

Stochastic Optimization Problems with Nondifferentiable Cost Functionals¹

D. P. BERTSEKAS²

Communicated by P. Varayia

Abstract. In this paper, we examine a class of stochastic optimization problems characterized by nondifferentiability of the objective function. It is shown that, in many cases, the expected value of the objective function is differentiable and, thus, the resulting optimization problem can be solved by using classical analytical or numerical methods. The results are subsequently applied to the solution of a problem of economic resource allocation.

1. Introduction

In this paper, we examine some questions which are important for the solution of a certain class of stochastic optimization problems. This class of problems is characterized by a cost functional of the form

$$F(x) = E\{f(x, \omega)\}, \quad (1)$$

where the vector x belongs to the n -dimensional Euclidean space R^n , ω is an uncertain quantity-element of a probability space, and $E\{\cdot\}$ denotes the expectation operation.

The problem of minimizing the cost functional F of Eq. (1) subject to constraints on the vector x can be viewed, of course, as a deterministic optimization problem. In the case where the function $f(\cdot, \omega)$ is a differentiable function of x for each ω , it can be shown under quite general assumptions that the gradient of the function F exists for each x and is given by

$$\nabla F(x) = E\{\nabla f(x, \omega)\}. \quad (2)$$

¹ This work was supported by the National Science Foundation under Grant No. GK 29237.

² Assistant Professor, Department of Engineering-Economic Systems, Stanford University, Stanford, California.

Thus, for this case, the well-known gradient-based analytical methods (such as the Kuhn–Tucker theorem), and numerical methods (such as steepest descent, conjugate gradient, methods of feasible directions, etc.) can be used effectively for the solution of the optimization problem for a quite general class of constraints on x .

The purpose of this paper is to consider optimization problems with a cost functional of the form (1) where the function $f(\cdot, \omega)$ is nondifferentiable for some values of ω and to examine the differentiability properties of the function F under these circumstances. It is shown that, in many cases, the function F is differentiable even when the function $f(\cdot, \omega)$ is nondifferentiable for some ω , and an explicit characterization is given for the gradient of F . In these cases, the classical methodology of optimization theory for differentiable cost functionals can be used for the solution.

Stochastic optimization problems where the function $f(\cdot, \omega)$ is nondifferentiable have not been considered in the literature, to the author's knowledge, although their deterministic counterparts have received considerable attention recently (Refs. 1–10). Such problems arise often in practice, particularly in economic applications. A typical case is examined in this paper in some detail as an illustration of the use of the results obtained.

In Section 2, we develop the main theoretical results of this paper for the case where the function $f(\cdot, \omega)$ is a real-valued convex function. It is shown, under quite general assumptions, that the function F of Eq. (1) is real-valued and convex with a directional derivative at a point $x \in R^n$ with respect to a vector $y \in R^n$ given by

$$F'(x; y) = E\{f'(x, \omega; y)\}. \tag{3}$$

The equivalent equation which characterizes the subdifferential (Ref. 11) $\partial F(x)$ of the convex function F at a point x in terms of the subdifferential $\partial f(x, \omega)$ of the function $f(\cdot, \omega)$ at the point x is

$$\partial F(x) = E\{\partial f(x, \omega)\}, \tag{4}$$

where the integral $E\{\partial f(x, \omega)\}$ of the set-valued function $\partial f(x, \cdot)$ is appropriately interpreted in what follows. Equation (4) is a direct generalization of Eq. (2), since the subdifferentials $\partial F(x)$, $\partial f(x, \omega)$ are sets consisting of single points, the gradients $\nabla F(x)$, $\nabla f(x, \omega)$, whenever the functions F and $f(\cdot, \omega)$ are differentiable at the point x for all values of ω . We also derive necessary and sufficient conditions for the function F to be differentiable at a point in terms of the given probability measure, and we demonstrate that, in many cases, likely to be encountered in

practice, the function F is in fact differentiable, and an explicit equation is given for the calculation of its gradient. As a result, methods which are valid for the solution of optimization problems with differentiable cost functionals can be directly applied in these cases. The specific application that we consider, the problem of production planning to meet a stochastic demand, falls in this category, under quite general assumptions on the probability measure characterizing the uncertainty. This problem is examined in some detail in Section 3.

