SUBDIFFERENTIAL CONVERGENCE IN STOCHASTIC PROGRAMS

John R. Birge
Department of Industrial & 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 91-12

April 1991

SUBDIFFERENTIAL CONVERGENCE IN STOCHASTIC PROGRAMS *

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

(Revised on April 15, 1991)

Abstract

In this paper, we discuss convergence behavior of subdifferentials in approximation schemes for stochastic programs. This information is useful for solving stochastic programs by nonlinear programming techniques. Wets [33] showed that epiconvergence of closed convex functions implies the set convergence of the graph of the subdifferentials of these functions. This conclusion is not true in general by a counterexample of Higle and Sen [14]. We show that epiconvergence of closed convex functions implies set convergence of subdifferentials of these functions at points where the limit function is G-differentiable and apply this result to convex stochastic programs. We also show that similar results can be achieved in three other cases of expectational functionals: piecewise smooth integrands, continuous probability distributions and loss functions. In the case of loss functions, we extend the existing results of Marti [19] to more general situations. Some basic methods using the approximate derivative information are also discussed.

Keywords: approximation, subdifferential, stochastic programming.

Section 1. Introduction

* This work is supported in part by the Australian Research Council. John Birge's work is also supported in part by Office of Naval Research Grant N0014-86-K-0628 and the National Science Foundation under Grant EECS-885101.

In general, a *stochastic programming problem* can be formulated as [36]:

$$
\begin{aligned}
&\text{minimize } E\{f_0(x,\xi)\} \\
&\text{subject to} E\{f_i(x,\xi)\} \le 0, \;\; i = 1, ..., s, \\
&\qquad\qquad E\{f_i(x,\xi)\} = 0, \;\; i = s+1, ..., m, \\
&\qquad\qquad x \in X \subseteq \Re^n,
\end{aligned}
\tag{1.1}
$$

where

- ξ is a random vector with support $\Xi \subseteq \Re^N$, and a probability distribution function P on $\Re^N$,
- $f_0 : \Re^n \times \Xi \to \Re \cup \{+\infty\}$,
- $f_i : \Re^n \times \Xi \to \Re$, $i = 1, ..., m$,
- X is closed,
- for all $x \in X$, and $i = 0, 1, ..., m$, the *expectational functional*

$$(Ef_i)(x) := E\{f_i(x,\xi)\} = \int_\Xi f_i(x,\xi) dP(\xi)$$

is finite. The *stochastic program with recourse* and the *stochastic program with chance constraints* are two special cases of this model [36]. The numerical solution for (1.1) is not an easy task [11][26]. One popular method is to use a discrete distribution to approximate P. In the case of the stochastic program with recourse, some approximations for the recourse function have also been proposed [5][6][16]. This results in approximation of $f_0(x,\xi)$. The resulting problem is a large-scale program, which can be solved by either decomposition methods [34], or perhaps variants of the interior point algorithm [2]. One may also propose methods to approximate $f_i(x,\xi)$, for $i = 1, ..., m$, and solve the resulting relatively tractable problem.

In the stochastic programming literature, most of the literature focuses on the convergence of the solution vector as a by-product of the epi-convergence of the approximating expectational functionals. Some recent papers on this topic include [5], [9], [12], [14], [15], [17], [18] and [30]. Can we also get some approximation information on the subdifferentials of the objective function and the constraints of (1.1)? This information will be useful for solving stochastic programming by nonlinear programming techniques which use subgradients [11] [19] [20] [23] [24]. In [31], Wang suggested to solve stochastic programs by an *approximate nonlinear programming method*, i.e., solving the original problem by a nonlinear programming method and at the kth step using the kth approximation function value and its derivative or subgradient as the approximation value of the original problem and its derivative or subgradient. For this approach, the approximation information of subdifferentials is especially useful. Wang used the term *differential stability* to describe this information.

It seems that Marti [19] first addressed this problem. He considered a special class of expectational functionals: convex loss functions. This class is the most common expectational functional in the recourse problem. Wets [33] proved that epiconvergence of closed convex functions implies the set convergence of the graph of the subdifferentials of these functions. In Section 2, we show that epiconvergence of closed convex functions implies set convergence of subdifferentials of these functions at points where the limit function is G-differentiable and apply this result to convex stochastic programs. In particular, combining with recent epiconsistency result for convex stochastic programs of King and Wets [17], we establish general conditions for subdifferential convergence of approximation schemes for those problems in the sense of probability one. According to this result, a stochastic quasigradient-type method (see [10]) is suggested at the end of that section.

Recently, Higle and Sen [14] gave examples that epi-convergence does not imply subdifferential convergence in the nonconvex case. They also introduced a notion of subdifferential convergence, which is weaker than the one we use here. Therefore, we cannot expect that nice subdifferential approximation behavior can be achieved as a bonus of epi-convergence in general. Since in practice we always use some approximation functions, it is thus important to ask in which cases nice subdifferential approximation behavior can be achieved. This behavior may be especially critical when subdifferential information is required by an algorithm or sensitivity analysis.

We show that nice approximation behavior of subdifferentials of expectational functionals can be achieved in three other cases, namely, piecewise smooth integrands, continuous distributions and loss functions. We study these three cases in Sections 3-5 respectively.

The significance of piecewise smooth integrands is clear. The continuous distribution approximation probably was first suggested by Wets [32]. In that paper, he suggested a piecewise linear distribution approximation. In [8], Dexter, Yu and Ziemba suggested using the linear combination of lognormal univariate distribution approximations. Wets also discussed the possibility of continuous distribution approximations in [34] and [35]. Gassmann [13] discussed applications of normal distributions in stochastic programming.

Important special cases of expectational functionals are loss functions. Marti [19] studied the subdifferential convergence behaviors of this class of functions as we commented before. Our results cover more general cases, and we show that everywhere strictly differentiable results can be obtained in reasonable conditions. Thus, in these special cases, differentiable methods may be used instead of nondifferentiable methods that are otherwise necessary.

The main uses for our results are in the ability to use general optimization techniques that do not require complete optimization for a single approximation and that allow differentiable techniques to be used in intermediate approximation iterations. We discuss possible applications in algorithms in Section 6. We give general methods and show how approximations may be used without requiring optimization for each approximation ν. Certainly, this section only gives general ideas for possible uses of subdifferential approximation results. For more sophisticated nonlinear programming approaches in stochastic programming, the readers may refer to Marti [20], Marti and Fuchs [21], Nazareth and Wets [23] [24]. For the approximate nonlinear programming method, see Wang [31].

As said before, the issue of subdifferential convergence has been investigated by Marti [19], Wets [33], Higle and Sen [14] and Wang [31] in different contexts. While we study this issue again in this paper, we wish not to depreciate other approaches, such as the stochastic quasigradient method [10] [11] which depends upon sampling. Each method has its advantages in particular situations. Moreover, sometimes they can be employed together, as in the case discussed in Section 2.

Section 2. Convex Stochastic Programs

Consider the *expectational functional* $E_f = E\{f(.,\xi))\}$, where ξ is a random vector with support $\Xi \subseteq \Re^N$ and f is an extended real-valued function on $\Re^n \times \Xi$, as described in the introduction. One has

$$E_f(x) = \int f(x,\xi)P(d\xi), \tag{2.1}$$

where P is a probability measure defined on $\Re^N$. It is usually difficult to calculate E_f and its derivative or subdifferential. A popular approach is to approximate (2.1) by

$$E_f^\nu(x) = \int f(x,\xi)P_\nu(d\xi), \tag{2.2}$$

where $\{P_\nu, \nu = 1, ...\}$ is a sequence of probability measures converging in distribution to the probability measure P.

