

The Michel-Penot Subdifferential and Stochastic Programming

John R. Birge
Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, MI 48109

Liqun Qi
School of Mathematics
The University of New South Wales
Kensington, N. S. W. 2033
AUSTRALIA

Technical Report 89-23

July 1989

This report is also issued as
Applied Mathematics Preprint 89/12
School of Mathematics
The University of New South Wales
Kensington, N. S. W.
AUSTRALIA

THE MICHEL-PENOT SUBDIFFERENTIAL AND STOCHASTIC PROGRAMMING

John R. Birge*

Department of Industrial and Operations Engineering

University of Michigan

Ann Arbor, MI 48109

and

Liqun Qi

School of Mathematics

The University of New South Wales

Kensington, N.S.W. 2033

Australia

Abstract

The Michel-Penot subdifferential of a locally Lipschitzian function is the principal part of the Clarke subdifferential. It does not contain extraneous “subgradients” at abnormal points. A locally Lipschitzian function can be determined by its Michel-Penot subdifferential uniquely up to an additive constant though this cannot be done by its Clarke subdifferential if the set of abnormal points is not negligible. A set-valued operator is the Michel-Penot subdifferential of a locally Lipschitzian function if and only if it is a maximal seminormal operator satisfying a cyclical condition. Various calculus rules hold for the Michel-Penot subdifferential. Equalities hold for these rules at a point under semiregularity, a weaker condition than regularity, and almost everywhere under conditions on the space alone. Applications in stochastic programming are discussed.

Section 1. Introduction

The Clarke subdifferential theory is a breakthrough in nonsmooth analysis. It extends the subdifferential theory of convex functions to nonconvex functions and has found applications in several fields [5]. However, at some abnormal points of certain nonsmooth functions, the Clarke subdifferential of this function may include some “extraneous subgradients”. These “extraneous subgradients” may be further expanded through calculus. These abnormal phenomena are not so important for “not seriously” nonsmooth func-

* Supported by Office of Naval Research Grant N0014-86-K-0628.

tions, such as semismooth functions and regular functions. However, for some “seriously” nonsmooth functions, this abnormality cannot be ignored. These “seriously” nonsmooth functions can even be Lipschitzian functions. The following examples demonstrate these abnormalities:

- 1.1. Even at a point where the G-derivative exists for a Lipschitzian function, the Clarke subdifferential of the function at this point may not be singleton. For example, let $f : \mathfrak{R} \rightarrow \mathfrak{R}$ be defined by $f(x) = x^2 \sin(1/x)$ for $x \neq 0$ and $f(0) = 0$. Then f is a globally Lipschitzian and differentiable function with $f'(x) = 2x \sin(1/x) - \cos(1/x)$ for $x \neq 0$ and $f'(0) = 0$. However, the Clarke subdifferential of f at 0 is $[-1, 1]$. Thus, the Clarke subdifferential includes some extra “subgradients” at this point. When the measure of the set of such abnormal points is not zero, these “extraneous subgradients” may overgrow to cover the real derivatives.

- 1.2. Such abnormalities make the characterization problem of the Clarke subdifferential very difficult in general, if not impossible. In fact, such a “seriously” nonsmooth function cannot be determined uniquely by its Clarke subdifferential up to an additive constant. For example [20], consider $f : \mathfrak{R} \rightarrow \mathfrak{R}$. Let S be a measurable subset of \mathfrak{R} with the property that for every nonempty open interval I , the intersection of I and S , and the intersection of I and $\mathfrak{R} \setminus S$ are both of positive measure. Define $g(x) = 1$ if $x \in S$ and $g(x) = -1$ if $x \in \mathfrak{R} \setminus S$. Let $f(x) = \int_0^1 g(t) dt$. Then f is globally Lipschitzian on \mathfrak{R} (with Lipschitzian constant 1) and in particular f is differentiable almost everywhere: $f'(x) = g(x)$ for almost every x . From the construction of S , however, we know that the Clarke subdifferential $\partial^\circ f(x) = [-1, 1]$ for every $x \in \mathfrak{R}$. If we let $f_t(x) = f(x + t)$, then we still have the same Clarke subdifferential. It cannot be true in general that $f_t = f + \text{const}$. Thus, it is not generally true that a function can be determined by its Clarke subdifferential uniquely up to an additive constant even in the one-dimensional Lipschitzian case.

- 1.3. The calculus of the Clarke subdifferential only has inclusion relationships. Under some regularity conditions, these inclusion relationships become equalities. One approach to treat these abnormalities is to assume that the set of the abnormal points is negligible for the function under consideration, i.e., the Clarke subdifferential of this function is singleton almost everywhere. In [16][17], Qi studied the characterization problem of

the Clarke subdifferential with this assumption. He characterized the Clarke subdifferential of such functions as cyclically maximal normal operators [16] and proved that semismooth functions satisfy this assumption [17]. One may argue that this assumption is artificial and thus trivializes the problem. However, according to a generalized version of the Rademacher Theorem [4], a locally Lipschitzian function is differentiable almost everywhere in a separable Banach space under the Haar measure. Thus, if we can exclude those “extraneous subgradients” from the Clarke subdifferential, this assumption would hold for all the locally Lipschitzian functions.

Several authors have proposed alternative definitions to remove these extraneous subgradients (see Treiman [21]). Michel and Penot [9][10] gave such a definition of subdifferential. The Michel-Penot subdifferential of a locally Lipschitzian function at a point is a nonempty, convex, weak*-compact set, contained in the Clarke subdifferential of the function at this point. It is singleton whenever the G-derivative of this function exists at this point. Thus, no “extraneous subgradients” are included in their definition.

The Michel-Penot subdifferential forms the principal part of the Clarke subdifferential and is the same as the Clarke subdifferential at normal points. It is also the same as Treiman’s B-subdifferential in finite dimensions and was also defined by Caligaris and Oliva [3] in 1986. We follow Borwein, Fitzpatrick and Giles [2] and Ward [22] in calling this subdifferential, the Michel-Penot subdifferential. We summarize the definition and basic properties of the Michel-Penot subdifferential in Section 2.

In Section 3, we prove various calculus rules for the Michel-Penot subdifferential. They include two chain rules, pointwise maxima, and basic calculus. We show that equalities hold for these rules under semiregularity, which is a condition weaker than regularity. In fact, a function is regular at a point if and only if it is semiregular and normal at this point. We also discuss some applications of these rules to the problem of minimizing the maximum eigenvalue of a symmetric matrix.

We study in Section 4 the characterization problem for the Michel-Penot subdifferential. We prove that a locally Lipschitzian function is determined uniquely by its Michel-Penot subdifferential up to an additive constant in a separable Banach space. We define the maximal seminormal operator there. The Michel-Penot subdifferentials of locally Lipschitzian functions and maximal monotone operators in finite-dimensional spaces are exam-

ples of maximal seminormal operators. A maximal seminormal operator is the Michel-Penot subdifferential of a locally Lipschitzian function if and only if it is cyclical. When a maximal seminormal operator is weak*-upper semicontinuous, it is a maximal normal operator.