2. Main Results

The questions that we shall consider in this section are related to the problem of minimizing, subject to some constraints, the function $F: R^n \rightarrow R$ given by

$$F(x) = E\{f(x, \omega)\} = \int_{\Omega} f(x, \omega) dP(\omega), \quad (5)$$

where $x \in R^n$ (n -dimensional Euclidean space), ω is an element of a probability space (Ω, \mathcal{F}, P) , $f: R^n \times \Omega \rightarrow R$ is a given real-valued function, and $E\{\cdot\}$ denotes the expectation operation. The set Ω is arbitrary, \mathcal{F} is a σ -algebra of subsets of Ω , and P is a probability measure defined on \mathcal{F} . In the applications considered in this paper, as well as in most practical situations, the set Ω is the Euclidean space R^m , the σ -algebra \mathcal{F} consists of the Borel sets of R^m , and the probability measure P is characterized by a probability distribution defined on R^m .

Concerning the function f , we assume that $f(\cdot, \omega)$ is a real-valued convex function of x for each $\omega \in \Omega$ and that $f(x, \cdot)$ is a P -integrable function of ω for each $x \in R^n$, i.e., we assume that $E\{|f(x, \omega)|\} < \infty$ for each $x \in R^n$. Under the above assumptions, the function F of Eq. (5) is well defined as a real-valued function on R^n .

Our primary concern in this section is to investigate the differentiability properties of the function F , particularly in the case where the function $f(\cdot, \omega)$ is not necessarily everywhere differentiable for some values of ω . First, let us establish the convexity of the function F .

Lemma 2.1. The function $F: R^n \rightarrow R$ of Eq. (5) is convex.

Proof. By using the convexity of $f(\cdot, \omega)$, we have, for any $x_1, x_2 \in R^n$, $0 \leq a \leq 1$, $\omega \in \Omega$,

$$f[ax_1 + (1-a)x_2, \omega] \leq af(x_1, \omega) + (1-a)f(x_2, \omega),$$

from which it follows that

$$E\{f[ax_1 + (1 - a)x_2, \omega]\} \leq aE\{f(x_1, \omega)\} + (1 - a)E\{f(x_2, \omega)\}.$$

We now have

$$\begin{aligned} F[ax_1 + (1 - a)x_2] &= E\{f[ax_1 + (1 - a)x_2, \omega]\} \\ &\leq aE\{f(x_1, \omega)\} + (1 - a)E\{f(x_2, \omega)\} \\ &= aF(x_1) + (1 - a)F(x_2). \end{aligned}$$

It is known (Ref. 11) that any convex function $g: R^n \rightarrow R$ is continuous and its directional derivative at a point x with respect to a vector y , defined by

$$g'(x; y) = \lim_{\lambda \rightarrow 0^+} \{[g(x + \lambda y) - g(x)]/\lambda\}, \tag{6}$$

exists for any $x, y \in R^n$. Furthermore, for any $x \in R^n$, $g'(x; \cdot)$ is a continuous convex real-valued function of y . It can be shown also (Ref. 11) that, for any given $x \in R^n$ and any sequence $\{\lambda_n\}$ such that $\lambda_n \rightarrow 0$, $\lambda_n \geq \lambda_{n+1} \geq \dots > 0$, we have

$$[g(x + \lambda_n y) - g(x)]/\lambda_n \geq [g(x + \lambda_{n+1} y) - g(x)]/\lambda_{n+1} \geq \dots \geq g'(x; y) \tag{7}$$

and, of course,

$$\lim_{n \rightarrow \infty} [g(x + \lambda_n y) - g(x)]/\lambda_n = g'(x; y). \tag{8}$$

By applying Eq. (8) to the convex function $f(\cdot, \omega)$, we have that $f'(x, \omega; y)$ is a measurable function of ω for every $x, y \in R^n$, since it is the pointwise limit of a sequence of measurable functions. Furthermore, we have the following proposition.