In the following, we use E_f^0 and P_0 instead of E_f and P for convenience. We denote the Lebesgue measure by m. If $C \subseteq \Re^n$ is a closed convex set, then Ψ_C^* is the support function of C, defined by $\Psi^*(h|C) = \sup\{< x, h >: x \in C\}$. A sequence of closed convex sets $\{C_\nu : \nu = 1, ...\}$ in $\Re^n$ is said to converge to a closed convex set C in $\Re^n$ if for any $h \in \Re^n$,

$$\lim_{\nu \to +\infty} \Psi^*(h|C_\nu) = \Psi^*(h|C).$$

One may easily prove the following proposition:

Proposition 2.1 *Suppose that C and C_ν, for $\nu = 1, ...,$ are closed convex sets in $\Re^n$. The following two statements are equivalent:*

(a) *C_ν converges to C as $\nu \to +\infty$;*

(b) *a point $x \in C$ if and only if there are $x_\nu \in C_\nu$ such that $x_\nu \to x$.* ∎

In the following, ∂ is the symbol for the generalized subdifferential in the sense of Clarke [7]. Suppose an extended real-valued function $g : \Re^n \to \Re \cup \{+\infty\}$ is locally Lipschitz at x, where x is in the essential domain of f. Then, the *Clarke directional derivative* of g at x with respect to $h \in \Re^n$ is $g^\circ(x;h)$, where

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

The subdifferential of g at x is then:

$$\partial g(x) := \{u \in \Re^n : < u, h > \le g^\circ(x;h), \forall h \in \Re^n\}.$$

When g is convex, the above definition coincides with the definition in convex analysis:

$$\partial g(x) := \{u \in \Re^n : \; g(x+h) \ge g(x) + < u, h >, \; \forall h \in \Re^n\}.$$

The conditions of the following theorem are the same as the conditions of Theorem 2.8 of [5]. We only slightly strengthen its conclusions for our further use.

Theorem 2.2 *Suppose that*

(i) *$\{P_\nu, \nu = 1, ...\}$ converges in distribution to P;*

(ii) *$f(x, \cdot)$ is continuous on Ξ for each $x \in D$, where*

$$D = \{x : E_f(x) < +\infty\} = \{x : f(x,\xi) < +\infty, a.s.\};$$

(iii) *$f(\cdot, \xi)$ is locally Lipschitz on D with Lipschitz constant independent of ξ;*

(iv) *for any $x \in D$ and $\epsilon > 0$, there exists a compact set S_ϵ and ν_ϵ such that for all $\nu \ge \nu_\epsilon$,*

$$\int_{\Xi \setminus S_\epsilon} |f(x,\xi)| P_\nu(d\xi) < \epsilon,$$

and with $V_x = \{\xi : f(x,\xi) = +\infty\}, P(V_x) > 0$ if and only if $P_\nu(V_x) > 0$ for $\nu = 0, 1, ...$

Then

(a) *E_f^ν epi- and pointwise converges to E_f; if $x, x^\nu \in D$ for $\nu = 1, 2, ...$ and $x^\nu \to x$, then*

$$\lim_{\nu \to \infty} E_f^\nu(x^\nu) = E_f(x); \tag{2.3}$$

(b) *E_f^ν, where $\nu = 0, 1, \ldots$, is locally Lipschitz on D; furthermore, for each $x \in D$, $\{\partial E_f^\nu(x) : \nu = 0, 1, \ldots.\}$ is bounded;*

(c) *if $x^\nu \in D$ minimizes E_f^ν for each ν and x is a limiting point of $\{x^\nu\}$, then x minimizes E_f.*

Proof:

The epi- and point convergence were established in [5]. Let $x, x^\nu \in D, x^\nu \to x$. Then

$$\begin{aligned}
&\lim_{\nu\to\infty} |E_f^\nu(x^\nu) - E_f^\nu(x)| \\
&\le \lim_{\nu\to\infty} \int |f(x^\nu,\xi) - f(x,\xi)|P_\nu(d\xi) \\
&\le \lim_{\nu\to\infty} \int L_x\|x^\nu - x\|P_\nu(d\xi) \\
&= \lim_{\nu\to\infty} L_x\|x^\nu - x\| \\
&=0,
\end{aligned}$$

where L_x is the Lipschitzian constant of $f(\cdot,\xi)$ near x, which is independent of ξ by (iii). By point convergence of E_f^ν at x,

$$\lim_{\nu\to\infty} E_f^\nu(x) = E_f(x).$$

Putting these two results together, we have (2.3). This proves (a).

For any $x \in D$, y and z close to x, $\nu = 0, 1, \ldots,$

$$\begin{aligned}
&|E_f^\nu(y) - E_f^\nu(z)| \\
&\le \int |f(y,\xi) - f(z,\xi)|P_\nu(d\xi) \\
&\le \int L_x\|y - z\|P_\nu(d\xi) \\
&=L_x\|y - z\|,
\end{aligned}$$

where L_x is the Lipschitzian constant of $f(\cdot,\xi)$ near x, which is independent of ξ by (iii). By [7], $\partial E_f^\nu(x)$ is a nonempty, compact convex set, for each ν; and the 2-norms of subgradients in these subdifferentials are bounded by L_x. This proves (b).

By (b), E_f^ν are lower semicontinuous functions. By (a), E_f^ν epi-converges to E_f. By Theorem 3.7 of [36], we get the conclusion of (c). This completes the proof. ∎

Certainly, the above proof is not difficult. We are interested in the conclusions. Since E_f^ν is locally Lipschitz on D, by [7], for each $x \in D$, $\partial E_f^\nu(x)$ exists and is a nonempty, compact convex set. Will $\partial E_f^\nu(x)$

converge to $\partial E_f(x)$ in the sense of set convergence we just defined? Furthermore, by the Rademacher Theorem, E_f^ν is G-differentiable a.s. on D. Will $\nabla E_f^\nu(x)$ converge to $\nabla E_f(x)$ for each $x \in D \setminus D_1$, where $m(D_1) = 0$? These are the topics of investigation of this paper.

In the case of closed convexity, we may invoke Theorem 3 of Wets [33]. He proved that if $g, g^\nu : \Re^n \to \Re \cup \{+\infty\}, \nu = 1, 2, ...,$ are closed convex functions and $\{g^\nu\}$ epiconverges to g, then the graphs of the subdifferentials of g^ν converge to the graph of the subdifferential of g, i.e., for any convergent sequence $\{(x^\nu, u^\nu) : u^\nu \in \partial g^\nu(x^\nu)\}$ with (x, u) as its limit, one has $u \in \partial g(x)$; for any (x, u) with $u \in \partial g(x)$, there exists at least one such sequence $\{(x^\nu, u^\nu) : u^\nu \in \partial g^\nu(x^\nu)\}$ converging to it.

However, in general it is not true that

$$\partial g(x) = \lim_{\nu \to \infty} \partial g^\nu(x) \tag{2.4}$$

even if $x \in \text{int}(\text{dom}(g))$. For example, let $n = 1, g(x) = |x|, g^\nu(x) = |x|$ if $|x| \geq \frac{1}{\nu}, g^\nu(x) = \frac{1}{\nu}$ if $|x| < \frac{1}{\nu}$, for $\nu = 1, 2, ...$ Then g and g^ν are closed convex functions, $\{g^\nu\}$ epiconverges to g. For $\nu = 1, 2, ..., g^\nu$ is differentiable at $x = 0$ with $\nabla g^\nu(0) = 0$. But $\partial g(0) = [-1, 1]$. However, if g is G-differentiable at x, (2.4) is true. We prove this fact here.

