of a set S by conv S.

From convex analysis [13, Sections 23 and 24] and [2, Proposition 3] we have the following:

Proposition 3.

If $F : R^n \to R$ is convex (concave) then F(F) is locally Lipschitz,

$$\partial F(x) = \{g \in R^n : F(y) \geqq (\leqq)\ F(x) + \langle g, y-x\rangle \text{ for all } y \in R^n\} \text{ for each } x \in R^n \quad ,$$

i.e. ∂F is the *subdifferential* of F, F(-F) is semiconvex on R^n and F(F) is semismooth on R^n.

From [2, Proposition 4] and the properties of continuously differentiable functions we have the following:

Proposition 4.

If $F : R^n \to R$ is continuously differentiable then F is locally Lipschitz, $\partial F(x) = \{\nabla F(x)\}$ for each $x \in R^n$, and F is quasidifferentiable and semismooth on R^n.

An example of a locally Lipschitz function on R that is not semismooth (nor quasidifferentiable) is the following differentiable function that is not continuously differentiable:

$$F(x) = \begin{cases} x^2 \sin(1/x) & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases} .$$

Note that $F'(0;1) = 0$ and $\partial F(0) = \text{conv}\{-1,1\}$ is the set of possible accumulation points of $F'(x;1)$ as $x \downarrow 0$.

An example of a function that is semiconvex and semismooth on R, but not convex nor differentiable, is

$$F(x) = \log(1+|x|)$$

where

$$\partial F(x) = \begin{cases} 1/(1+x) & \text{for } x > 0 \\ \text{conv}\{-1,1\} & \text{for } x = 0 \\ -1/(1-x) & \text{for } x < 0 \end{cases} .$$

Note that in a neighborhood of $x = 0$

$$F(x) = \max[\log(1+x), \log(1-x)] ,$$

i.e. F is a pointwise maximum of smooth functions. General functions of this type are the subject of the next section.

3. SEMISMOOTH AND SEMICONVEX EXTREMAL-VALUED FUNCTIONS

In this section we supplement developments in Feuer [3] and Clarke [1] to show that certain extremal-valued functions E are semismooth and/or semiconvex.

Suppose $E : R^n \to R$ is defined on B, an open subset of R^n, as follows in terms of $f : R^n \times T \to R$ where T is a topological space:

Suppose there exists a sequentially compact subspace U of T such that

(a) $f(x,u)$ is continuous for $(x,u) \in B \times U$

(b) $f(x,u)$ is Lipschitz for $x \in B$ uniformly for $u \in U$

(c) $\partial_x f(x,u)$ is uppersemicontinuous for $(x,u) \in B \times U$

and for each $x \in B$

either

(d) $E(x) = \max\ [f(x,u) : u \in U]$

and

(e) $f'_x(x,u;d) = f^0_x(x,u;d)$ for all $(u,d) \in U \times R^n$

or

(d') $E(x) = \min\ [f(x,u) : u \in U]$

and

(e') $f'_x(x,u;d) = -f^0_x(x,u;-d)$ for all $(u,d) \in U \times R^n$

For each $x \in B$ let

$$A(x) = \{u \in U : E(x) = f(x,u)\} \quad .$$

Note that E and A are well defined by the continuity and compactness assumptions. Furthermore, for each $x \in B$, $A(x)$ is compact and $\partial_x f(x,\cdot)$ is uppersemicontinuous and bounded on U, and a direct consequence of [1, Theorem 2.1] is the following:

Theorem 1. Let the above assumptions on E *and* f *hold. Then* E *is Lipschitz on* B *and for each* $x \in B$

$$\partial E(x) = \text{conv}\ \{\partial_x f(x,u) : u \in A(x)\} \quad ,$$

and for each $d \in R^n$

$$E'(x;d) = E^0(x;d) = \max\ [\langle g,d\rangle : g \in \partial_x f(x,u), u \in A(x)]$$

if (d) *and* (e) *hold, or*

$$E'(x;d) = -E^0(x;-d) = \min\ [\langle g,d\rangle : g \in \partial_x f(x,u), u \in A(x)]$$

if (d') *and* (e') *hold.*

Remark: Feuer [3] shows the results of Theorem 1 under the stronger assumptions of our next theorem and proves a result [3, p. 57] close to semismoothness from which our next proof is adapted.

Theorem 2. Suppose that (a) *and* (d) *or* (d') *hold and that* $f(\cdot,u)$ *is differentiable on* B *for each* $u \in U$ *and* $\nabla_x f(\cdot,\cdot)$ *is continuous and bounded on* $B \times U$. *Then* E *is semismooth on* B.