Proposition 2.1. The directional derivative of the convex function F is given for any $x, y \in R^n$ by the equation

$$F'(x; y) = E\{f'(x, \omega; y)\} = \int_{\Omega} f'(x, \omega; y) dP(\omega). \tag{9}$$

Proof. We have, by Eqs. (5), (8),

$$\begin{aligned} F'(x; y) &= \lim_{n \rightarrow \infty} \{[F(x + \lambda_n y) - F(x)]/\lambda_n\} \\ &= \lim_{n \rightarrow \infty} \int_{\Omega} \{[f(x + \lambda_n y, \omega) - f(x, \omega)]/\lambda_n\} dP(\omega). \end{aligned} \tag{10}$$

Consider now the sequence of functions

$$f_n(\omega) = [f(x + \lambda_n y, \omega) - f(x, \omega)]/\lambda_n, \quad n \geq 1.$$

Clearly, $f_n(\omega)$ are integrable functions, since $f(x, \cdot)$ is an integrable function of ω for each $x \in R^n$. Furthermore, we have

$$\lim_{n \rightarrow \infty} f_n(\omega) = f'(x, \omega; y)$$

for each $\omega \in \Omega$ and

$$f_n(\omega) \geq f_{n+1}(\omega) \geq \dots \geq f'(x, \omega; y)$$

for all n by (7). Therefore, by the monotone convergence theorem (Refs. 12–14), we have

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n(\omega) dP(\omega) = \int_{\Omega} f'(x, \omega; y) dP(\omega)$$

and, substituting in (10), we obtain

$$F'(x; y) = \int_{\Omega} f'(x, \omega; y) dP(\omega) = E\{f'(x, \omega; y)\}.$$

The expression (9) for the directional derivative of the function F can be used to obtain an expression for the subdifferential of F . For any real-valued convex function $g: R^n \rightarrow R$, the subdifferential of g at a point x is defined as the set

$$\partial g(x) = \{x^* \mid g(z) \geq g(x) + \langle z - x, x^* \rangle, \forall z \in R^n\}. \quad (11)$$

It can be shown (Ref. 11) that the set $\partial g(x)$ is nonempty, convex, and compact, and its support function is the directional derivative $g'(x; y)$ given by

$$g'(x; y) = \max_{x^* \in \partial g(x)} \langle x^*, y \rangle, \quad (12)$$

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product in R^n . The subdifferential $\partial g(x)$ generalizes the notion of the ordinary gradient. If g is differentiable at a point x , then $\partial g(x)$ consists of a single point, the gradient $\nabla g(x)$. The properties of subdifferentials are discussed in great detail in Ref. 11. Consider now, for any $x \in R^n$, $\omega \in \Omega$, the subdifferential $\partial f(x, \omega)$. For any fixed x , the mapping which assigns to $\omega \in \Omega$ the compact set $\partial f(x, \omega)$ is a set-valued mapping; in fact, it is a measurable set-valued mapping according to the terminology of Ref. 15 (see Ref. 15, Corollary 4.6). Measurable set-valued mappings have received considerable attention in the mathematical literature (see Ref. 15–17 and the references quoted therein). For example, the integral of some measurable set-valued

mappings has been considered in Refs. 16–17. Here, we define for a fixed $x \in R^n$ the integral of the mapping $\omega \rightarrow \partial f(x, \omega)$ as the subset of R^n

$$\begin{aligned}
 E\{\partial f(x, \omega)\} &= \int_{\Omega} \partial f(x, \omega) dP(\omega) \\
 &= \left\{ x^* \in R^n \mid x^* = \int_{\Omega} x^*(\omega) dP(\omega), x^*(\cdot): \text{measurable}, x^*(\omega) \in \partial f(x, \omega) \text{ a.e.} \right\},
 \end{aligned}
 \tag{13}$$