One of our motivations for this study is stochastic programming. One major method for solving stochastic programs is the subgradient method [7][15]. However, the objective function of a stochastic program generally is a multi-integral of several variables [1]. The number of variables ranges from 5 to 100 [15]. The “extraneous subgradients” may cause catastrophic effects to the computation. We study ways to avoid such “extraneous subgradients” by using the Michel-Penot subdifferential in integral functionals in Section 5. We first establish an inclusion relationship for the Michel-Penot subdifferential of integral functionals. Again, this inclusion relationship is an equality relationship under semiregularity. Furthermore, by the generalized Rademacher Theorem, equality holds almost everywhere in finite dimensions and, in Haar measure, in separable Banach spaces. (This is not true in general for the Clarke subdifferential.) We may estimate the subdifferentials at other points using this information.

In Section 6 we study applications of the above discussions in stochastic programming. In particular, we establish general optimality conditions and propose a framework for the use of subgradient methods for finding stationary points of Lipschitzian functions.

Section 2. The Michel-Penot Subdifferential

Suppose X is a Banach space and Y is an open subset of X . Consider a locally Lipschitzian function $f : Y \rightarrow \mathfrak{R}$. The Michel-Penot directional derivative of f at $x \in Y$ in the direction $h \in X$, is

$$f^\circ(x; h) := \sup_{k \in X} \{ \limsup_{t \downarrow 0} [f(x + th + tk) - f(x + tk)]/t \}.$$

The Michel-Penot subdifferential of f at x is the set

$$\partial^\circ f(x) := \{ u \in X^* : \langle u, h \rangle \leq f^\circ(x; h), \forall h \in X \}.$$

Most of the followings properties of the Michel-Penot directional derivative and the Michel-Penot subdifferential were given in [9] and [10]. The others are obvious.

$$2.1. f^-(x; h) \leq f^+(x; h) \leq f^\circ(x; h) \leq f^\circ(x; h),$$

where $f^-(x; h)$ is the lower directional derivative:

$$f^-(x; h) := \liminf_{t \downarrow 0} [f(x + th) - f(x)]/t;$$

$f^+(x; h)$ is the upper directional derivative:

$$f^+(x; h) := \limsup_{t \downarrow 0} [f(x + th) - f(x)]/t;$$

$f^\circ(x; h)$ is the Clarke directional derivative,

$$f^\circ(x; h) := \limsup_{y \rightarrow x, t \downarrow 0} [f(y + th) - f(y)]/t.$$

When $f^-(x; h) = f^+(x; h)$, then the usual directional derivative

$$f'(x; h) := \lim_{t \downarrow 0} [f(x + th) - f(x)]/t$$

exists and $f'(x; h) = f^-(x; h) = f^+(x; h)$.

2.2. The mapping $f^\circ(x; \cdot)$ is sublinear; $f^\circ(x; -h) = (-f)^\circ(x; h)$; $\partial^\circ f(x)$ is a nonempty, convex, weak*-compact set; $\partial^\circ f$ is locally bounded in Y ; for any real number s , $\partial^\circ(sf) = s\partial^\circ f$.

$$2.3. f^\circ(x; h) = \max\{\langle u, h \rangle : u \in \partial^\circ f(x)\}.$$

$$2.4. \sup_{k \in X} \{f^+(x; h+k) - f^+(x; k)\} \leq f^\circ(x; h) \leq \sup_{k \in X} \{f^+(x; h+k) - f^-(x; k)\}.$$

In particular if $f'(x; k)$ exists for all k at x , then

$$f^\circ(x; h) = \sup_{k \in X} f'(x; h+k) - f'(x; k).$$

Function f is G-differentiable at x if and only if $\partial^\circ f$ is singleton at x . In this case,

$$\partial^\circ f(x) = \{f'(x)\}.$$

2.5. If f is a convex function, then $\partial^\circ f = \partial f$, the subdifferential of f in the sense of convex analysis.

2.6. $f^\circ(x; h) = \limsup_{y \rightarrow x} f^\circ(y; h)$; $\partial^\circ f(x) \subset \partial f(x)$; $\partial^\circ f(x) = \partial f(x)$ if and only if $\partial^\circ f$ is weak*-upper semicontinuous at x .

2.7. Suppose that $\{f_i : i \in S\}$ is a collection of locally Lipschitzian functions on Y . Let $f : Y \rightarrow \mathfrak{R}$ be defined by

$$f(y) = \sup\{f_i(y) : i \in S\}.$$

Then for any x in Y ,

$$\partial^\circ f(x) \subset \text{weak}^* - \text{cl conv}\{\partial^\circ f_i(x) : i \in S(x)\},$$

where $S(x) = \{i : f_i(x) = f(x)\}$.

2.8. If f attains a local minimum or maximum at x , then $0 \in \partial^\circ f(x)$.

For example, let $f : \mathfrak{R} \rightarrow \mathfrak{R}$ be defined by $f(x) = x^2 \sin(1/x) + 0.5x$ when x is not 0 and $f(0) = 0$. Then $\partial^\circ f(0) = 0.5$. Thus, 0 is not an optimum point. On the other hand, $\partial^\circ f(0) = [-0.5, 1.5]$. We cannot draw a conclusion about the optimality from $\partial^\circ f(0)$.

Section 3. Semiregularity and Calculus Rules

Suppose that X, Y and f are the same as in Section 2. Let x be a point in Y . We say that f is *normal* at x if $\partial^\circ f(x) = \partial^\circ f(x)$. We say that f is *semiregular* at x provided for all h , the usual directional derivative $f'(x; h)$ exists and $f'(x; h) = f^\circ(x; h)$.

Proposition 3.1. *Function f is regular at x if and only if f is normal and semiregular at x . Function f is strictly differentiable at x if and only if f is normal and G-differentiable at x .*

Proof: By 2.3, f is normal at x if and only if for any h , $f^\circ(x; h) = f^\circ(x; h)$. Recall that f is said to be regular at x provided that for all h , $f'(x; h)$ exists and $f'(x; h) = f^\circ(x; h)$ [5]. By 2.1, $f'(x; h) \leq f^\circ(x; h) \leq f^\circ(x; h)$. The first conclusion follows. By Proposition 2.2.4 of [5], f is strictly differentiable at x if and only if $\partial^\circ f(x)$ is singleton. However, $\partial^\circ f(x)$ is singleton if and only if $\partial^\circ f(x)$ is singleton and f is normal at x . By 2.4, $\partial^\circ f(x)$ is singleton if and only if f is G-differentiable at x . The second conclusion follows. ■

Therefore, semiregularity is a property weaker than regularity. By 2.4, if f is G-differentiable at x , then f is semiregular at x . According to the generalized Rademacher Theorem [4], a locally Lipschitzian function is differentiable almost everywhere in a finite dimensional space and, in Haar measure, in a separable Banach space. Recently, Preiss [14] showed that if X is a Banach space with an equivalent norm G-(F-) differentiable

away from the origin, then every locally Lipschitzian function defined on an open subset Y of X is G-(F-) differentiable on a dense subset of Y . These show that semiregularity holds for locally Lipschitzian functions either almost everywhere or in a dense subset under conditions on the space alone. On the other hand a locally Lipschitzian function need not be regular almost everywhere as Example 1.2 shows. Certainly, if f is convex, then f is semiregular at x , since a convex function is regular at any point.