Proof: Note that the additional assumption implies (b), (c), (e), and (e') and that $\partial_x f = \nabla_x f$ on $B \times U$. Suppose E has the max form (d). (The proof of semismoothness for the min form (d') is similar.) Let $x \in B$, $d \in R^n$, $x_k = x + t_k d + \theta_k$ and $g_k \in \partial E(x_k)$ where $\{t_k\} \downarrow 0$ and $\{\theta_k/t_k\} \to 0 \in R^n$. From Theorem 1 and Proposition 1 we have that

$$E'(x;d) = E^0(x;d) = \max\ [\langle g,d\rangle : g \in \partial E(x)]$$

and ∂E is bounded and uppersemicontinuous on a ball about x, so

$$\limsup_{k\to\infty} \langle g_k,d\rangle \leq E'(x;d) \quad .$$

Suppose

$$\liminf_{k\to\infty} \langle g_k,d\rangle < E'(x;d) \quad ,$$

i.e. there is an $\varepsilon > 0$ and a subsequence of $\{g_k\}$ such that on this subsequence

$$\{\langle g_k,d\rangle\} \to E'(x;d) - \varepsilon \quad . \tag{3.1}$$

For each k corresponding to this subsequence choose $\bar{g}_k \in \partial E(x_k)$ and $u_k \in A(x_k)$ such that

$$\bar{g}_k = \nabla_x f(x_k,u_k) \in \text{conv}\ \{\nabla_x f(x_k,u) : u \in A(x_k)\} = \partial E(x_k)$$

and

$$\langle \bar{g}_k,d\rangle = \min\ [\langle g,d\rangle : g \in \partial E(x_k)] \leqq \langle g_k,d\rangle \quad . \tag{3.2}$$

Since $\nabla_x f$ is continuous on $B \times U$, $\{x_k\} \to x$ and $\{u_k\}$ is in the compact set U, $\{\bar{g}_k\}$ and $\{u_k\}$ have accumulation points $\bar{g}$ and $\bar{u}$, respectively, such that

$$\bar{g} = \nabla_x f(x,\bar{u}) \quad .$$

Thus, by (3.1) and (3.2),

$$\langle \nabla_x f(x,\bar{u}),d\rangle = \langle \bar{g},d\rangle \leqq E'(x;d) - \varepsilon \quad .$$

Let $u^* \in A(x)$ be such that

$$E'(x;d) = E^0(x;d) = \max\ [\langle \nabla_x f(x,u),d\rangle : u \in A(x)] = \langle \nabla_x f(x,u^*),d\rangle \quad .$$

Then

$$\langle \nabla_x f(x,\bar{u}),d\rangle \leqq \langle \nabla_x f(x,u^*),d\rangle - \varepsilon$$

and, since $\langle \nabla_x f(\cdot,\cdot),\cdot\rangle$ is continuous, there exist neighborhoods $B(x)$, $V(\bar{u})$ and $D(d)$ such that

$$\langle \nabla_x f(z,u),\delta\rangle \leqq \langle \nabla_x f(z,u^*),\delta\rangle - \varepsilon/2 \text{ for all } (z,u,\delta) \in B(x) \times V(\bar{u}) \times D(d) \quad .$$

Choose k so large that $u_k \in V(\bar{u})$, $t_k|d| + |\theta_k|$ is less than the radius of a ball about x contained in $B(x)$ and $2|\theta_k/t_k|$ is less than the radius of a ball about d contained in $D(d)$. Then for all $t \in [0,t_k]$,

$$x(t) \equiv x + td + (t/t_k)^2\theta_k \in B(x) \quad ,$$

and

$$x'(t) = d + 2(t/t_k)(\theta_k/t_k) \in D(d) \quad .$$

Then

$$<\nabla_x f(x(t),u_k),x'(t)> \leqq <\nabla_x f(x(t),u^*),x'(t)> - \varepsilon/2 \quad \text{for all } t \in [0,t_k] \quad .$$

Integrating from $t = 0$ to $t = t_k$ gives

$$f(x(t_k),u_k) - f(x(0),u_k) \leqq f(x(t_k),u^*) - f(x(0),u^*) - t_k\varepsilon/2 \quad .$$

But $x(t_k) = x_k$, $x(0) = x$, $u_k \in A(x_k)$ and $u^* \in A(x)$,

so $$E(x_k) - f(x,u_k) \leqq f(x_k,u^*) - E(x) - t_k\varepsilon/2 \quad ,$$

or

$$E(x_k) + E(x) \leqq f(x_k,u^*) + f(x,u_k) - t_k\varepsilon/2 \quad .$$

But this leads to a contradiction, because $f(x_k,u^*) \leqq E(x_k)$, $f(x,u_k) \leqq E(x)$, $t_k > 0$ and $\varepsilon > 0$. Thus, $\lim_{k\to\infty} <g_k,d> = E'(x;d)$, so E is semismooth at x.□