where the initials a.e. denote almost everywhere. Notice that, by Eq. (12), for each $\omega \in \Omega, y \in R^n$, we have

$$|\langle x^*(\omega), y \rangle| \leq \max\{f'(x, \omega; y), f'(x, \omega; -y)\}
 \tag{14}$$

whenever $x^*(\omega) \in \partial f(x, \omega)$. As a result, since the functions $f'(x, \omega; y), f'(x, \omega; -y)$ are integrable for each $x, y \in R^n$, we have that the function $x^*(\cdot): \Omega \rightarrow R^n$ is integrable whenever it is measurable and $x^*(\omega) \in \partial f(x, \omega)$ a.e., in the sense that each coordinate of $x^*(\cdot)$ is an integrable function. Therefore, the set $E\{\partial f(x, \omega)\}$ of Eq. (13) is well defined. Furthermore, $E\{\partial f(x, \omega)\}$ is nonempty by Corollary 1.1 of Ref. 15. We now have the following proposition.

Proposition 2.2. The subdifferential of the function F at a point x is given by

$$\partial F(x) = E\{\partial f(x, \omega)\},
 \tag{15}$$

where the set $E\{\partial f(x, \omega)\}$ is defined in Eq. (13).

Proof. Let y be any fixed vector in R^n . Since the mapping $\omega \rightarrow \partial f(x, \omega)$ is measurable and $f(x, \omega) = \langle y, x \rangle$ is a normal convex integrand (according to the terminology of Ref. 15), it follows from Corollary 4.3 of Ref. 15 that there exists a measurable function $\bar{x}^*(\cdot)$, with $\bar{x}^*(\omega) \in \partial f(x, \omega)$ for all $\omega \in \Omega$, such that

$$f'(x, \omega; y) = \max_{x^*(\omega) \in \partial f(x, \omega)} \langle y, x^*(\omega) \rangle = \langle y, \bar{x}^*(\omega) \rangle.$$

This fact, together with Proposition 2.1 and Eq. (12), implies that

$$\begin{aligned}
 \max_{x^* \in \partial F(x)} \langle y, x^* \rangle &= \int_{\Omega} f'(x, \omega; y) dP(\omega) \\
 &= \int_{\Omega} \max_{x^*(\omega) \in \partial f(x, \omega)} \langle y, x^*(\omega) \rangle dP(\omega) \\
 &= \max_{\substack{x^*(\omega) \in \partial f(x, \omega) \text{ a.e.} \\ x^*(\cdot): \text{measurable}}} \left\langle y, \int_{\Omega} x^*(\omega) dP(\omega) \right\rangle \\
 &= \max_{x^* \in E\{\partial f(x, \omega)\}} \langle y, x^* \rangle, \quad \forall y \in R^n.
 \end{aligned}$$

Therefore, we have (Ref. 11)

$$\partial F(x) = \text{closure } [E\{\partial f(x, \omega)\}],$$

since the set $\partial F(x)$ is convex and compact and the set $E\{\partial f(x, \omega)\}$ is convex by the convexity of $\partial f(x, \omega)$ for each $\omega \in \Omega$. Thus, in order to prove that $\partial F(x) = E\{\partial f(x, \omega)\}$, it is sufficient to prove that $E\{\partial f(x, \omega)\}$ is a closed set.