The following Mean-Value Theorem for the Michel-Penot subdifferential was proved in [2]. Since the Michel-Penot subdifferential of f at a point is contained, and sometimes properly contained, in its Clarke subdifferential at this point, this theorem is stronger than the corresponding version for the Clarke subdifferential, the Lebourg Theorem (Theorem 2.3.7 of [5]). This again implies that those “subgradients” in the Clarke subdifferential but not in the Michel-Penot subdifferential are extraneous in a certain sense.

Theorem 3.2 (Mean-Value Theorem) . (Borwein, Fitzpatrick and Giles) *Suppose that $[x, y]$ is a line segment in Y . Then there exists a point z in (x, y) such that*

$$f(y) - f(x) \in \langle \partial^\circ f(z), y - x \rangle .$$

■

We now discuss the chain rule for the Michel-Penot subdifferential. Suppose that $f : Y \rightarrow \mathbb{R}^n$ is given by $f(y) = g(p(y))$, where $p : Y \rightarrow \mathbb{R}^n$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ are locally Lipschitzian. Then f is also locally Lipschitzian. Denote the components of p by $p_i, i = 1, 2, \dots, n$. Assume that g is normal at $p(x)$, where x is a point in Y . The inclusion relationship part of the following chain rule is similar to the corresponding part of Theorem 2.3.9 of [5] for the Clarke subdifferential but the proofs are not all the same. The difference is that in general $\partial^\circ p$ is not weak*-upper semicontinuous at x . Thus, we cannot invoke here a lemma as the one in the proof of Theorem 2.3.9 of [5]. The equality part of this chain rule is stronger than the corresponding part of Theorem 2.3.9 of [5] by only requiring semiregularity for each p_i in (i) and G-differentiability for p in (ii).

Theorem 3.3 (Chain Rule I). *One has*

$$\partial^\circ f(x) \text{ is contained in weak}^* - \text{cl conv } S, \tag{1}$$

where

$$S = \left\{ \sum_i a_i v_i : v_i \in \partial^\circ p_i(x), a \in \partial^\circ g(p(x)) \right\},$$

and equality holds under any one of the following additional hypotheses:

- (i) g is regular at $p(x)$, each $p_i(x)$ is semiregular at x , and every element of $\partial^\circ g(p(x))$ has nonnegative components. In this case, it follows that f is semiregular at x .
- (ii) g is semiregular at $p(x)$ and p is G -differentiable at x . In this case it follows that f is regular at x , and the weak*-cl conv is superfluous.

Proof: The set S is weak*-compact, and therefore its convex hull has compact closure.

The support function of S evaluated at a point h in X is

$$r(h) := \max \left\{ \sum_i a_i \langle v_i, h \rangle : v_i \in \partial^\circ p_i(x), a \in \partial^\circ g(p(x)) \right\}. \quad (2)$$

By 2.3, it suffices to prove that $r(h)$ majorizes $f^\circ(x; h)$ for any h . Applying the Mean-Value Theorem on g , for any h and k in X and small positive t , we have

$$\begin{aligned} [f(x + th + tk) - f(x + tk)]/t &= [g(p(x + th + tk)) - g(p(x + tk))]/t \\ &= \sum a_i(z) [p_i(x + th + tk) - p_i(x + tk)]/t, \end{aligned} \quad (3)$$

where $a(z) \in \partial^\circ g(z)$, $z \in [p(x + tk), p(x + th + tk)]$. Fix h and k in X . On one hand,

$$\begin{aligned} \max \{ \langle v_i, h \rangle : v_i \in \partial^\circ p_i(x) \} &= p_i^\circ(x; h) \\ &\geq \limsup_{t \downarrow 0} \frac{[p_i(x + th + tk) - p_i(x + tk)]}{t}. \end{aligned}$$

On the other hand,

$$\begin{aligned} \min \{ \langle v_i, h \rangle : v_i \in \partial^\circ p_i(x) \} &= - \max \{ \langle v_i, -h \rangle : v_i \in \partial^\circ p_i(x) \} = -p_i^\circ(x; -h) \\ &\leq - \limsup_{t \downarrow 0} \frac{[p_i(x - th + tk') - p_i(x + tk')]}{t}, \\ &\quad \text{where } k' := k + h \\ &= - \limsup_{t \downarrow 0} \frac{[p_i(x + tk) - p_i(x + th + tk)]}{t} \\ &= \liminf_{t \downarrow 0} \frac{[p_i(x + th + tk) - p_i(x + tk)]}{t}. \end{aligned}$$

Thus, for $\epsilon > 0$, there exists $\delta > 0$ such that $\forall t \in (0, \delta]$,

$$[p_i(x + th + tk) - p_i(x + tk)]/t \in \langle \partial^\circ p_i(x), h \rangle + [-\epsilon, \epsilon]. \quad (4)$$

Since g is normal at $p(x)$, we may choose δ small enough such that for all $t \in (0, \delta]$,

$$a(z) \in \partial g^\circ(p(x)) + [-\epsilon, \epsilon]^n. \quad (5)$$

Combining (2), (3), (4) and (5), we have

$$[f(x + th + tk) - f(x + tk)]/t \leq r(h) + n(L + L_i \|h\| + \epsilon)\epsilon,$$

for any $t \in (0, \delta]$ where L is the local bound of $\partial^\circ g$ near $p(x)$, L_i are the local bounds of $\partial^\circ p_i$ near x . Therefore, for any h in X ,

$$\begin{aligned} f^\circ(x; h) &= \sup_{k \in X} \{ \limsup_{t \downarrow 0} [f(x + th + tk) - f(x + tk)]/t \} \\ &\leq r(h). \end{aligned}$$

This proves (1). The proof for the equality under condition (i) is almost identical with the proof of the corresponding part of the proof of Theorem 2.3.9 of [5] except we use semiregularity and the Michel-Penot subdifferential for p_i and f instead of regularity and the Clarke subdifferential there.

When p is G-differentiable at x , each p_i is semiregular at x by 2.4. Thus, condition (ii) is a special case of (i). In this case, each $\partial^\circ p_i(x)$ is single-valued. Therefore, S is already convex and weak*-closed. The weak*-cl conv is hence superfluous in this case. This completes the proof. ■

Remark 3.4 If g is not normal at $p(x)$, we may use $\partial^\circ g(p(x))$ to replace $\partial^\circ g(p(x))$ in the expression of S . The claims still hold. ■

Instead of assuming that g is normal at x , we assume now that g is G-differentiable at $p(x)$. Notice that for locally Lipschitzian functions in a finite dimensional space, G-differentiability is the same as F-differentiability. The first two examples in the introduction show that G-differentiability in general does not guarantee normality.

Theorem 3.5 (Chain Rule II). *In this case, one has*

$$\partial^\circ f(x) \subset S := \left\{ \sum a_i v_i : v_i \in \partial^\circ p_i(x), a = g'(p(x)) \right\}, \quad (6)$$

and the equality holds if each $p_i(x)$ is semiregular at x , and $g'(p(x))$ has nonnegative components. In this case, it follows that f is semiregular at x .