Theorem 2.3 *Suppose that* $g, g^\nu : \Re^n \to \Re \cup \{+\infty\}, \nu = 1, 2, ...,$ *are closed convex functions and* $\{g^\nu\}$ *epiconverges to* g. *Suppose further that* g *is G-differentiable at* x. *Then*

$$\nabla g(x) = \lim_{\nu \to \infty} \partial g^\nu(x). \tag{2.5}$$

Proof:

By Theorem 3 of [33], any convergent subsequence of $\{u^\nu \in \partial g^\nu(x)\}$ converges to $\nabla g(x)$. What we need to prove is that there exists ν_x such that for any $\nu \geq \nu_x$, $\partial g^\nu(x)$ is nonempty, and $\{\partial g^\nu(x) : \nu \geq \nu_x\}$ is bounded. We now prove these. Since g is G-differentiable at x, $x \in \text{int}(\text{dom}(g))$. By Corollary 5 of [33], $g^\nu(x)$ converges to $g(x)$. Thus, for ν big enough, $x \in \text{dom}(g^\nu)$, i.e., $\partial g^\nu(x)$ is nonempty since g^ν is closed convex. For any $\epsilon > 0$, since g is continuous at x, there exists a $t > 0$ such that

$$g(x + te_i) \leq g(x) + \epsilon,$$

$$g(x - te_i) \leq g(x) + \epsilon,$$

for $i = 1, 2, ..., n$, where $e_i, i = 1, 2, ..., n$, are the unit vectors in $\Re^n$. Since g^ν epiconverges to g, by the properties of epiconvergence [32] [33] [35], there exist $y_i^\nu \to x + te_i$ and $z_i^\nu \to x - te_i$ such that

$$\limsup_{\nu \to \infty} g^\nu(y_i^\nu) \leq g(x + te_i) \leq g(x) + \epsilon$$

and

$$\limsup_{\nu \to \infty} g^\nu(z_i^\nu) \leq g(x - te_i) \leq g(x) + \epsilon.$$

Then, there exists ν_x such that for all $\nu \geq \nu_x$, $x \in \text{dom}(g^\nu)$,

$$g^\nu(y_i^\nu) \leq g(x) + 2\epsilon, \tag{2.6}$$

$$g^\nu(z_i^\nu) \leq g(x) + 2\epsilon, \tag{2.7}$$

$$g^\nu(x) \geq g(x) - \epsilon, \tag{2.8}$$

$$||y_i^\nu - (x + te_i)|| \leq \frac{t}{n}, \tag{2.9}$$

and

$$||z_i^\nu - (x - te_i)|| \leq \frac{t}{n}, \tag{2.10}$$

where (2.8) holds because $g^\nu(x)$ converges to $g(x)$. By (2.9) and (2.10), for any y satisfying $||y - x|| \leq \frac{t}{n}$,

$$y \in \text{conv}\,\{y_1^\nu, ..., y_n^\nu, z_1^\nu, ..., z_n^\nu\}.$$

By (2.6), (2.7) and convexity of g^ν, for any y satisfying $||y - x|| \leq \frac{t}{n}$,

$$g^\nu(y) \leq g(x) + 2\epsilon. \tag{2.11}$$

Now, for any $u \in \partial g^\nu(x)$ where $\nu \geq \nu_x$, if $u \neq 0$, let

$$y = x + \frac{tu}{n||u||}.$$

By convexity of g^ν,

$$g^\nu(y) \geq g^\nu(x) + < u, y - x > = g^\nu(x) + \frac{t}{n}||u||. \tag{2.12}$$

By (2.12), (2.11) and (2.8),

$$||u|| \leq [g^\nu(y) - g^\nu(x)]/t \leq \frac{3n\epsilon}{t}.$$

This proves that $\{\partial g^\nu(x) : \nu \geq \nu_x\}$ is bounded. The proof of this theorem is completed. ∎

Remark: In fact, for any $x \in \text{int}(\text{dom}(g))$, there exists ν_x such that for any $\nu \geq \nu_x$, $\partial g^\nu(x)$ is nonempty, and $\{\partial g^\nu(x) : \nu \geq \nu_x\}$ is bounded. Thus, for any $x \in \text{int}(\text{dom}(g))$, the right hand side of (2.4) is nonempty and always contained in the left hand side of (2.4). But equality does not necessarily hold by our example. ∎

We now apply Theorem 2.3 to expectational functionals.

Corollary 2.4 *In (2.1) and (2.2), suppose that $f(\cdot,\xi)$ is closed convex for each $\xi \in \Xi$ and that E_f^ν epiconverges to E_f. Then for $D = dom(E_f)$,*

(d) *there is a Lebesgue zero-measure set $D_1 \subseteq D$ such that E_f is G-differentiable on $D \setminus D_1$, E_f is not G-differentiable on D_1, and for each $x \in D \setminus D_1$*

$$\lim_{\nu\to\infty} \partial E_f^\nu(x) = \nabla E_f(x);$$

(e) *for each $x \in D$,*

$$\partial E_f(x) = \{\lim_{\nu\to\infty} u^\nu : u^\nu \in \partial E_f^\nu(x^\nu), x^\nu \to x\}.$$

Proof:

By closed convexity of $f(\cdot,\xi)$, E_f^ν are also closed convex for all ν. Now (d) follows Theorem 2.3 and the differentiability property of convex functions, and (e) follows Theorem 3 of [33]. ∎

One now can combine Corollary 2.4 with any epiconvergence results for convex stochastic programs. For example, one may combine Corollary 2.4 with Theorem 2.2.

Corollary 2.5 *In the setting of Theorem 2.2, if $f(\cdot,\xi)$ is convex for each $\xi \in \Xi$, then (d) and (e) hold.*

Proof:

By Theorem 2.2, E_f^ν epi-converges to E_f. By convexity of $f(\cdot,\xi)$, E_f^ν are also convex for all ν. By (b), E_f^ν are lower semicontinuous, thus closed convex. Now the conclusions follow Corollary 2.4. ∎

One may also combine Corollary 2.4 with some recent results of epiconvergence, such as the results of Lepp [18]. Perhaps an interesting combination is with the results of King and Wets [17]. We now let P_ν be an empirical measure derived from an independent series of random observations $\{\xi_1, ..., \xi_\nu\}$ each with common distribution P. Then for all x,

$$E_f^\nu(x) = \frac{1}{\nu}\sum_{i=1}^{\nu} f(x,\xi_i). \tag{2.13}$$

Let $(\Xi, \mathcal{A}, P)$ be a probability space completed with respect to P. A closed-valued multifunction G mapping Ξ to $\Re^n$ is called *measurable* if for all closed subsets $C \subseteq \Re^n$, one has

$$G^{-1}(C) := \{\xi \in \Xi : G(\xi) \cap C \neq \emptyset\} \in \mathcal{A}.$$

In the following, "with probability one" refers to the sampling probability measure on $\{\xi_1, ..., \xi_\nu, ...\}$ that is consistent with P (see [17] for details). Applying Theorem 2.3 of [17] and Corollary 2.4 of this paper, we have:

Corollary 2.6 *Suppose for each $\xi \in \Xi$, $f(\cdot, \xi)$ is closed convex and the epigraphical multifunction $\xi \mapsto$ epi $f(\cdot, \xi)$ is measurable. Let E_f^ν be calculated by (2.13). If there exist one point $\bar{x} \in dom(E_f)$and a measurable selection $\bar{u}(\xi) \in \partial f(\bar{x}, \xi)$ with $\int \|\bar{u}(\xi)\| P(d\xi)$ finite, then the conclusions of Corollary 2.4 hold with probability one.* ∎

King and Wets [17] applied their results to the *stochastic program with fixed recourse*

$$\begin{aligned} &\text{minimize } cx + \int Q(x, \xi) P(d\xi) \\ &\text{subject to } Ax = b, \\ &\qquad x \geq 0, \end{aligned} \tag{2.14}$$

where $x \in \Re^n$ and

$$Q(x, \xi) = \inf\{q(\xi)y : Wy = T(\xi)x - \zeta(\xi), y \in \Re^m_+\} \tag{2.15}$$

It is a fixed recourse problem since W is deterministic. Combining their Theorem 3.1 with our Corollary 2.4, we have

Corollary 2.7 *Suppose that the stochastic program (2.14) has fixed recourse (2.15) and that for all i, j, k, the random variables $q_i \zeta_j$ and $q_i T_{jk}$ have finite first moments. If there exists a feasible point $\bar{x}$ of (2.14) with the objective function of (2.14) finite, then the conclusions of Corollary 2.4 hold with probability one for*

$$f(x, \xi) = cx + Q(x, \xi) + \delta(x),$$

where $\delta(x) = 0$ if $Ax = b, x \geq 0$, $\delta(x) = +\infty$ otherwise. ∎

By Theorem 3.1 of [17], one may solve the approximation problem

$$\begin{aligned} &\text{minimize } cx + \frac{1}{\nu}\sum_{i=1}^{\nu} Q(x,\xi_i) \\ &\text{subject to } Ax = b, \\ &x \geq 0, \end{aligned} \tag{2.16}$$

instead of solving (2.14). If the solution of (2.16) converges as ν tends to infinity, then the limiting point is a solution of (2.14). Alternatively, by Corollary 2.7, one may directly solve (2.14) with a nonlinear programming method and use

$$cx + \frac{1}{\nu}\sum_{i=1}^{\nu} Q(x,\xi_i)$$

and

$$c + \frac{1}{\nu}\sum_{i=1}^{\nu} \partial_x Q(x,\xi_i)$$

as approximate objective function values and subdifferentials of (2.14) with $\nu = \nu(k)$ at the kth step. Notice that $u \in \partial_x Q(x,\xi_i)$ if and only if u is an optimal dual solution of (2.15) with $\xi = \xi_i$. In this way, one may directly solve the original problem but avoid the multiple integrals. Certainly, this approach needs further investigation for its convergence and practical performance.