Theorem 3. *Let* X *be a subset of* B. *Suppose that* (a), (b), (c), (d), *and* (e) *hold, i.e.* E *is a max function, and suppose that* $f(\cdot,u)$ *is semiconvex at* $x \in X$ *(with respect to* X*) for each* $u \in U$. *Then* E *is semiconvex at* $x \in X$ *(with respect to* X*).*

Proof: By Theorem 1, E is Lipschitz on a ball about x, quasidifferentiable at x, and for $d \in R^n$ there exist $\bar{u} \in A(x)$ and $\bar{g} \in \partial_x f(x,\bar{u})$ such that

$$E'(x;d) = <\bar{g},d> = \max\ [<g,d> : g \in \partial_x f(x,u), u \in A(x)] \quad .$$

Suppose $x + d \in X$ and $E'(x;d) \geqq 0$. Then, by the quasidifferentiability of $f(\cdot,\bar{u})$ at x, we have

$$f'_x(x,\bar{u};d) = f^0_x(x,\bar{u};d) = \max\ [\langle g,d\rangle : g \in \partial_x f(x,\bar{u})] \geqq \langle \bar{g},d\rangle \geqq 0 \quad .$$

Thus, by the semiconvexity of $f(\cdot,\bar{u})$ at x,

$$f(x+d,\bar{u}) \geqq f(x,\bar{u}) \quad .$$

But $x + d \in X \subset B$ and assumption (d) imply

$$E(x+d) \geqq f(x+d,\bar{u})$$

and $\bar{u} \in A(x)$ implies

$$E(x) = f(x,\bar{u}) \quad ,$$

so

$$E(x+d) \geqq f(x+d,\bar{u}) \geqq f(x,\bar{u}) = E(x)$$

and the semiconvexity of E at x is established.□

The following function F is an example of a semismooth function on R^2 which is not an extremal-valued function in the sense of Theorem 2, because in any ball about (0,0) there is a point at which the value of F is neither the maximum nor the minimum of the three underlying linear functions that define F:

$$F(x_1,x_2) = \begin{cases} x_1 & \text{for } x_2 \geqq 0 \text{ and } x_2 \geqq x_1 \geqq 0 \\ x_2 & \text{for } x_1 \geqq 0 \text{ and } x_1 \geqq x_2 \geqq 0 \\ 0 & \text{for } x_1 \leqq 0 \text{ or } x_2 \leqq 0 \quad . \end{cases}$$

Note that $F(x_1,x_2) = \max\ [0,\min(x_1,x_2)]$. This raises the question of whether or not a finite extremal composition of extremal-valued functions is a semismooth function. This is indeed the case, as is shown in more generality in the next section.

4. SEMISMOOTH COMPOSITION

In this section we show that a semismooth composition of semismooth functions results in a semismooth function. In order to prove this useful result we first establish a type of "chain rule" for generalized gradient sets. For $v^1, v^2, \ldots, v^m \in R^n$ let $[v^1 v^2 \cdots v^m]$ denote the $n \times m$ matrix whose i^{th} column is v^i for $i = 1, 2, \ldots, m$.

Theorem 4. *Let* $f_i : R^n \to R$ *for* $i = 1, 2, \ldots, m$ *and* $E : R^m \to R$ *be locally Lipschitz. For* $x \in R^n$ *define*

$$Y(x) = (f_1(x), f_2(x), \ldots, f_m(x)) \quad ,$$
$$F(x) = E(Y(x))$$

and

$$G(x) = \text{conv}\,\{g \in R^n : g = [g^1 g^2 \cdots g^m]w,\ g^i \in \partial f_i(x),\ i = 1, 2, \ldots, m,\ w \in \partial E(Y(x))\} \quad .$$

Then F *is locally Lipschitz and*

$$\partial F(x) \subset G(x) \text{ for each } x \in R^n \quad . \tag{4.1}$$

Remarks: Clarke [2] establishes (4.1) for the three cases where (1) E is continuously differentiable and $m = 1$, (2) $E(y_1, y_2) = y_1 + y_2$ and (3) $E(y) = \max\,[y_i : i \in \{1, 2, \ldots, m\}]$ for $y = (y_1, y_2, \ldots, y_m)$.

Note that the containment in (4.1) may be strict, because, as suggested to us by M.J.D. Powell, for $E(y_1, y_2) = y_1 - y_2$, $x \in R$ and $f_1(x) = f_2(x) = |x|$, we have $\partial F(0) = \{0\}$ and $G(0) = \text{conv}\,\{-2, 2\}$.