Proof: We only point out the difference between this proof and the proof of Theorem 3.3. Instead of (3), we have for fixed h and k ,

$$\begin{aligned} [f(x+th+tk) - f(x+tk)]/t &= [g(p(x+th+tk)) - g(p(x+tk))]/t \\ &= [\langle g'(p(x)), p(x+th+tk) - p(x+th) \rangle \\ &\quad + o(\|p(x+th+tk) - p(x+tk)\|)]/t \\ &= \sum a_i [p_i(x+th+tk) - p_i(x+th)]/t + o(1). \end{aligned} \quad (7)$$

Combining (7) and (4), we obtain the conclusion. In this case, S is already convex and weak*-closed. Hence, we do not need to take the weak*-closure of the convex hull of S in (6). The proof for equality is similar to the corresponding part of the proof of Theorem 2.3.9 of [5]. ■

Remark 3.6. Michel and Penot [9] also proved a chain rule which requires that p is differentiable at x . Our chain rules are more useful to derive the following calculus rules. ■

Proposition 3.7 (Pointwise Maxima). *Suppose that $\{f_i : i = 1, \dots, n\}$ is a finite collection of locally Lipschitzian functions on Y . Let $f : Y \rightarrow \mathfrak{R}$ be defined by*

$$f(y) = \max\{f_i(y) : i = 1, \dots, n\}.$$

Then for any x in Y ,

$$\partial^\circ f(x) \subset \text{conv}\{\partial^\circ f_i(x) : i \in I(x)\},$$

where $I(x) = \{i : f_i(x) = f(x)\}$. If f is semiregular at x for each i in $I(x)$, then equality holds and f is semiregular at x .

Proof: The arguments are similar to those for Proposition 2.3.12 of [5] except we use Theorem 3.3 here. ■

Remark 3.8. The inclusion relationship is a special case of 2.7.

Proposition 3.9 (Basic Calculus). *Suppose that f and g are two locally Lipschitzian functions on Y . Then*

$$\partial^\circ(f + g)(x) \subset \partial^\circ f(x) + \partial^\circ g(x);$$

$$\partial^\circ(f \cdot g)(x) \subset g(x)\partial^\circ f(x) + f(x)\partial^\circ g(x);$$

$$\partial^\circ(f/g)(x) \subset [g(x)\partial^\circ f(x) - f(x)\partial^\circ g(x)]/[g(x)]^2, \text{ if } g(x) \neq 0.$$

Furthermore, if f and g are semiregular at x , then equality holds for the first formula and $f + g$ is semiregular at x ; if f and g are semiregular at x , and $f(x) \geq 0$, $g(x) \geq 0$, then equality holds for the second formula and fg is semiregular at x ; if f and g are semiregular at x , and $f(x) \geq 0$, $g(x) > 0$, then equality holds for the third formula and f/g is semiregular at x .

Proof: The arguments are similar to those for Propositions 2.3.13 and 2.3.14 of [5] except we use Theorem 3.3 here. ■

The following shows that equality can be obtained in Proposition 3.9 almost everywhere in certain spaces.

Proposition 3.10. *Suppose Y is an open subset of X , a separable Banach space. Then, equality holds for the formulas in Proposition 3.9 almost everywhere in Y in the Haar measure.*

Proof: When X is a separable Banach space, a generalized version of the Rademacher Theorem under the Haar measure holds [4]. Thus, f and g are differentiable almost everywhere in the sense of the Haar measure. The right hand sides of each formula in Proposition 3.9 are, therefore, singleton almost everywhere in the Haar measure on Y , proving the result. ■

We now give some applications of the above discussions.

Theorem 3.11 (Maximum Eigenvalue). *Let $A(x)$ be a G -differentiable $n \times n$ symmetric matrix-valued function of a variable x in \mathfrak{R}^m . Let $f(x)$ be the maximum eigenvalue of $A(x)$. Then f is a locally Lipschitzian, semiregular function, and*

$$\partial^\circ f(x) = \text{conv}\{[\langle A'_1(x)u, u \rangle, \dots, \langle A'_m(x)u, u \rangle] : u \in S\},$$

where S is the set of unit eigenvectors belonging to the maximum eigenvalue of $A(x)$. If the maximum eigenvalue of $A(x)$ has multiplicity 1, then f is G -differentiable at x , and

$$f'(x) = [\langle A'_1(x)u, u \rangle, \dots, \langle A'_m(x)u, u \rangle],$$

where u is the unit eigenvector belonging to the maximum eigenvalue of $A(x)$.

Proof: Let g be a function mapping the $n \times n$ symmetric matrix space to \Re , such that $g(y)$ is the maximum eigenvalue of y . Then g is a convex function [12] and

$$\partial g(y) = \text{conv}\{u^T u : u \in S(y)\},$$

where $S(y)$ is the set of unit eigenvectors belonging to the maximum eigenvalue of y . Combining this and Theorem 3.3, we get our result. ■

This theorem generalizes Proposition 2.8.8 of [5], which is valid when A is smooth. We thus can calculate the directional derivatives of f at x since they are the same as the Michel-Penot directional derivatives in this case. This is useful to find a decent direction of f at x if optimality has not been achieved [12] [13]. For more discussions on eigenvalues and singular values of nonsmooth matrix-valued functions, see [19].

Section 4. Characterization of the Michel-Penot Subdifferential

Suppose again that X is a separable Banach space, that Y is an open subset of X and that f is a locally Lipschitzian function on Y . Thus, f is differentiable almost everywhere in the sense of the Haar measure as given above. By 2.4, $\partial^\circ f(x)$ is equal to $\{f'(x)\}$ almost everywhere. Let $[x, y]$ be a line segment in Y and $x' \rightarrow x$, $y' = y + (x' - x)$. It follows from the Fubini Theorem that there are points x' approximating x with the property that $\partial^\circ f$ is single-valued almost everywhere on the line segment $[x', y']$. On $[x', y']$,

$$\int_0^1 \langle \partial^\circ f(x' + t(y - x)), y - x \rangle dt = f(y') - f(x').$$

Define

$$\begin{aligned} & \int_0^1 \langle \partial^\circ f(x + t(y - x)), y - x \rangle dt \\ &= \lim_{x' \rightarrow x} \int_0^1 \langle \partial^\circ f(x' + t(y - x)), y - x \rangle dt, \end{aligned}$$

where x' are those points such that $\partial^\circ f$ is single-valued almost everywhere on $[x', y']$. Then

$$\begin{aligned} \int_0^1 \langle \partial^\circ f(x + t(y - x)), y - x \rangle dt &= \lim_{x' \rightarrow x} [f(y') - f(x')] \\ &= f(y) - f(x). \end{aligned}$$

Suppose now that for any two points x and y in Y , there exist a finite number of line segments $[x_0 = x, x_1], [x_1, x_2], \dots, [x_{m-1}, x_m = y]$ in Y , connecting them. We call such a set Y *linearly connected*.