Section 3. Piecewise Smooth Integrands

In the nonconvex case, if f is a "piecewise smooth" function in x, then we may still get similar results. We say a mapping $(x,\xi) \mapsto f(x,\xi) : \Re^n \times \Xi \to \Re \cup \{+\infty\}$ is a piecewise smooth function in x if $\Re^n \times \Xi$ can be partitioned into countable pieces of convex polyhedra such that for any (x,ξ) in the interior of a piece, $\nabla_x f(x,\xi)$ exists and is continuous in ξ,

$$m(K(\xi)) = 0,$$

for each $\xi \in \Xi$, where $K(\xi) = \{x : (x,\xi) \text{ is on the boundary of a piece of } f\}$. We assume that any polyhedral set in $\Re^n \times \Re^N$ is measurable under the product measure $m \times P_\nu$ for $\nu = 0, 1, \ldots$.

Theorem 3.1 *In the setting of Theorem 2.2, we only assume (i) and (iii) hold. If f is a piecewise smooth function in x as defined above, then the conclusion (b) holds and*

(f) *there is a Lebesgue zero-measure set $D_1 \subseteq D$ such that E_f^ν, for $\nu = 0,1,2,\ldots$ are G-differentiable on $D \setminus D_1$ and for each $x \in D_2 = D \setminus D_1$*

$$\lim_{\nu\to\infty} \nabla E_f^\nu(x) = \nabla E_f(x). \tag{3.1}$$

Proof:

The conclusion (b) holds since in the proof of Theorem 2.2 (b) only (iii) is invoked. Let J be the set of all the boundary (facet) points of polyhedron pieces of f. Then J is measurable under the product measure $m \times P_\nu$, for $\nu = 0, 1, \ldots$

Let $J(x) = \{\xi \in \Xi : (x, \xi) \in J\}$ for each $x \in D$. Let $D^\nu = \{x \in D : J(x) \text{ has positive } P_\nu \text{ measure}\}$. By (2.5) and the Fubini Theorem, $m(D^\nu) = 0$. Let

$$D_1 = \cup_{\nu=0,1,\ldots} D^\nu$$

and $D_2 = D \setminus D_1$. Then the Lebesgue measure of D_1 is zero. We now prove (f) is true for the sets D_1 and D_2.

Let $x \in D_2$. By the property of D_2, the set $J(x)$ has zero measure in P_ν and m, $\nabla_x f(x, \xi)$ exists and is continuous in ξ for $\xi \in \Xi \setminus J(x)$. Let $h \in \Re^n$. Denote $f_\xi(x) = f(x, \xi)$. By the dominated convergence theorem, we have the existence of $(E_f^\nu)'(x; h)$ and the equality

$$(E_f^\nu)'(x; h) = \int f'_\xi(x; h) P_\nu(d\xi).$$

But

$$f'_\xi(x; h) = < \nabla_x f(x, \xi), h >,$$

for $\xi \in \Xi \setminus J(x)$. The existence of $\nabla E_f^\nu(x)$ follows.

Let $\epsilon > 0$ be given. Let $h \in \Re^n$. By (b) and the property of D_2, $f'_\xi(x; h)$ is bounded by L_x and continuous in $\Xi \setminus J(x)$, where L_x is the Lipschitz constant in the proof of Theorem 2.2 (b). Let

$$C_k = \{\xi \in \Xi : d(\xi, J(x)) < \frac{1}{k}\},$$

where δ is a positive number. Since $J(x)$ is a union of polyhedra of dimension less than N, C_k is also a union of polyhedra, thus, measurable in P^ν for $\nu = 0, 1, \ldots$ But

$$\cap_{k=1}^\infty C_k = J(x).$$

Since $P(J(x)) = 0$, sufficiently large k, we have

$$P(C_{k+1}) \leq P(C_k) < \frac{\epsilon}{5L_x}.$$

We may define a bounded continuous function $\phi : \Xi \to \Re_+$ such that $\phi(\xi) = 1$ if $\xi \in C_{k+1}$, $0 \le \phi(\xi) \le 1$ if $\xi \in C_k \setminus C_{k+1}$, and $\phi(\xi) = 0$ otherwise. Then

$$|\int \phi(\xi)P(d\xi)| \le P(C_k) < \frac{\epsilon}{5L_x}.$$

and

$$P_\nu(C_{k+1}) \le \int \phi(\xi)P_\nu(d\xi)$$

for $\nu = 1, 2, \ldots$ By (i)

$$\lim_{\nu\to\infty} \int \phi(\xi)P_\nu(d\xi) = \int \phi(\xi)P(d\xi).$$

Combining these, we see that there is a ν'_ϵ such that for all $\nu \ge \nu'_\epsilon$,

$$P_\nu(C_{k+1}) < \frac{\epsilon}{5L_x}.$$

Denote such a C_{k+1} by C. We may also define a continuous bounded function $g : \Xi \to \Re$ such that

$$g(\xi) = f'_\xi(x; h)$$

for $\xi \in \Xi \setminus C$, and g is also bounded by L_x. By (i), there is a ν''_ϵ such that for all $\nu \ge \nu''_\epsilon$,

$$|\int g(\xi)P_\nu(d\xi) - \int g(\xi)P(d\xi)| < \frac{\epsilon}{5}.$$

Now, for all $\nu \ge \max\{\nu'_\epsilon, \nu''_\epsilon\}$,

$$\begin{aligned}
&|(E^\nu_f)'(x;h) - (E_f)'(x;h)| \\
=&|\int f'_\xi(x;h)P_\nu(d\xi) - \int f'_\xi(x;h)P(d\xi)| \\
\le&\int |f'_\xi(x;h) - g(\xi)|P_\nu(d\xi) + \int |f'_\xi(x;h) - g(\xi)|P(d\xi) + |\int g(\xi)P_\nu(d\xi) - \int g(\xi)P(d\xi)| \\
\le&2L_xP_\nu(C) + 2L_xP(C) + \frac{\epsilon}{5} \\
\le&\epsilon.
\end{aligned}$$

Thus,

$$\lim_{\nu\to\infty} (E^\nu_f)'(x;h) = (E_f)'(x;h).$$

Since this is true for all $h \in \Re^n$, (3.1) holds. This completes the proof. ∎

Section 4. Continuous Distributions

We say that a function $p : \Re^N \rightarrow \Re$ is piecewise continuous if $\Re^N$ can be partitioned into countable pieces of convex polyhedra such that p is continuous in the interior of each piece. We say a probability measure P on $\Re^N$ is *continuous* if there is a piecewise continuous function $p : \Re^N \rightarrow \Re_+$ such that

$$P(d\xi) = p(\xi)d\xi.$$

One may anticipate that many practical probability measures are continuous. Piecewise linear and piecewise constant probability measures may be regarded as possible candidates for approximating measures P_ν. In the following, we will see that if P is a continuous probability measure and we use continuous probability measures to approximate it, then we have similar results to Corollary 2.3 and Theorem 3.1, even slightly stronger.