Proof: It is not difficult to show that F is locally Lipschitz and to show that G is uppersemicontinuous. Hence, by part (c) of Proposition 1, F is differentiable almost everywhere, and if we show

$$\nabla F(\bar{x}) \in G(\bar{x}) \tag{4.2}$$

where $\bar{x}$ is any point of differentiability of F, then (4.1) follows from the convexity and uppersemicontinuity of G.

In order to show (4.2), let $\nabla F(\bar{x})$ exist, $d \in R^n$ and $\{t_k\} \downarrow 0$. Then

$$\langle \nabla F(\bar{x}), d\rangle = F'(\bar{x};d)$$
$$= \lim_{k\to\infty} [F(\bar{x}+t_k d) - F(\bar{x})]/t_k$$
$$= \lim_{k\to\infty} [E(Y(\bar{x}+t_k d)) - E(Y(\bar{x}))]/t_k \quad . \tag{4.3}$$

Choose a subsequence of $\{t_k\}$ such that for each $i = 1,2,\ldots,m$

$$\{[f_i(\bar{x}+t_k d) - f_i(\bar{x})]/t_k\} \to f_i^*$$

on the subsequence. By Lemma 1,

$$f_i^* = \langle g^i, d\rangle \text{ for some } g^i \in \partial f_i(\bar{x}) \quad , \tag{4.4}$$

so

$$\{[f_i(\bar{x}+t_k d) - f_i(\bar{x}) - t_k\langle g^i, d\rangle]/t_k\} \to 0$$

on the subsequence. Let

$$v = (f_1^*, f_2^*, \ldots, f_m^*) = (\langle d,g^1\rangle, \langle d,g^2\rangle, \ldots, \langle d,g^m\rangle) . \tag{4.5}$$

Then

$$\{[Y(\bar{x}+t_k d) - Y(\bar{x}) - t_k v]/t_k\} \to 0 \in R^m$$

and, by the Lipschitz continuity of E,

$$\{[E(Y(\bar{x}+t_k d)) - E(Y(\bar{x})+t_k v)]/t_k\} \to 0 \tag{4.6}$$

on the subsequence. Now choose a sub-subsequence of $\{t_k\}$ such that

$$\{[E(Y(\bar{x})+t_k v) - E(Y(\bar{x}))]/t_k\} \to E^* \tag{4.7}$$

on this sub-subsequence. Then, by combining (4.6) and (4.7),

$$\{[E(Y(\bar{x}+t_k d)) - E(Y(\bar{x}))]/t_k\} \to E^*$$

on the sub-subsequence and, by (4.3),

$$\langle \nabla F(\bar{x}), d\rangle = E^* \quad . \tag{4.8}$$

From (4.7) and Lemma 1,

$$E^* = \langle v,w\rangle \text{ for some } w \in \partial E(Y(\bar{x})) \quad . \tag{4.9}$$

Let

$$g = [g^1 g^2 \cdots g^m] w \quad ,$$

so that combining (4.8), (4.9), (4.5) and (4.4) and recalling the definition of G yields

$$\langle \nabla F(\bar{x}),d\rangle = E^* = \langle v,w\rangle = \langle (\langle d,g^1\rangle,\langle d,g^2\rangle,\ldots,\langle d,g^m\rangle),w\rangle = \langle d,g\rangle$$

where $g \in G(\bar{x})$. Since this result holds for each $d \in R^n$, and $G(\bar{x})$ is convex, we have that the desired result (4.2) holds, for, if not, then a strict separation theorem [9, Theorem 3.2.6] gives a contradiction.□

<u>Theorem 5</u>. Suppose, in addition to the assumptions of Theorem 4, that f_i *for each* $i = 1,2,\ldots,m$ *is semismooth at* $x \in R^n$ *and* E *is semismooth at* $Y(x) \in R^m$. *Then* F *is semismooth at* x.

Proof: Suppose $x_k = x + t_k d + \theta_k$ and $g_k \in \partial F(x_k)$ where $d \in R^n$, $\{t_k\} \downarrow 0$ and $\{\theta_k/t_k\} \to 0 \in R^n$. Since $\partial F(x_k)$ is contained in the compact convex set $G(x_k)$, by minimizing and maximizing the linear function $\langle \cdot ,d\rangle$ over $G(x_k)$, we may find $\bar{g}_k, \hat{g}_k \in G(x_k)$ such that

$$\langle \bar{g}_k,d\rangle \leqq \langle g_k,d\rangle \leqq \langle \hat{g}_k,d\rangle$$

and

$$\bar{g}_k = [\bar{g}_k^1 \bar{g}_k^2 \cdots \bar{g}_k^m]\bar{w}_k \quad , \qquad \hat{g}_k = [\hat{g}_k^1,\hat{g}_k^2,\ldots,\hat{g}_k^m]\hat{w}_k$$