Theorem 4.1 (Unique Determination Theorem). *Under the above assumptions, f is determined by $\partial^\circ f$ uniquely up to an additive constant.*

Proof: Fix a point x in Y . For any y in Y , let $[x_0 = x, x_1], [x_1, x_2], \dots, [x_{m-1}, x_m = y]$ be a finite collection of line segments in Y , connecting x and y . Then for any such f with the same $\partial^\circ f$,

$$\begin{aligned} f(y) &= f(x) + \sum_{i=1}^m [f(x_i) - f(x_{i-1})] \\ &= f(x) + \sum_{i=1}^m \int_0^1 \langle \partial^\circ f(x_{i-1} + t(x_i - x_{i-1})), x_i - x_{i-1} \rangle dt, \end{aligned}$$

which is uniquely determined up to an additive constant, $f(x)$. ■

Corollary 4.2 . *For any cycle of line segments $[x_0 = x, x_1], [x_1, x_2], \dots, [x_m, x_{m+1} = x]$ in Y , $\sum_{i=0}^{m+1} \int_0^1 \langle \partial^\circ f(x_{i-1} + t(x_i - x_{i-1})), x_i - x_{i-1} \rangle dt = 0$. ■*

We now study the characterization problem for the Michel-Penot subdifferential, i.e., what kind of set-valued operators are the Michel-Penot subdifferentials of locally Lipschitzian functions? Suppose that X is a separable Banach space and that Y is a linearly connected open set in X . A set-valued operator $F : Y \rightarrow X^*$ is called a *maximal seminormal operator* if

- (i) F is single-valued almost everywhere in Y in the Haar measure;
- (ii) F is locally bounded in Y ;
- (iii) For any $x \in Y$, $F(x)$ is a nonempty, convex, weak*-compact set;
- (iv) For any $x \in Y$ and $h \in X$,

$$\max\{\langle u, h \rangle : u \in F(x)\} = \sup_{k \in X} \limsup_{t \downarrow 0} \int_0^1 \langle F(x + t(k + sh)), h \rangle ds, \quad (8)$$

where the integral applies to the intervals in which F is single-valued almost everywhere in the corresponding one-dimensional measure.

By Section 2 and Corollary 4.2, it is easy to see the following.

Proposition 4.3. *The Michel-Penot subdifferential of a locally Lipschitzian function defined on Y is a maximal seminormal operator. ■*

Another candidate is the maximal monotone operator.

Proposition 4.4. *Suppose that X is a finite dimensional space. A maximal monotone operator F defined on Y is a maximal seminormal operator.*

Proof: According to Mignot [11], we have (i). Condition (ii) is commonly known for maximal monotone operators. It suffices now to prove (iii). Let x be in Y , h be in X , $t > 0$ and $s \in (0, 1]$. Let $u \in F(x)$ and $v \in F(x + tsh)$. Then

$$\langle v - u, tsh \rangle \geq 0,$$

i.e.,

$$\langle u, h \rangle \leq \langle v, h \rangle.$$

Thus,

$$\langle u, h \rangle \leq \limsup_{t \downarrow 0} \int_0^1 \langle F(x + tsh), h \rangle ds,$$

Therefore, the left hand side of (8) is not greater than its right hand side. On the other hand, let x be in Y , h and k in X , $t > 0$ and $s \in [0, 1)$. Let $v \in F(x + t(k + sh))$ and $w \in F(x + tk + th)$. Then

$$\langle w - v, t(1 - s)h \rangle \geq 0,$$

i.e.,

$$\langle v, h \rangle \leq \langle w, h \rangle.$$

Thus,

$$\begin{aligned} \limsup_{t \downarrow 0} \int_0^1 \langle F(x + t(k + sh)), h \rangle ds &= \limsup_{t \downarrow 0} \langle w, h \rangle \\ &\leq \max\{\langle u, h \rangle : u \in F(x)\}. \end{aligned}$$

Therefore, the right hand side of (8) is not greater than its left hand side. Hence (8) holds. ■

Let $[x, y]$ be a line segment in Y and $x' \rightarrow x$, $y' = y + (x' - x)$. It follows from the Fubini Theorem that there are points x' approximating x with the property that F is single-valued almost everywhere on the line segment $[x', y']$. Define

$$\int_0^1 \langle F(x + t(y - x)), y - x \rangle dt = \limsup_{x' \rightarrow x} \int_0^1 \langle F(x' + t(y - x)), y - x \rangle dt,$$

where x' are those points such that F is single-valued almost everywhere on $[x', y']$. This quantity is finite because of the local boundedness of F . We call F *cyclical* if for any cycle of line segments $[x_0 = x, x_1], [x_1, x_2], \dots, [x_m, x_{m+1} = x]$ in Y ,

$$\sum_{i=0}^{m+1} \int_0^1 \langle F(x_{i-1} + t(x_i - x_{i-1})), x_i - x_{i-1} \rangle dt = 0.$$

This property characterizes the Michel-Penot subdifferential of a locally Lipschitzian function in Y .

Theorem 4.5 (Characterization Theorem). *A maximal seminormal operator F , defined on Y , is the Michel-Penot subdifferential if and only if it is cyclical.*

Proof: The necessity of this condition is due to Corollary 4.2. We now prove the sufficiency of this condition. Suppose that F is cyclical. Fix a point x in Y . For any $y \in Y$, let $[x_0 = x, x_1], [x_1, x_2], \dots, [x_m, x_{m+1} = x]$ be a finite collection of line segments in Y , connecting x and y . Define $f : Y \rightarrow \mathfrak{R}$ by

$$f(y) = \sum_{i=1}^m \int_0^1 \langle F(x_{i-1} + t(x_i - x_{i-1})), x_i - x_{i-1} \rangle dt,$$

for any y in Y . Since F is cyclical, this definition is unique, no matter what collection of line segments we choose to connect x and y . By the local boundedness of F , we see that f is locally Lipschitzian. For fixed k and h , and very small t , we see that

$$\int_0^1 \langle F(y + t(k + sh)), h \rangle ds = [f(y + th + tk) - f(y + tk)]/t.$$

Thus by (8) and the definitions of the Michel-Penot directional derivative and the Michel-Penot subdifferential,

$$\max\{\langle u, h \rangle : u \in F(y)\} = f^\circ(y; h)$$

for any $y \in Y$ and any $h \in X$, and

$$\partial^\circ f(y) = F(y)$$

for any $y \in Y$. This completes the proof. ■

Theorem 4.5 is thus parallel to the characterization theorem of the subdifferential of convex functions. It is also parallel to the discussion of Qi [16] for the Clarke subdifferential of nonsmooth functions under the assumption mentioned in our introduction. If a set-valued operator F satisfies (i) and

(v) For any $x \in Y$,

$$F(x) = \text{weak}^* - \text{cl conv}\{\text{weak}^* - \lim_i u_i : u_i \in F(x_i), x_i \rightarrow x\}, \quad (9)$$

then F is called a *maximal normal operator* [16].