In the proofs of the following theorem and Theorem 5.2, we use an alternative subdifferential of non-convex functions besides the Clarke subdifferential to establish G-differentiability results. We let $\partial^\diamond$ be the symbol for the generalized subdifferential in the sense of Michel and Penot [22]. Again, suppose an extended real-valued function $g : \Re^n \rightarrow \Re \cup \{+\infty\}$ is locally Lipschitz at x, where x is in the essential domain of f. Then, the *Michel-Penot directional derivative* of g at x with respect to $h \in \Re^n$ is $g^\diamond(x; h)$, where

$$g^\diamond(x;h) := \sup_{k\in\Re^n} \{\limsup_{t\downarrow 0} [g(x+tk+th) - g(x+tk)]/t\}.$$

The Michel-Penot subdifferential of g at x is then:

$$\partial^\diamond g(x) := \{u \in \Re^n : < u, h > \leq g^\diamond(x;h), \forall h \in \Re^n\}.$$

An advantages of the Michel-Penot subdifferential is that $\partial^\diamond g(x)$ is singleton if and only if g is G-differentiable at x. Thus, g is G-differentiable at x if and only if

$$g^\diamond(x;e_i) + g^\diamond(x;-e_i) = 0,$$

for $i = 1, 2, ..., n$, where e_i is the ith unit vector. This fact is used in the proof of the following theorem.

Theorem 4.1 *In the setting of Theorem 2.2, we only assume (iii) holds. If $\{P_\nu, \nu = 0, 1, ...\}$ are continuous probability measures with*

$$P_\nu(d\xi) = p_\nu(\xi)d\xi,$$

where $p_\nu : \Xi \rightarrow \Re_+$, $\nu = 0, 1, 2, ...$ are piecewise continuous functions,

(v)

$$\lim_{\nu\rightarrow\infty} \int |p_\nu(\xi) - p(\xi)|d\xi = 0;$$

(vi) *the map* $(x,\xi) \mapsto f(x,\xi)$ *is Lebesgue measurable in* $D \times \Re^N$;

then the conclusions (f) and (g) hold:

(f) *the same as in Theorem 3.1.*

(g) *for any* $x \in \text{int}(D)$,

$$\lim_{\nu \to +\infty} \partial E_f^\nu(x) = \partial E_f(x). \tag{4.1}$$

Proof:

(f). By (iii), $f(.,\xi)$ is locally Lipschitz for each $\xi \in \Xi$. By the Rademacher Theorem, $f(.,\xi)$ is G-differentiable almost everywhere in D for each $\xi \in \Xi$. Let

$$\bar{J} = \{(x,\xi) \in D \times \Xi : \nabla_x f(x,\xi) \,\text{exists}\}.$$

We will show that $\bar{J}$ is Lebesgue measurable;

Let e_i be the ith unit vector in $\Re^n$. Denote $f(x,\xi)$ by $f_\xi(x)$. Consider

$$f_\xi^\diamond(x;e_i) = \sup_{k \in \Re^n} \{\limsup_{t \downarrow 0} [f(x+te_i+tk,\xi) - f(x+tk,\xi)]/t\}, \tag{4.2}$$

the Michel-Penot directional derivative of f_ξ at x with respect to e_i. Since $f(.,\xi)$ is continuous, we may let $t \downarrow 0$ and k only take rational values in (4.2). By (vi), $(x,\xi) \mapsto f(x+te_i+tk,\xi)$ and $(x,\xi) \mapsto f(x+tk,\xi)$ are Lebesgue measurable. Therefore, $f_\xi^\diamond(x;e_i)$, as the "countable sup limsup" of Lebesgue measurable functions of (x,ξ), is Lebesgue measurable. Similarly $f_\xi^\diamond(x;-e_i)$ is also Lebesgue measurable. So the set

$$\bar{J}_i = \{(x,\xi) \in D \times (\Re^N \setminus V) : f_\xi^\diamond(x;e_i) + f_\xi^\diamond(x;-e_i) = 0\}$$

is also Lebesgue measurable. However, by the discussion on the Michel-Penot subdifferential before this theorem, $\bar{J} = \cap_{i=1}^n \bar{J}_i$. Thus, $\bar{J}$ is also Lebesgue measurable.

Let $J(x) = \{\xi \in \Xi : (x,\xi) \notin \bar{J}\}$ for each $x \in D$. Let $D_1 = \{x \in D : m(J(x)) > 0\}$. By the Fubini Theorem, $m(D_1) = 0$. Let $D_2 = D \setminus D_1$. We now prove (f) is true for the sets D_1 and D_2.

First we show that the G-derivative $\nabla E_f^\nu(x)$ exists for $x \in D_2$. Let $x \in D_2$. Then $\nabla_x f(x,\xi)$ exists a.s. Let $h \in \Re^n$. Similarly, we may show that the map $\xi \mapsto f_\xi'(x;h)$ is Lebesgue measurable and bounded. By the dominated convergence theorem, we have the existence of $(E_f^\nu)'(x;h)$ and the equality

$$(E_f^\nu)'(x;h) = \int f_\xi'(x;h) p_\nu(\xi) d\xi = \int f_\xi'(x;h) P_\nu(d\xi). \tag{4.3}$$

But

$$f'_\xi(x;h) = < \nabla_x f(x,\xi), h >,$$

for $\xi \notin J(x)$. The existence of $\nabla E^\nu_f(x)$ follows. By (4.3),

$$\begin{aligned} |(E^\nu_f)'(x;h) - E'_f(x;h)| &\leq \int |f'_\xi(x;h)| \cdot |p_\nu(\xi) - p(\xi)| d\xi \\ &\leq L_x \|h\| \int |p_\nu(\xi) - p(\xi)| d\xi, \end{aligned} \tag{4.4}$$

where L_x is the Lipschitz constant of $f(\cdot,\xi)$ near x, which is independent of ξ by (iii). By (4.4) and (v), (3.1) holds. This proves (f).

(g). Let $x \in$ int (D). Notice that (b) still holds since it only relies on (iii). By (b), $\partial E^\nu_f(x)$ are nonempty, compact and convex sets for $\nu = 0, 1, \ldots.$ Let h be an arbitrary vector in $\Re^n$ and ϵ be a small positive number. Let $U(x)$ be a very small neighborhood of x. For any y in $U(x) \cap D_2$, by (4.4), $(E^\nu_f)'(y;h)$ tends to $E'_f(y;h)$ uniformly. Then there is a $\nu(\epsilon)$ such that for any $\nu \geq \nu(\epsilon)$ and $y \in U(x) \cap D_2$,

$$|(E^\nu_f)'(y;h) - E'_f(y;h)| \leq \epsilon. \tag{4.5}$$

By 2.5.1 of [7],

$$\partial E^\nu_f(x) = \text{conv}\{\lim \nabla E^\nu_f(y) : y \to x, y \in D_2\}, \tag{4.6}$$

for $\nu = 0, 1, \ldots.$ Then there is a sequence $y_j \to x, y_j \in U(x) \cap D_2$ such that

$$\Psi^*(h|\partial E_f(x)) = \lim_{j\to\infty} E'_f(y_j;h). \tag{4.7}$$

By (4.6),

$$\limsup_{j\to\infty} (E^\nu_f)'(y_j;h) \leq \Psi^*(h|\partial E^\nu_f(x)), \tag{4.8}$$

for $\nu = 1, \ldots.$ By (4.5), (4.7) and (4.8), for any $\nu \geq \nu(\epsilon)$,

$$\Psi^*(h|\partial E_f(x)) \leq \Psi^*(h|\partial E^\nu_f(x)) + \epsilon. \tag{4.9}$$

Similarly, by (4.6), there is a sequence $\bar{y}_j \to x, \bar{y}_j \in U(x) \cap D_2$ such that

$$\Psi^*(h|\partial E^\nu_f(x)) = \lim_{j\to\infty} (E^\nu_f)'(\bar{y}_j;h), \tag{4.10}$$

and

$$\limsup_{j\to\infty} E'_f(\bar{y}_j;h) \leq \Psi^*(h|\partial E_f(x)), \tag{4.11}$$

for $\nu = 1, \ldots$. By (4.5), (4.10) and (4.11), for any $\nu \geq \nu(\epsilon)$,

$$\Psi^*(h|\partial E_f^\nu(x)) \leq \Psi^*(h|\partial E_f(x)) + \epsilon. \tag{4.12}$$