where

$$\bar{g}_k^i, \hat{g}_k^i \in \partial f_i(x_k) \text{ for each } i = 1,2,\ldots,m$$

and

$$\bar{w}_k, \hat{w}_k \in \partial E(Y(x_k)) \quad .$$

By the uppersemicontinuity and local boundedness of the various maps, $\{\bar{g}_k\}$ and $\{\hat{g}_k\}$ are bounded and there are accumulation points $\bar{g}$ of $\{\bar{g}_k\}$ and $\hat{g}$ of $\{\hat{g}_k\}$ and corresponding accumulation points $\bar{g}^i$ of $\{\bar{g}_k^i\}$ and $\hat{g}^i$ of $\{\hat{g}_k^i\}$ for each $i = 1,2,\ldots,m$ and $\bar{w}$ of $\{\bar{w}_k\}$ and $\hat{w}$ of $\{\hat{w}_k\}$ such that

$$\bar{g} = [\bar{g}^1 \bar{g}^2 \cdots \bar{g}^m]\bar{w} \quad , \qquad \hat{g} = [\hat{g}^1 \hat{g}^2 \cdots \hat{g}^m]\hat{w}$$

and

$$\langle \bar{g}, d\rangle \leqq \liminf_{k\to\infty} \langle g_k, d\rangle \leqq \limsup_{k\to\infty} \langle g_k, d\rangle \leqq \langle \hat{g}, d\rangle \quad .$$

By the semismoothness of each f_i, we have

$$\langle d, \bar{g}^i\rangle = \langle d, \hat{g}^i\rangle = f_i'(x;d) \quad ,$$

so, by defining

$$z = (f_1'(x;d), f_2'(x;d), \ldots, f_m'(x;d)) \quad ,$$

we have

$$\langle d, \bar{g}\rangle = \langle d, [\bar{g}^1 \bar{g}^2 \cdots \bar{g}^m]\bar{w}\rangle = \langle z, \bar{w}\rangle \quad ,$$

$$\langle d, \hat{g}\rangle = \langle d, [\hat{g}^1 \hat{g}^2 \cdots \hat{g}^m]\hat{w}\rangle = \langle z, \hat{w}\rangle \quad ,$$

and, thus,

$$\langle z, \bar{w}\rangle \leqq \liminf_{k\to\infty} \langle g_k, d\rangle \leqq \limsup_{k\to\infty} \langle g_k, d\rangle \leqq \langle z, \hat{w}\rangle \quad .$$

So, if we show that

$$\langle z, \bar{w}\rangle = \langle z, \hat{w}\rangle \quad , \tag{4.10}$$

then $\{\langle g_k, d\rangle\}$ has only one accumulation point and we are done.

To show (4.10) we will show that

$$Y(x_k) = Y(x) + t_k z + \phi_k \qquad (4.11)$$

where

$$\{\phi_k/t_k\} \to 0 \in R^m \quad , \qquad (4.12)$$

and then, since $\bar{w}_k$, $\hat{w}_k \in \partial E(Y(x_k))$, we have, by the semismoothness of E, that $\{<\bar{w}_k,z>\}$ and $\{<\hat{w}_k,z>\}$ have the same limit, which implies (4.10), because $\bar{w}$ and $\hat{w}$ are accumulation points of $\{\bar{w}_k\}$ and $\{\hat{w}_k\}$, respectively.

For each $i = 1,2,\ldots,m$ let

$$\phi_k^i = f_i(x_k) - f_i(x) - t_k f_i'(x;d) \quad ,$$

so that (4.11) is satisfied with $\phi_k = (\phi_k^1, \phi_k^2, \ldots, \phi_k^m)$ and

$$\phi_k^i/t_k = [f_i(x_k) - f_i(x)]/t_k - f_i'(x;d) \quad . \qquad (4.13)$$

Note that, by using the definition of x_k and adding and subtracting $f_i(x+t_k d)$, we have

$$[f_i(x_k) - f_i(x)]/t_k = [f_i(x+t_k d+\theta_k) - f_i(x+t_k d)]/t_k + [f_i(x+t_k d) - f_i(x)]/t_k \quad . \qquad (4.14)$$

As $k\to\infty$, the first term of the right-hand side of (4.14) converges to zero, because each f_i is Lipschitz and $\{\theta_k/t_k\} \to 0 \in R^n$. The second term converges to $f_i'(x;d)$, so we have that

$$\{[f_i(x_k) - f_i(x)]/t_k\} \to f_i'(x;d) \quad ,$$

which, by (4.13), implies (4.12) and completes the proof.□

5. STATIONARITY AND OPTIMALITY

Consider the following problem that is equivalent to the optimization problem of section 1:

$$\text{minimize } f(x)$$
$$\text{subject to } h(x) \leqq 0$$

where

$$h(x) = \max_{1 \leqq i \leqq m} h_i(x) \qquad \text{for } x \in R^n \quad .$$

We say that $x \in R^n$ is *feasible* if $h(x) \leqq 0$ and *strictly feasible* if $h(x) < 0$. We say that $\bar{x} \in R^n$ is *optimal* if $\bar{x}$ is feasible and $f(\bar{x}) \leqq f(x)$ for all feasible x.