Proposition 4.6. *If a maximal seminormal operator F , defined on Y , is weak*-upper semicontinuous at a point x in Y , then (9) holds at this point.*

Proof: Denote the right hand side of (9) by C . For any $h \in X$, by (9),

$$\max\{\langle u, h \rangle : u \in C\} \leq \max\{\langle u, h \rangle : u \in F(x)\}.$$

On the other hand, by (8),

$$\begin{aligned} \max\{\langle u, h \rangle : u \in F(x)\} &= \sup_{k \in X} \{\limsup_{t \downarrow 0} \int_0^1 \langle F(x + t(k + sh)), h \rangle ds\} \\ &\leq \max\{\langle u, h \rangle : u \in C\}. \end{aligned}$$

Thus, $F(x) = C$, i.e., (9) holds. ■

Therefore, a maximal seminormal operator is a maximal normal operator, if it is weak*-upper semicontinuous on Y . A maximal normal operator is the Clarke subdifferential of a locally Lipschitzian function if and only if it is cyclical [16]. Theorem 4.5 generalizes this result. For other discussions on maximal normal operators, see [17] and [18].

Section 5. Integral Functionals

Again, we assume that X is a separable Banach space and Y is an open subset of X . Suppose that (W, \mathcal{A}, P) is a positive measure space. Consider the following integral functional:

$$f(x) = \int_W f_w(x) P(dw),$$

where

- (i) for each x in Y , the map $w \rightarrow f_w(x)$ is measurable;
- (ii) there exists a integrable function $g : W \rightarrow \mathfrak{R}$, such that for any x and $y \in Y$ and $w \in W$, one has

$$|f_w(x) - f_w(y)| \leq g(w) \|x - y\|.$$

Consider the Michel-Penot subdifferential of f . We first establish an inclusion relationship

$$\partial^\circ f(x) \subset \int_W \partial^\circ f_w(x) P(dw). \quad (10)$$

Then we try to get exact formulas for $\partial^\circ f(x)$. As in [5], (10) has the following interpretation. For every $\xi \in \partial^\circ f(x)$, there exists a mapping $w \rightarrow \xi_w$ from W to X^* , where $\xi_w \in \partial^\circ f_w(x)$ a.e. relative to P , and such that $\forall v \in X$, the function $w \rightarrow \langle \xi_w, v \rangle$ is in $L^1(X, \mathfrak{R})$ and $\langle \xi, v \rangle = \int_W \langle \xi_w, v \rangle P(dw)$.

Lemma 5.1. *Suppose that f is defined at some point $x \in Y$. Then f is defined and Lipschitzian in Y .*

Proof: Let y be any point in Y . Then the maps $w \rightarrow f_w(x)$ and $w \rightarrow f_w(y)$ are measurable. Hence,

$$\begin{aligned} |f(x) - f(y)| &\leq \int_{\mathcal{W}} |f_w(x) - f_w(y)| P(dw) \\ &\leq L \|x - y\|, \end{aligned}$$

where $L = \int_{\mathcal{W}} g(w) P(dw)$.

Thus f is defined and Lipschitzian on Y . ■

Theorem 5.2. *Under the above assumptions, for any $x \in Y$, (10) holds. Further, if f_w is semiregular at x for each w , then f is semiregular at x and equality holds in expression (10).*

Proof: This proof is similar to the proof of Theorem 2.7.2 of [5]. Let $u \in \partial^\circ f(x)$, and let h be any vector in X . Then by the definitions of $\partial^\circ f(x)$ and $f^\circ(x; h)$,

$$\begin{aligned} \langle u, h \rangle &\leq f^\circ(x; h) \\ &= \sup_{k \in X} \{ \limsup_{t \downarrow 0} \int_{\mathcal{W}} ([f_w(x + th + tk) - f_w(x + tk)]/t) P(dw) \}. \end{aligned}$$

According to conditions (i) and (ii) on f and Fatou's Lemma, the right hand side of the second inequality is not greater than

$$\int_{\mathcal{W}} \sup_{k \in X} \{ \limsup_{t \downarrow 0} [f_w(x + th + tk) - f_w(x + tk)]/t \} P(dw) = \int_{\mathcal{W}} f^\circ(x; h) P(dw).$$

By 2.2, the mapping $f^\circ(x; \cdot)$ is convex. Let

$$r(h) = \int_{\mathcal{W}} f^\circ(x; h) P(dw).$$

Then r is also a convex function and $r(0) = 0$. Thus,

$$\langle u, h \rangle \leq r(h) - r(0)$$

for any $h \in X$. Thus, $u \in \partial r(0)$. If the map $w \rightarrow r_w(h) = f^\circ(x; h)$ is measurable for each h , then $\partial r(0)$ is contained in $\int_W \partial r_w(0) P(dw)$ [5][8]. However, $\partial r_w(0) = \partial^\circ f(x)$ by construction, so the result would follow. Consider the map $w \rightarrow r_w(h)$ for each h ,

$$r_w(h) = \sup_{k \in X} \{ \limsup_{t \downarrow 0} [f_w(x + th + tk) - f_w(x + tk)]/t \}.$$

Since $f_w(\cdot)$ is continuous, we may let $t \downarrow 0$ only take rational values and k take values in a countable dense subset of X . Therefore, $r_w(h)$ as a function of w , is the countable supremum of a countable lim sup of measurable functions of w . Thus, it is also a measurable function of w . This proves (10). The proof of the second assertion is identical to the corresponding part of the proof of Theorem of 2.7.2 of [5] using semiregularity and the Michel-Penot subdifferential instead of regularity and the Clarke subdifferential. ■

Theorem 5.2 is essentially a Michel-Penot subdifferential version of Theorem 2.7.2 of [5]. The advantage obtained in this case is that under moderate requirements on X and W , the right hand side of (10) is singleton almost everywhere in Y , thus (10) becomes equality almost everywhere in Y . In stochastic programming, X is a space of functions on W . If W is countable or X is restricted to continuous functions or a finite dimensional space, we may use Theorem 5.2 directly. In general, however, X is an L^p space. The next theorem states the results for L^∞ . The result for general p follows closely.

Theorem 5.3. *Let (W, \mathcal{A}, P) be a σ -finite positive measure space and let $L^\infty(W, \mathfrak{R}^n)$ be the space of measurable essentially bounded functions mapping W to \mathfrak{R}^n . Suppose X is a closed subspace of $L^\infty(W, \mathfrak{R}^n)$ and $f_w : \mathfrak{R}^n \rightarrow \mathfrak{R}$ such that (i) holds and*

(ii') $\exists \epsilon > 0$, a function $g \in L^1(W, \mathfrak{R})$, such that, $\forall w \in W$, $k_1, k_2 \in B(x(w), \epsilon)$,

$$|f_w(k_1) - f_w(k_2)| \leq g(w) \|k_1 - k_2\|, \quad (11)$$

where $B(x(w), \epsilon)$ is the ball of radius ϵ about $x(w)$ in \mathfrak{R}^n . Then, for any x satisfying the conditions, f is Lipschitzian near x and (10) holds. Furthermore, equality holds in (10) if f_w is semiregular at $x(w)$ for all w .

Proof: First to show that f is Lipschitzian near x , suppose $\|x - y\|_\infty = \delta < \epsilon$, so that

$P\{w : \|x(w) - y(w)\| \geq \delta\} = 0$. Then,

$$\begin{aligned}
|f(x) - f(y)| &\leq \int_W |f_w(x(w)) - f_w(y(w))| P(dw) \\
&= \int_{\{w : \|x(w) - y(w)\| < \delta\}} |f_w(x(w)) - f_w(y(w))| P(dw) \\
&\quad + \int_{\{w : \|x(w) - y(w)\| \geq \delta\}} |f_w(x(w)) - f_w(y(w))| P(dw) \\
&\leq L\delta
\end{aligned} \tag{12}$$

where $L = \int_W g(w) P(dw)$. Hence, (12) implies f is Lipschitzian for all $y \in B_{L\infty}(x, \epsilon)$.