By (4.9) and (4.12),

$$\lim_{\nu \to +\infty} \Psi^*(h|\partial E_f^\nu(x)) = \Psi^*(h|\partial E_f(x)).$$

Then (4.1) follows. This completes the proof of this theorem. ∎

Section 5. Loss Functions

We now discuss a special case of (2.1), in which

$$f(x,\xi) = u(T(\xi)x - \zeta(\xi)),$$

where $u : \Re^m \to \Re \cup \{+\infty\}$ is a loss function, $\zeta(\xi) \in \Re^m$ and $T(\xi) \in \Re^{m\times n}$. Consider

$$E_u(x) = \int u(T(\xi)x - \zeta(\xi))P(d\xi), \tag{5.1}$$

where P is a probability measure on $\Re^N$. The expectational functional (5.1) arises in stochastic program with recourse. See [5][6]. It also appears in other applications such as error optimization and optimal design. See [21]. In stochastic linear programming with recourse, u is piecewise linear and convex. In more general problems, however, u may not even be convex. This situation occurs, for example, when u is a loss if inventory, $T(\xi)x$, is less than demand, $\zeta(\xi)$. This penalty will generally become flat for very low values of $T(\xi)x - \zeta(\xi)$, voiding convexity. Nevertheless, it is useful to characterize the convergence to critical points as we do here.

Again, approximate (5.1) by

$$E_u^\nu(x) = \int u(T(\xi)x - \zeta(\xi))P_\nu(d\xi),$$

and denote E_u by E_u^0 sometimes. In particular, we consider continuous approximations as in Theorem 4.1.

Theorem 5.1 *Suppose that*

(i) *for any $x \in D$, where D is an open set in $\Re^n$, $E_u^\nu(x) < +\infty$, for $\nu = 0, 1, \ldots$;*

(ii) *u is locally Lipschitz in its support Ω, and strictly differentiable in $\Omega_1 \subseteq \Omega$ such that $\Omega_2 = \Omega \setminus \Omega_1$ is a zero Lebesgue measure set;*

(iii) *for any $\xi \in \Re^N$, there is a constant L such that $\|T(\xi)\| \leq L$;*

(iv) *for any $x \in D$, the map $\xi \mapsto T(\xi)x - \zeta(\xi)$ maps a set in $\Re^N$ to a Lebesgue zero measure set in $\Re^m$ only if this set has Lebesgue measure zero in $\Re^N$;*

(v) *$\{P_\nu, \nu = 0, 1, ...\}$ are continuous probability measures with*

$$P_\nu(d\xi) = p_\nu(\xi)d\xi,$$

where $p_\nu : \Xi \to \Re_+$, $\nu = 0, 1, 2, ...$ are piecewise continuous functions;

(vi) *there is a positive number δ and a function $\eta : \Omega \to \Re$, such that*

$$\eta(w) \geq \sup\{||d|| : d \in \partial u(\tilde{w}),\ ||\tilde{w} - w|| \leq \delta\},$$

and for any $x \in D$, $\eta(T(.)x - \zeta(.))p_\nu(.)$ is Lebesgue integrable, for $\nu = 0, 1, ...;$

(vii) *assumption (iv) of Theorem 2.2 holds;*

(viii) *either u is convex or piecewise smooth, or Theorem 4.1 (v) holds;*

Then

(a) *E_u^ν are strictly differentiable in D for $\nu = 0, 1, ...;$*

(b) *for any $x \in D$ and $x^\nu \to x$,*

$$\lim_{\nu \to +\infty} E_u^\nu(x^\nu) = E_u(x);$$

(c) *if $x^\nu \in D$ minimizes E_u^ν for each ν and x is a limiting point of $\{x^\nu\}$, then x minimizes E_u;*

(d) *for any $x \in D$,*

$$\lim_{\nu \to +\infty} \nabla E_u^\nu(x) = \nabla E_u(x). \tag{5.2}$$

Proof:

By (ii) (iii) (v) and (vii), the conditions of Theorem 2.2 hold. We have conclusions (b) and (c). It suffices now to prove (a) and (d). Suppose $x \in D$. Let

$$U = \{\bar{x} \in D : ||\bar{x} - x|| < \delta/L\}$$

and $g_\nu(\xi) = L\eta(T(\xi)x - \zeta(\xi))p_\nu(\xi)$. Then for $\bar{x}, \hat{x} \in U$, by (ii) and 2.3.7 of [7],

$$u(T(\xi)\bar{x} - \zeta(\xi)) - u(T(\xi)\hat{x} - \zeta(\xi)) \in < \partial u(\tilde{w}), T(\xi)(\bar{x} - \hat{x}) >,$$

where $\tilde{w}$ is a point in $[T(\xi)\bar{x} - \zeta(\xi), T(\xi)\hat{x} - \zeta(\xi)]$. Let $w = T(\xi)x - \zeta(\xi)$. Then,

$$||T(\xi)\bar{x} - \zeta(\xi) - w|| \leq ||T(\xi)(\hat{x} - x)|| \leq \delta,$$

$$\|T(\xi)\hat{x} - \zeta(\xi) - w\| \le \|T(\xi)(\hat{x} - x)\| \le \delta.$$

Thus, $\|\tilde{w} - w\| \le \delta$. By (iii), (v) and (vi),

$$|u(T(\xi)\bar{x} - \zeta(\xi))p_\nu(\xi) - u(T(\xi)\hat{x} - \zeta(\xi))p_\nu(\xi)| \le g_\nu(\xi)\|\bar{x} - \hat{x}\|,$$

where g_ν is Lebesgue integrable on $\Re^N$. However,

$$E_u^\nu(x) = \int u(T(\xi)x - \zeta(\xi))p_\nu(\xi)d\xi.$$

Hence, by 2.7.2 of [7], for $x \in D$, $\nu = 0, 1, \ldots$, E_u^ν is locally Lipschitz and

$$\partial E_u^\nu(x) \subseteq \int \partial u(T(\xi)x - \zeta(\xi))T(\xi)p_\nu(\xi)d\xi. \tag{5.3}$$

But the right hand side of (5.3) is singleton by (ii), (iii) and (iv). Thus $\partial E_u^\nu(x)$ is also a singleton and equality holds for (5.3). By 2.2.4 of [7], E_u^ν is strictly differentiable at x for $\nu = 0, 1, \ldots.$ This proves (a).

Now, by Corollary 2.5, Theorems 3.1 and 4.1, (5.2) follows (viii). This completes the proof of this theorem. ■

We note that Marti [19] establishes similar results for u continuous, positively homogeneous, and subadditive. We extend his results to this more general loss function. We now justify the conditions of Theorem 5.1.

If T is deterministic, then conditions (iii) and (iv) of this theorem hold naturally. Condition (ii) is equivalent to saying that u is locally Lipschitz and that the Clarke subdifferential of u is single-valued almost everywhere. Such functions are called *primal functions* in [27][28][29]. Convex functions, concave functions, differences of convex functions, regular functions and semismooth functions are all examples of primal functions. All primal functions defined on an open set form a linear space, i.e., the class of primal functions is closed in addition and scalar multiplication. For more discussion on primal functions, see [27][28][29].

As said before, in stochastic linear programming with recourse, u is a convex piecewise linear function. Thus, conditions (ii) and (viii) hold. Furthermore, condition (vi) also holds as long as the p_ν are integrable for $\nu = 0, 1, \ldots.$

For condition (v), as we said before, the piecewise linear distribution approximation suggested by Wets [32] and the linear combination of lognormal univariate distribution suggested by Dexter, Yu and Ziemba [8] are good examples of continuous approximations. In [4], we will discuss this topic more.