Let X be a subset of R^n and for each $x \in R^n$ let

$$A(x) = \{i \in \{1,2,\ldots,m\} : h(x) = h_i(x)\} \quad .$$

Then, from Theorems 4,5,1 and 3, we have the following:

Theorem 6. *Suppose* $h_1, h_2, \ldots, h_m$ *are locally Lipschitz. Then*

(a) h *is locally Lipschitz and for each* $x \in R^n$

$$\partial h(x) \subset \text{conv}\{\partial h_i(x) : i \in A(x)\} \quad .$$

(b) *If* $h_1, h_2, \ldots, h_m$ *are semismooth on* X *then* h *is semismooth on* X.

(c) *If* $h_1, h_2, \ldots, h_m$ *are semiconvex (quasidifferentiable) on* X *then* h *is semiconvex (quasidifferentiable) on* X *and for each* $x \in R^n$

$$\partial h(x) = \text{conv}\{\partial h_i(x) : i \in A(x)\} \quad .$$

A key idea for dealing with the above optimization problem is to define the point-to-set map $M : R^n \to 2^{R^n}$ by

$$M(x) = \begin{cases} \partial f(x) & \text{if } h(x) < 0 \\ \text{conv}\,\{\partial f(x) \cup \partial h(x)\} & \text{if } h(x) = 0 \\ \partial h(x) & \text{if } h(x) > 0 \end{cases} \quad \text{for } x \in R^n \quad .$$

This map was introduced and used by Merrill [10, Chapter 12] for problems with differentiable and/or convex functions, i.e. problems with functions having gradients and/or subgradients. It is used by our algorithm in [11] for problems with functions having generalized gradients.

We say that $\bar{x} \in R^n$ is *stationary* for the optimization problem if $h(\bar{x}) \leqq 0$ and $0 \in M(\bar{x})$. Our algorithm in [11] is shown to converge to stationary points for problems with semismooth functions. The next result shows that stationarity is necessary for optimality. It follows from a very general theorem in Clarke [2]. Here we give an independent proof using a strict separation theorem for convex sets.

Theorem 7. Suppose f *and* h *are locally Lipschitz. If* $\bar{x}$ *is optimal then* $\bar{x}$ *is stationary.*

Proof: Consider the case where $h(\bar{x}) = 0$. Suppose, for contradiction purposes, that $\bar{x}$ is not stationary. Then $0 \notin M(\bar{x})$. Since $\partial f(\bar{x})$ and $\partial h(\bar{x})$ are compact, $M(\bar{x})$ is closed and convex and, thus, from a strict separation theorem [9, Cor. 3.2.4], there exists a $d \in R^n$ such that

$$\langle g,d\rangle < 0 \text{ for all } g \in M(\bar{x}) \quad . \tag{5.1}$$

Since $\bar{x}$ is optimal, it must be the case that either $f^0(\bar{x};d) \geqq 0$ or $h^0(\bar{x};d) \geqq 0$, for if not, we can find a $t > 0$ such that $f(\bar{x}+td) < f(\bar{x})$ and $h(\bar{x}+td) < h(\bar{x}) = 0$, which contradicts the optimality of $\bar{x}$. Thus, by Proposition 1, there is a $\bar{g} \in (\partial f(\bar{x}) \cup \partial h(\bar{x})) \subset M(\bar{x})$ such that $\langle\bar{g},d\rangle \geqq 0$. But this contradicts (5.1). So $0 \in M(\bar{x})$. We omit the proof of the case where $h(\bar{x}) < 0$ which is similar, but simpler.□

Remark: This theorem, when specialized, gives two well-known necessary optimality theorems. If $h_1, h_2, \ldots, h_m$ and f are differentiable then the above result combined with part (a) of Theorem 6 shows that an optimal $\bar{x}$ solves the Karush [5]-John [4] stationary point problem [9, p. 93]. Alternatively, if $h_1, h_2, \ldots, h_m$ and f are convex then Theorems 6 and 7 and Proposition 3 show that an optimal $\bar{x}$ solves the corresponding saddle-point problem [9, p. 71].