The proof of Theorem 5.2 can now be applied in exactly the same way, noting that X is separable. ■

The results in Theorems 5.2 and 5.3 lead to the complete characterization of subgradients of the integral functional through subgradients of the integrand. In the first case, we assume that X and W are finite dimensional and use the Lebesgue measure. In stochastic programming, this corresponds to decisions X that must be taken before the realization of random outcomes. A common example is the simple recourse stochastic program (see Wets [23]).

Proposition 5.4. *If X and W are finite dimensional, then under the assumptions for Theorem 5.2, the right hand side of (10) is single-valued over a subset Y' of Y such that $Y'' = Y \setminus Y'$ has zero measure. Thus equality holds for (10) on Y .*

Proof: Since f is Lipschitzian in Y for each w , $\partial^\circ f_w(x)$ is single-valued almost everywhere in Y for each x by the Rademacher Theorem. Let

$$Y' = \{x \in Y : \partial^\circ f_w(x) \text{ is single-valued almost everywhere in } W\}.$$

Then, the right hand side is single-valued over Y' . By the Fubini Theorem, the measure of $Y'' = Y \setminus Y'$ should be zero. ■

In the more general case, the equality in (10) holds almost everywhere in the Haar measure around x using Theorem 5.3.

Proposition 5.5. *Under the assumptions for Theorem 5.3, equality holds in (10) almost everywhere in the Haar measure on $B_{L\infty}(x, \epsilon)$.*

Proof: The proof follows similarly to the proof of Proposition 5.4. Note that f_w is Lipschitzian on \mathfrak{R}^n near $x(w)$ for each w . Hence, $\partial^\circ f_w$ is single valued almost everywhere in $B_{\mathfrak{R}^n}(x(w), \epsilon)$. The assumptions imply that $x(w)$ is finite for all w so we may translate to the origin without loss of generality by suitably redefining $f_w(v)$ as $f_w(v - x(w))$ for any $v \in \mathfrak{R}^n$. We follow this translation (which does not affect the Haar measure on L^∞) and assume $x \equiv 0$.

Consider the product measure on $\mathfrak{R}^n \times W$. From the Fubini Theorem, the set, $\{(w, y), \text{ where } \|y\| < \epsilon \text{ and } \partial^\circ f_w(y) \text{ is not single valued}\}$, must have zero measure. Moreover, for almost all $y \in \mathfrak{R}^n, \|y\| < \epsilon, \{w : \partial^\circ f_w(y) \text{ is not single valued}\}$ has zero measure.

We will now construct finite dimensional subspaces of L^∞ on which (10) holds almost everywhere near x . Since W is separable, we may construct a sequence of finite sigma fields $U_i \subset \mathcal{A}$ such that $U_i \subset U_{i+1}$ and for any $B \in \mathcal{A}, \exists B' \subset \cup_i U_i$ such that $B \supset B'$ and $P\{B \setminus B'\} = 0$. For each i , consider $V_i \subset L^\infty(W, \mathfrak{R}^n)$, where $V_i = \{y \text{ such that } y \text{ is measurable on } U_i\}$. This subspace is finite dimensional and equivalent to a $|U_i|$ -fold product of \mathfrak{R}^n . Also, note that the union of subspaces V_i is dense in L^∞ .

Again, $\partial^\circ f_w$ is single valued almost everywhere in $B_{V_i}(\epsilon)$ using the Lebesgue product measure on V_i . The Fubini Theorem gives that for almost all $y \in V_i, \|y\| < \epsilon, \{w : \partial^\circ f_w(y) \text{ is not single valued}\}$, has zero measure. Thus, the set of $\|y\| < \epsilon$, where $\{w : \partial^\circ f_w(y) \text{ is not single valued}\}$ does not have zero measure, has zero measure in finite dimensional subspaces whose union is dense in $L^\infty(W, \mathfrak{R}^n)$. Hence, for almost all $\|y\| < \epsilon$ in the Haar measure on $L^\infty(W, \mathfrak{R}^n)$, $\{w : \partial^\circ f_w(y) \text{ is not single valued}\}$ has zero measure (using [4, Theorem 7.2]). Therefore, (10) holds almost everywhere near x . ■

Notice that $\partial^\circ f(x) = \{f'(x)\}$ a.e. by 2.4. The next result relates this to the Clarke subdifferential.

Proposition 5.6. *For $x \in Y''$, as defined above, the Clarke subdifferential*

$$\partial^\circ f(x) = \text{conv}\{\lim_i f'_i(x_i) : x_i \rightarrow x, x_i \in Y'\};$$

the Michel-Penot subdifferential can be decided by

$$\max\{\langle u, h \rangle : u \in \partial^\circ f(x)\}$$

$$= \sup_{k \in X} \{ \limsup_{t \downarrow 0} \int_0^1 \langle \partial^\circ f(x + t(k + sh)), h \rangle ds \},$$

where the integral is only defined where the intersection of Y'' and $[x + tk, x + tk + th]$ has one-dimensional zero measure.

Proof: The first claim is due to Theorem 2.5.1 of [5]. The second is due to an argument similar to that in Section 4. ■

Section 6. Applications in Stochastic Programming

Much of our motivation is for analysis and computation in stochastic programming. Useful optimality conditions for these problems follow directly from the results above. The general form of the stochastic program that we consider is:

$$\inf_{x \in X} F(x) \text{ subject to } g_i(x, w) \leq 0, a.e., \forall i \in I, \quad (SP)$$

where $F(x) = \int f(x)P(dx)$, X is a separable Banach space, I is a finite index set, and (W, \mathcal{A}, P) is a positive measure space. In the stochastic programming context, X is a space of functions on W . The functions f and g may be written $f(x(w), w)$ and $g_i(x(w), w)$. To distinguish a random vector, we use the bold-face notation convention, e.g., \mathbf{x} .

In general, X is a closed subspace of $L^p(W, \mathcal{A}, P; \mathfrak{R}^n)$ for $1 \leq p \leq \infty$. The random vector, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ where $\mathbf{x}_1 = x_1 : w \rightarrow \mathfrak{R}^{n_1}$ and $\mathbf{x}_2 = x_2 : w \rightarrow \mathfrak{R}^{n_2}$, $n = n_1 + n_2$. This corresponds to an initial, or first-period, decision x_1 followed by a future or second-period decision x_2 , which may in turn be partitioned into decisions at distinct points in time. To obtain implementable decisions in this context, all points in X satisfy *nonanticipativity*, defined by:

$$(\mathbf{x}_1 - E \{ \mathbf{x}_1 | \mathcal{A}_1 \}) = 0, a.e. \quad (13)$$

where $E \{ \cdot | \mathcal{A}_1 \}$ represents conditional expectation with respect to a sigma field, $\mathcal{A}_1 \subset \mathcal{A}$. Practically, this corresponds to enforcing a decision before the realization of some random outcome to depend on data that is available at that time. In multiple period problems, several subfields are nested to correspond to the resolution of random outcomes as time proceeds.