Theorem 5.2 *If we delete the almost everywhere strict differentiability requirement on u in Theorem 5.1, then all the conclusions still hold except conclusion (a) is changed to:*

(a) E_u^ν *are G-differentiable in D for* $\nu = 0, 1, ...$

Proof: We use the Michel-Penot subdifferential instead of the Clarke subdifferential in the proof of Theorem 5.1. Instead of invoking 2.7.2 of [7], we invoke Theorem 5.2 of [3] now. This implies that the left-hand-side of (5.3), now the Michel-Penot subdifferential of E_u^ν at x, is a singleton. By the properties of the Michel-Penot subdifferential [3], E_u^ν is G-differentiable at $x \in D$. This completes the proof. ∎

Another approach to approximate (2.1) is by using

$$E_{f^\nu}(x) = \int f^\nu(x, \xi)P(d\xi),$$

where $\{f^\nu, \nu = 1, ...\}$ is a sequence of extended real-valued functions converging pointwise to f. The convergence of E_{f^ν} to E_f was also discussed in [5] in the context of epi-convergence. See Theorem 2.7 of [5]. One may also derive results similar to Corollary 2.5, Theorems 3.1, 4.1, 5.1 and 5.2 by considering these four cases.

Section 6. Application in Algorithms

The major motivation in the development of continuous approximation schemes is for uses in algorithms. In this section, we demonstrate that some basic algorithms based on gradient and subgradient information can incorporate continuous approximations into convergent procedures. We first consider methods based on the result in Corollary 2.5. In this case, we only suppose that subgradients are available. The major result is that we obtain convergence to a value below some goal or else demonstrate that no value below the goal exists.

The algorithm is a direct modification of the subgradient method given by Polyak [25]. It can also be strengthened to allow for other step sizes as in Allen *et al.* [1]. We state the algorithm as follows. We assume that f is a convex function of x. The objective is to minimize $E_f(x)$ over $x \in D$.

Subgradient Algorithm

Step 0. Suppose sequences $\epsilon_\nu \to 0$, $\gamma_k \to 0$ such that $\sum_{k=0}^{\infty} \gamma_k = +\infty$, a goal value G, a maximum iteration count per approximation of I_{max}, and an initial point x^0. Let $k = 0, i = 0$, and $\nu = 0$.

Step 1. Pick $\zeta \in \partial E_f^\nu(x^k)$. If $||\zeta|| < \epsilon_\nu$, let $x^k = x^{k+1}$ and go to Step 2. Otherwise, let

$$x^{k+1} = x^k - \gamma_k(\zeta/||\zeta||). \tag{6.1}$$

If $E_f^\nu(x^{k+1}) < G$, go to Step 2. Otherwise, go to Step 3.

Step 2. Let $\nu = \nu + 1$, $i = 0$, and $k = k + 1$. Go to Step 1.

Step 3. Let $i = i + 1$. If $i > I_{max}$, go to Step 2. Otherwise, $k = k + 1$, go to Step 1.

We note that alternative choices for the step length parameter (here $\gamma_k/||\zeta||$) are possible but we will show convergence just for this case. The basic idea of the algorithm is to follow the subgradient algorithm steps unless a small subgradient is found, an approximate function value is less than the goal value, or neither has been found within I_{max} iterations with a single approximation. In these cases, the approximation is refined. From Corollary 2.5, we have conditions for the subdifferential sets to converge. We need an additional condition about the boundedness of the sequence of iterates:

(vii) The iterates x^k of the subgradient algorithm all belong to D and the diameter of D is a finite value, $M = \sup_{y,x \in D}\{||y - x||\}$.

This last condition can be guaranteed by adding a barrier or penalty function near the boundary of D. We assume it here for simplicity. Together the conditions lead to convergence in this algorithm as in the following theorem.

Theorem 6.1. *Suppose the conditions of Theorem 4.1 and (vii) above, then the subgradient algorithm given above produces a sequence $\{x^k\}$ such that $E_f(\bar{x}) \leq G$ for some limit point $\bar{x}$ of $\{x^k\}$ or $E_f(x) \geq G$ for all $x \in D$.*

Proof: The proof follows the same form as in a proof of the subgradient method's convergence. Suppose that there exists x^* such that $E_f(x^*) < G$ but $E_f(\bar{x}) > G$ for any limit point $\bar{x}$ of $\{x^k\}$. Since D is bounded, there exists $\bar{K}$ such that for all $k > \bar{K}$, $E_f(x^k) > G$. Otherwise, there exists $\{x^{k_i}\}$ with $E_f(x^{k_i}) \leq G$ and some $\bar{x} = \lim_i x^{k_i}$ with $E_f(\bar{x}) \leq G$. By the continuity of E_f, for all sufficiently small ϵ, there exists some δ such that for all $||x - x^*|| < \delta$, $E_f(x) \leq G - \epsilon M$. Suppose K is such that for all $k \geq K$, there exists $\eta^\nu \in \partial E_f^\nu(x^k)$ (where ν is the corresponding approximation index used by the algorithm at iteration k) with $||\eta^\nu - \eta|| < \epsilon M$ and any $\eta \in \partial E_f(x^k)$. This is always possible for any $\epsilon > 0$ by Theorem 4.1 and noting that the algorithm always increases ν if a value below G, a small subgradient, or neither condition has been found in I_{max} iterations.

Next, we suppose that $y = x^* + \delta(\eta^\nu/||\eta^\nu||)$. From this for $k > \max\{\bar{K}, K\}$, we obtain:

$$\begin{aligned} E_f(y) &\geq E_f(x^k) + < \eta, y - x^k > \\ &\geq G + < \eta - \eta^\nu, y - x^k > + < \eta^\nu, x^* - x^k > + < \eta^\nu, y - x^* > \\ &\geq G - ||\eta - \eta^\nu|| ||y - x^k|| + < \eta^\nu, x^* - x^k > + \delta ||\eta^\nu|| \\ &\geq G - \epsilon M + < \eta^\nu, x^* - x^k > + \delta ||\eta^\nu||, \end{aligned} \tag{6.2}$$

which implies that $< \eta^\nu, x^* - x^k > /||\eta^\nu|| \leq \delta$. But, we also have that $||x^{k+1} - x^*||^2 = ||x^k - x^*||^2$ $-2\gamma_k < \eta^\nu/||\eta^\nu||, x^k - x^* > + \gamma_k^2$. Hence,

$$\gamma^k(2\delta - \gamma^k) \leq -||x^{k+1} - x^*||^2 + ||x^k - x^*||^2, \tag{6.3}$$

but summing yields infinity on the left hand side of (6.3) and $||x^k - x^*||^2$ on the right, a contradiction. The result follows. ∎

This result shows that subgradient information may be sufficient to produce an algorithm that achieves an optimal value. If continuous distribution approximations are used however, we may have the results of Theorem 5.1 and convergence of derivatives. This result allows the use of any method that uses derivatives. It is contained in the following general method. Now, in the context of Theorem 5.1, we suppose that E_u has a finite minimum and that the solution set is compact.

Gradient Method

Step 0. Suppose a sequence $\epsilon^\nu \to 0$ and an initial point $x^0 \in D$.

Step 1. Follow a convergent descent algorithm for unconstrained minimization of a differentiable convex function that uses $\nabla E_u^\nu(x^k)$ to generate a new point x^{k+1} (i.e., an algorithm that generates a sequence of points with decreasing function values that either terminates with an optimal solution or has all limit points as optimal solutions to the unconstrained minimization problem such as in steepest descent). Go to Step 2.

Step 2. If $\nabla E_u^\nu(x^{k+1}) \leq \epsilon^\nu$, let $\nu = \nu + 1$, $k = k + 1$ and go to Step 1. Otherwise, let $k = k + 1$ and go to Step 1.