As usual, in order to have stationarity be sufficient for optimality, we need stronger assumptions on the problem functions. We now proceed to show that if the problem functions are semiconvex and there is a strictly feasible point then stationarity implies optimality. In order to demonstrate this we require the following preliminary result for semiconvex functions on convex sets:

Theorem 8. If F *is semiconvex on a convex set* $X \subset R^n$, $x \in X$ *and* $x + d \in X$ *then*

$$F(x+d) \leqq F(x) \text{ implies } F'(x;d) \leqq 0 \quad .$$

Proof: Suppose, for contradiction purposes, $F(x+d) \leqq F(x)$ and $F'(x;d) > 0$. Then there exists $t > 0$ such that $t < 1$ and $F(x+td) > F(x)$ Let $\bar{t} \in (0,1)$ maximize the continuous function $a(t) = F(x+td)$ over $t \in [0,1]$. Clearly, by the maximality of $\bar{t}$,

$$a(1) = F(x+d) \leqq F(x) = a(0) < a(\bar{t}) = F(x+\bar{t}d) \quad , \tag{5.2}$$

$$F'(x+\bar{t}d;d) \leqq 0 \text{ and } F'(x+\bar{t}d;-d) \leqq 0 \quad .$$

Now by the quasidifferentiability of F there exist $g^+ \in \partial F(x+\bar{t}d)$ and $g^- \in \partial F(x+\bar{t}d)$ such that

$$0 \geqq F'(x+\bar{t}d;d) = F^0(x+\bar{t}d;d) = \langle g^+,d\rangle \geqq \langle g^-,d\rangle$$

and

$$0 \geqq F'(x+\bar{t}d;-d) = F^0(x+\bar{t}d;-d) = \langle g^-,-d\rangle \geqq \langle g^+,-d\rangle \quad .$$

So,

$$F'(x+\bar{t}d;d) = 0 \quad ,$$

and, by the positive homogeneity of $F'(x+td;\cdot)$, since $1-\bar{t}>0$, we have

$$F'(x+\bar{t}d;(1-\bar{t})d) = (1-\bar{t})F'(x+\bar{t}d;d) = 0 \quad .$$

Then the semiconvexity of F implies

$$F(x+d) \geqq F(x+\bar{t}d)$$

which contradicts (5.2).□

Remark: The above proof follows one in Mangasarian [9, pp. 143-144] and a slight modification shows that a semiconvex function on a convex set is "strictly quasiconvex" and, hence, "quasiconvex" [9, Ch. 9].

Theorem 9. Suppose f *and* h *are semiconvex on* R^n *and* $\bar{x} \in R^n$ *is such that* $0 \in M(\bar{x})$.

(a) If $h(\bar{x}) > 0$ *then* $h(x) \geqq h(\bar{x}) > 0$ *for all* $x \in R^n$, *i.e. the optimization problem has no feasible points.*

(b) If $h(\bar{x}) \leqq 0$ *then at least one of the following holds:*

(i) $\bar{x}$ *is optimal*

(ii) $h(x) \geqq 0$ *for all* $x \in R^n$, *i.e. the optimization problem has no strictly feasible points.*

Proof: If $h(\bar{x}) > 0$ then $0 \in \partial h(\bar{x})$ and it is clear from the semiconvexity of h that $\bar{x}$ minimizes h over R^n and the desired result (a) follows. If $h(\bar{x}) < 0$ then $0 \in \partial f(\bar{x})$ and similar reasoning shows that $\bar{x}$ minimizes f over R^n which implies b(i). Suppose $h(\bar{x}) = 0$. Then there exist $\lambda \in [0,1]$, $\bar{g} \in \partial f(\bar{x})$ and $\hat{g} \in \partial h(\bar{x})$ such that

$$\lambda\bar{g} + (1-\lambda)\hat{g} = 0 \quad .$$

If $\lambda = 0$, then $\hat{g} = 0$, $\bar{x}$ minimizes h over R^n and b(ii) holds. Alternatively, if $\lambda > 0$ then

$$\bar{g} + [(1-\lambda)/\lambda]\hat{g} = 0 \quad ,$$

and for all $x \varepsilon R^n$

$$\langle\bar{g},x-\bar{x}\rangle + [(1-\lambda)/\lambda]\langle\hat{g},x-\bar{x}\rangle = 0 \quad .$$

For all $x \varepsilon R^n$ such that $h(x) \leqq 0 = h(\bar{x})$, we have, by the semiconvexity of h, Theorem 8 and the fact that $\hat{g} \varepsilon \partial h(\bar{x})$, that

$$0 \geqq h'(\bar{x};x-\bar{x}) = h^0(\bar{x},x-\bar{x}) \geqq \langle\hat{g},x-\bar{x}\rangle \quad .$$