The following theorem gives a necessary condition for an optimal solution to the general problem in (SP).

Theorem 6.1. *If \bar{x} is an optimal solution to (SP), and f and $g_i, i \in I$, satisfy conditions 5.i and 5.ii', then there exists some $\lambda \geq 0$ and measurable function, $\nu \in X^0$, the annihilator of X in X^* , such that*

$$\nu(w) = \xi(w) + \lambda\eta_i(w), \quad (14)$$

where $\xi(w) \in \partial f_w(\bar{x}(w))$ and $\eta_i(w) \in \text{conv}_{i \in I(\bar{x}(w), w)}\{\partial g_i(\bar{x}(w), w)\}$, $I(\bar{x}(w), w) = \{i | g_i(\bar{x}(w), w) = 0\}$.

Proof: We begin by rewriting the constraints as

$$G(x) = \int_{\mathcal{W}} \{\max_{i \in I} g_i(x(w), w)\} P(dw) \leq 0. \quad (15)$$

Now, (SP) becomes an optimization problem in x with a single constraint. From Michel and Penot ([9]) (a direct application of 2.8 and Proposition 3.7), if \bar{x} is optimal in (SP), then

$$0 \in \partial^\circ F(\bar{x}) + \lambda \partial^\circ G(\bar{x}), \quad (16)$$

for some $\lambda \geq 0$. Using Theorem 5.3, we have

$$\partial^\circ F(x) \subset \int_{\mathcal{W}} \partial^\circ f_w(x) P(dw)$$

and

$$\partial^\circ G(x) \subset \int_{\mathcal{W}} \partial^\circ [\max_{i \in I} g_i(x(w), w)] P(dw).$$

From Proposition 3.7,

$$\partial^\circ [\max_{i \in I} g_i(x(w), w)] \subset \text{conv}\{\partial^\circ g_i(x(w), x) : i \in I(x(w), w)\}, \quad (17)$$

where $I(x(w), w) = \{i | g_i(x(w), w) = 0\}$. The application of Theorem 5.3 and (17) to the right hand side of (16) yields

$$\partial^\circ F(x) + \lambda \partial^\circ G(x) \subset \int_{\mathcal{W}} \{\partial^\circ f_w(x) P(dw) + \lambda \text{conv}\{\partial^\circ g_i(x(w), x) : i \in I(x(w), w)\}\} P(dw). \quad (18)$$

Inclusions (16) and (18) imply that

$$0 = \int_{\mathcal{W}} \langle \xi(w) + \lambda\eta_i(w), v \rangle P(dw), \quad (19)$$

for all $v \in X$ and some $\xi(w) \in \partial f_w(\bar{x}(w))$ and $\eta_i(w) \in \text{conv}_{i \in I(x(w), w)} \{\partial^\circ g_i(\bar{x}(w), w)\}$.

This proves the result. ■

The theorem is strengthened if we can guarantee that $\nu \equiv 0$ is included in (14). This is true if $\mathcal{A}_1 = \mathcal{A}$, but not generally true otherwise. An additional assumption of some feasible extension of any initial decision is required to obtain this stronger result (see Dempster [6] for a discussion.)

Theorem 6.1 can, however, be used in conjunction with 5.5 for the construction of a class of algorithms for (SP). If we can find $\xi(w)$ and $\eta_i(w)$ such that the right hand side of (19) is less than some ϵ , then Proposition 5.5 assures us that we have an ϵ -subgradient of $F + \lambda G$. This can be combined with a subgradient method to obtain stationary points of $F + \lambda G$.

ACKNOWLEDGEMENT

The authors are thankful to Dr. Jeyakumar for references [3] and [22].

REFERENCES

- [1] J.R. Birge and L. Qi, "Computing block-angular Karmarkar projections with applications to stochastic programming", *Management Science* 34 (1988) 1472-1479.
- [2] J.M. Borwein, S.P. Fitzpatrick and J.R. Giles, "The differentiability of real functions on normed linear space using generalized subgradients", *Journal of Mathematical Analysis & Applications* 128 (1987) 512-534.
- [3] O. Caligaris and P. Oliva, "Necessary conditions for local Pareto minima", *Bollettino U.M.I.* 5-B (1986) 721-750.
- [4] J.P.R. Christensen, **Topology and Borel structure**, North-Holland, Amsterdam, 1974.
- [5] F.H. Clarke, **Optimization and Nonsmooth Analysis**, Wiley, New York, 1983.
- [6] M.A.H. Dempster, "On stochastic programming II: dynamic problems under risk," *Stochastics* 25 (1988) 15-42.
- [7] Y. Ermoliev and R. Wets, **Numerical Techniques in Stochastic Programming**, Springer-Verlag, Berlin, 1988.

- [8] A.D. Ioffe and V.M. Tihomirov, **Theory of Extremal Problems**, North-Holland, Amsterdam, 1974.
- [9] P. Michel and J.-P. Penot, “Calcul sous-différentiel pour des fonctions lipschitziennes et non lipschitziennes”, *C. R. Acad. Sc. Paris.* 298 (1984) 269-272.
- [10] P. Michel and J.-P. Penot, “A generalized derivative for calm and stable functions”, preprint, 1988.
- [11] F. Mignot, “Control dans les inequations variationelles elliptiques”, *J. Functional Analysis* 22 (1976) 130-185.
- [12] M.L. Overton, “On minimizing the maximum eigenvalue of a symmetric matrix”, *SIAM J. Matrix Analysis & Application* 9 (1988) 256-268.
- [13] M.L. Overton and R.S. Womersley, “On minimizing the spectral radius of a non-symmetric matrix: Optimality conditions and duality theory”, *SIAM J. Matrix Analysis & Application* 9 (1988) 473-498.
- [14] D. Preiss, “Differentiability of Lipschitz functions on Banach spaces”, to appear in: *Journal of Functional Analysis*.
- [15] A. Prekopa and R.J.-B. Wets, **Stochastic Programming 84**, *Mathematical Programming Study* 27 & 28 (1986).
- [16] L. Qi, “The maximal normal operator space and integration of subdifferentials of nonconvex functions”, to appear in: *Nonlinear Analysis, Theory Methods and Applications*.
- [17] L. Qi, “Semismoothness and decomposition of maximal normal operators”, to appear in: *Journal of Mathematical Analysis & Applications*.
- [18] L. Qi, “Quasidifferentials and maximal normal operators”, to appear in: *Mathematical Programming*.
- [19] L. Qi and R.S. Womersley, “On extremal singular value functions”, in preprint.
- [20] R.T. Rockafellar, **The Theory of Subgradients and Its Applications to Problems of Optimization: Convex and Nonconvex Functions**, Helderman-Verlag, West Berlin, 1981.
- [21] J.S. Treiman, “Shrinking generalized gradients,” *Nonlinear Analysis, Theory, Methods and Applications* 12 (1988) 1429–1450.
- [22] D. Ward, “Convex directional derivatives in optimization”, preprint, 1988.

[23] R. J-B. Wets, "Solving stochastic programs with simple recourse," *Stochastics* 10 (1983) 219-242.