Theorem 6.2. *Suppose the conditions of Theorem 5.1, then the gradient method given above produces a sequence x^k such that all limit points x^* of $\{x^k\}$ minimize E_u.*

Proof: When ν is not updated, the method just follows whatever algorithm has been employed in Step 1. By the differentiability assumption, the algorithm will generate some $x^{k(\nu)}$ such that $\|\nabla E_u^\nu(x^{k(\nu)})\| < \epsilon^\nu$. Thus, the algorithm continues to refine the approximation. At each $x^{k(\nu)}$, the bound on the norm of the gradient decreases so we have that there exists some $\mathcal{N}_1$ such that for all $\nu > \mathcal{N}_1$, $\|\nabla E_u^\nu(x^{k(\nu)})\| \leq \epsilon/2$. From Theorem 5.1, for any ϵ, there exists $\mathcal{N}_2$ such that for all $\nu \geq \mathcal{N}_2$, $\|\nabla E_u^\nu(x^{k(\nu)}) - \nabla E_u(x^{k(\nu)})\| \leq \epsilon/2$. Hence, for any $\nu \geq \max\{\mathcal{N}_1, \mathcal{N}_2\}$, $\|\nabla E_u(x^{k(\nu)})\| \leq \epsilon$. Thus, any limit point x^* of $\{x^{k(\nu)}\}$ minimizes E_u. Moreover, for any k, $k(\nu) \leq k \leq k(\nu+1)$, $E_u^\nu(x^{k(\nu)}) \geq E_u^\nu(x^k)$. Thus, we also have $E_u(x^{k(\nu)}) + \delta_k \geq E_u(x^k)$ for some $\delta_k \to 0$. Hence, $E_u(x^k) \to \min E_u(x)$, proving the result. ∎

These results are given to show that continuous approximation schemes have the advantage of allowing optimization with methods that require derivatives and that the resulting algorithms may converge under suitable assumptions. Specific forms of the approximations and computational results with these procedures will be reported in a following paper [4]. We note that similar results are also reported in Wang [31] for several algorithms. His procedure relies on the epiconvergence of E_f^ν to E_f and requires descent when using subgradients. Since this may not occur at some approximation, refinement may be necessary before optimality at that approximation and before any steps are possible. Our procedure allows progress at each approximation either through differentiability or through the subgradient method.

REFERENCES

[1] E. Allen , R. Helgason, J. Kennington and B. Shetty, "A generalization of Polyak's convergence result for subgradient optimization," *Mathematical Programming* 37 (1987) 309-317.

[2] J.R. Birge and L. Qi, "Computing block-angular Karmarkar projections with applications to stochastic programming," *Management Science* 34 (1988) 1472-1479.

[3] J.R. Birge and L. Qi, "The Michel-Penot subdifferential and stochastic programming," Applied Mathematics Preprint 89/12, School of Mathematics, The University of New South Wales, 1989.

[4] J.R. Birge and L. Qi, "Continuous approximation schemes for stochastic programs," in preparation.

[5] J.R. Birge and R.J-B. Wets, "Designing approximation schemes for stochastic optimization problems, in particular, for stochastic programs with recourse," *Mathematical Programming Study 27* (1986) 54-102.

[6] J.R. Birge and R.J-B. Wets, "Sublinear upper bounds for stochastic programs with recourse," *Mathematical Programming* 43 (1989) 131-149.

[7] F.H. Clarke, **Optimization and Nonsmooth Analysis**, Wiley, New York, 1983.

[8] A.S. Dexter, J.N.W. Yu, and W. T. Ziemba, "Portfolio selection in a lognormal market when the inversot has a power utility function : computationals results," in: **Stochastic Programming**, ed. by M.A.H. Dempster, Academic Press, New York, 1980, pp. 507-523.

[9] J. Dupačová and R.J-B. Wets, "Asymptotic behavior of statistical estimators adn of optimal solutions of stochastic optimization problems," *Annals of Statistics* 16 (1988) 1517-1549.

[10] Y. Ermoliev, "Stochastic quasigradient methods and their applications to systems optimization," *Stochastics* 9 (1983) 1–36.

[11] Y. Ermoliev and R.J-B. Wets, **Numerical Techniques in Stochastic Programming**, Springer-Verlag, Berlin, 1988.

[12] K. Frauendorfer, "Solving SLP recourse problem with arbitrary multivariate distributions - the dependent case," *Mathematics of Operations Research* 13 (1988) 377 - 394.

[13] H. Gassmann, "Conditional probability and conditional expectation of a random variable", in: **Numerical Techniques in Stochastic Programming**, eds. by Y. Ermoliev and R. Wets, Springer-Verlag, Berlin, 1988, pp. 237-254.

[14] J.L. Higle and S. Sen, "Statistical verification of optimality conditions for stochastic programs with recourse," Working Paper 89-028, Department of Systems and Industrial Engineering, University of Arizona (Tucson, AZ 85721, November 1989).

[15] P. Kall, "On approximations and stability in stochastic programming," in: **Parametric Optimization and Related Topics**, eds. by J. Guddat, et al., Akademie-Verlag, Berlin, 1987, pp. 387-407.

[16] P. Kall, A. Ruszczyński and K. Frauendorfer, "Approximation technicales in stochastic programming", in: **Numerical Techniques in Stochastic Programming**, eds. by Y. Ermoliev and R. Wets, Springer-Verlag, Berlin, 1988, pp. 33-64.

[17] A. J. King and R.J-B. Wets, "Epi-consistency of convex stochastic programs," *Stochastics and Stochastics Reports* 34 (1991) 83-92.

[18] R. Lepp, "Approximations to stochastic programs with complete recourse," *SIAM J. Control and Optimization* 28 (1990) 382–394.

[19] K. Marti, "Approximationen von Entscheidungsproblemen mit linearer Ergebnisfunktion und positiv homogener, subadditiver Verlusfunktion," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 31 (1975) 203–233.

[20] K. Marti, "Computation of descent directions and efficient points in stochastic optimization problem with invariant distributions," *ZAMM* 65 (1985) 132-156.

[21] K. Marti and E. Fuchs, "Computation of descent directions and efficient points in stochastic optimization problems without using derivatives," *Mathematical Programming Study 28* (1986) 132-156.

[22] 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.

[23] J. L. Nazareth and R.J-B. Wets, "Algorithms for stochastic programs: the case of nonstochastic tenders," *Mathematical Programming Study* 28 (1986) 1–28.

[24] J. L. Nazareth and R.J-B. Wets, "Nonlinear programming techniques applied to stochastic programs with recourse," in: **Numerical Techniques in Stochastic Programming**, eds. by Y. Ermoliev and R. Wets, Springer-Verlag, Berlin, 1988, pp. 95-119.

[25] B.T. Polyak, "Subgradient methods: A survey of Soviet research," in: **Nonsmooth Optimization**, eds. by C. Lemarechal and R. Mifflin, Pergaman Press, Oxford, 1978, pp. 5-29.

[26] A. Prékopa and R.J-B. Wets, **Stochastic Programming 84**, *Mathematical Programming Study* 27 & 28 (1986).

[27] L. Qi, "The maximal normal operator space and integration of subdifferentials of nonconvex functions", *Nonlinear Analysis, Theory Methods and Applications* 13 (1989) 1003-1011.

[28] L. Qi, "Semismoothness and decomposition of maximal normal operators", *Journal of Mathematical Analysis & Applications* 146 (1990) 271-279.

[29] L. Qi, "Quasidifferentials and maximal normal operators", *Mathematical Programming* 49 (1991) 263-271.

[30] S. M. Robinson and R.J-B. Wets,"Stability in two-stage stochastic programming," *SIAM J. Control and Optimization* 25 (1987) 1409-1416.

[31] J. Wang, "Approximate nonlinear programming algorithms for solving stochastic programs with recourse," *Annals of Operations Research*, to appear.

[32] R.J-B. Wets, "Solving stochastic programs with simple recourse, II," in: *Proceedings of 1975 Conference on Information Sciences and Systems* (John Hopkins University, Baltimore, Maryland, 1975).

[33] R.J-B. Wets, "Convergence of convex functions, variational inequalities and convex optimization problems," in: **Variational Inequalities and Complementarity Problems**, eds. by R.W. Cottle, F. Giannessi and J.-L. Lions, John Wiley, New York, 1980, pp. 375-404.

[34] R.J-B. Wets, "Stochastic programming: Solution techniques and approximation schemes," in: **Mathematical Programming: The State of the Art - Bonn 1982**, eds. by A. Bachem, M. Grötschel

and B. Korte, Springer-Verlag, Berlin, 1983, pp. 566-603.

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

[36] R.J-B. Wets, "Stochastic programming," in: **Handbooks in Operations Research and Management Science, Volume 1: Optimization**, eds. by G.L. Nemhauser, A.H.G. Rinnooy Kan, and M.J. Todd, North-Holland, Amsterdam, 1989, pp. 573-630.