Thus, since $[(1-\lambda)/\lambda] \geqq 0$, we have that

$$\langle\bar{g},x-\bar{x}\rangle \geqq 0 \text{ for all } x \text{ such that } h(x) \leqq 0 \quad .$$

So, by the semiconvexity of f, since $\bar{g} \varepsilon \partial f(x)$, we have that

$$f'(\bar{x};x-\bar{x}) = f^0(\bar{x};x-\bar{x}) \geqq \langle\bar{g},x-\bar{x}\rangle \geqq 0$$

and, hence,

$$f(x) \geqq f(\bar{x}) \text{ for all } x \text{ such that } h(x) \leqq 0 \quad .$$

Thus, $\bar{x}$ is optimal and we have that $\lambda > 0$ implies that b(i) holds.□

Remark: If $h(\bar{x}) = 0$ and $\lambda > 0$ in the above proof then, in order to show optimality of $\bar{x}$, we need only assume that h is quasidifferentiable and satisfies the conclusion of Theorem 8 rather than assume h is semiconvex. This observation corresponds to a sufficient optimality theorem in Mangasarian [9, Theorem 10.1.1] and says that if $\bar{x}$ satisfies generalized Karush [5] - Kuhn-Tucker [6] conditions, f is semiconvex and h is quasidifferentiable and "quasiconvex" [9, Chapter 9] then $\bar{x}$ is optimal. A constraint qualification that implies $\lambda > 0$ is that $0 \notin \partial h(\bar{x})$.

6. ACKNOWLEDGEMENT

I wish to thank Claude Lemarechal for his many helpful suggestions.

The research was sponsored, in part, by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-74-2695. The United States Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation hereon.

REFERENCES

[1] Clarke, F.H., Generalized Gradients and Applications, *Trans. Amer. Math. Soc.*, 205 (1975), 247-262.

[2] Clarke, F.H., A New Approach to Lagrange Multipliers, *Mathematics of Operations Research*, 1 (1976), 165-174.

[3] Feuer, A., *An Implementable Mathematical Programming Algorithm for Admissible Fundamental Functions*, Ph.D. Dissertation, Department of Mathematics, Columbia University, New York, 1974.

[4] John, F., Extremum Problems with Inequalities as Subsidiary Conditions, in K.O. Friedrichs, O.E. Neugebauer and J.J. Stoker, eds., *Studies and Essays: Courant Anniversary Volume*, Interscience Publishers, New York, 1948, 187-204.

[5] Karush, W., *Minima of Functions of Several Variables with Inequalities as Side Conditions*, M.S. Dissertation, Department of Mathematics, University of Chicago, Chicago, Ill., 1939.

[6] Kuhn, H.W. and A.W. Tucker, Nonlinear Programming, in J. Neyman, ed., *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, Calif., 1951.

[7] Lebourg, G., Valeur moyenne pour gradient généralisé, *C.R. Acad. Sc. Paris*, 281 (1975), 795-797.

[8] Lemarechal, C., An Extension of Davidon Methods to Nondifferentiable Problems, in M.L. Balinski and P. Wolfe, eds., *Nondifferentiable Optimization*, Mathematical Programming Study 3, North-Holland, Amsterdam, 1975, 95-109.

[9] Mangasarian, O.L., *Nonlinear Programming*, McGraw-Hill, New York, 1969.

[10] Merrill, O.H., *Applications and Extensions of an Algorithm that Computes Fixed Points of Certain Upper Semicontinuous Point to Set Mappings*, Ph.D. Dissertation, University of Michigan, Ann Arbor, Mich., 1972.

[11] Mifflin, R., *An Algorithm for Constrained Optimization with Semismooth Functions*, International Institute for Applied Systems Analysis, Laxenburg, Austria, forthcoming.

[12] Pshenichnyi, B.N., *Necessary Conditions for an Extremum*, Marcel Dekker, New York, 1971.

[13] Rockafellar, R.T., *Convex Analysis*, Princeton University Press, Princeton, N.J., 1970.

[14] Shor, N.Z., A Class of Almost-Differentiable Functions and a Minimization Method for Functions of this Class, *Cybernetics*, July (1974), 599-606; *Kibernetika*, 4 (1972), 65-70.

[15] Tuy, Hoáng, Sur les inégalites linéaires, *Colloquium Mathematicum*, 13 (1964), 107-123.

[16] Wolfe, P., A Method of Conjugate Subgradients for Minimizing Nondifferentiable Functions, in M.L. Balinski and P. Wolfe, eds., *Nondifferentiable Optimization*, Mathematical Programming Study 3, North-Holland, Amsterdam 1975, 145-173.