Let $x_n^* \rightarrow x^*$ be a convergent sequence in R^n , with $x_n^* \in E\{\partial f(x, \omega)\}$ for all n . Then, there exists a sequence of integrable functions $x_n^*(\cdot)$, with $x_n^*(\omega) \in \partial f(x, \omega)$ a.e. and such that $x_n^* = \int_{\Omega} x_n^*(\omega) dP(\omega)$. These functions (or more accurately their equivalence classes) can be viewed as elements of the space $L_1^n(\Omega, \mathcal{F}, P)$ of P -integrable functions $z: \Omega \rightarrow R^n$ with the norm

$$\|z\| = \int_{\Omega} \{|z_1(\omega)| + \dots + |z_n(\omega)|\} dP(\omega),$$

where z_1, \dots, z_n are the n coordinate functions of z . Furthermore, the sequence $\{x_n^*(\cdot)\}$ belongs to the subset C of $L_1^n(\Omega, \mathcal{F}, P)$ characterized by the property

$$z(\cdot) \in C \quad \text{iff} \quad z(\omega) \in \partial f(x, \omega) \text{ a.e. .}$$

By using Eq. (14), it is easy to verify the assumption of Theorem IV.8.9 of Ref. 13 and conclude that C is a weakly sequentially compact subset of $L_1^n(\Omega, \mathcal{F}, P)$. Therefore, the sequence $\{x_n^*(\cdot)\}$ contains a subsequence $\{x_{n_k}^*\}$ which converges weakly to a function $x^*(\cdot)$ in $L_1^n(\Omega, \mathcal{F}, P)$. Furthermore,

$$x^* = \lim_{n \rightarrow \infty} x_n^* = \lim_{n \rightarrow \infty} \int_{\Omega} x_n^*(\omega) dP(\omega) = \int_{\Omega} x^*(\omega) dP(\omega).$$

We shall show that $x^*(\omega) \in \partial f(x, \omega)$ a.e., and hence $x^* \in E\{\partial f(x, \omega)\}$. Since $\{x_{n_k}^*(\cdot)\}$ converges to $x^*(\cdot)$ weakly, some sequence of convex combinations of $x_{n_k}^*(\cdot)$ converges strongly to $x^*(\cdot)$ (see Ref. 13, Corollary V.3.14). Let $\{z_n(\cdot)\}$ be this sequence. We have $z_n(\omega) \in \partial f(x, \omega)$ a.e. for all n by the convexity of $\partial f(x, \omega)$. Since the sequence $\{z_n(\cdot)\}$ converges strongly to $x^*(\cdot)$, it also converges in measure, and therefore a subsequence $\{z_{n_k}(\cdot)\}$ converges to $x^*(\cdot)$ almost everywhere (Refs. 12, 14). Since $z_{n_k}(\omega) \in \partial f(x, \omega)$ a.e., we conclude that $x^*(\omega) \in \partial f(x, \omega)$ a.e. Therefore, $x^* = \int_{\Omega} x^*(\omega) dP(\omega) \in E\{\partial f(x, \omega)\}$, and the set $E\{\partial f(x, \omega)\}$ is closed.

We turn now our attention to deriving necessary and sufficient conditions for the function F to be differentiable at a point. It is clear

from Proposition 2.2 that a sufficient condition for the function F to be differentiable at a point x is that the gradient $\nabla f(x, \omega)$ exists for every $\omega \in \Omega$, except for a subset of Ω with probability zero. The following proposition shows that in fact this condition is also necessary.

Loosely speaking, one can interpret this proposition by saying that, whenever the set of points of discontinuity of the gradient $\nabla f(x, \omega)$ has probability zero, these discontinuities are averaged out.

Proposition 2.3. For every $x \in R^n$, the set

$$D_\omega(x) = \{\omega \in \Omega \mid f(\cdot, \omega) \text{ is nondifferentiable at } x\} \tag{16}$$

is measurable, i.e., $D_\omega(x) \in \mathcal{F}$. Furthermore, F is differentiable at the point x iff $P[D_\omega(x)] = 0$. In this case, we have

$$\nabla F(x) = \int_{\Omega - D_\omega(x)} \nabla f(x, \omega) dP(\omega), \tag{17}$$

where

$$[\Omega - D_\omega(x)] = \{\omega \in \Omega \mid \omega \notin D_\omega(x)\}.$$

Proof. We have (Ref. 11)

$$\begin{aligned} D_\omega(x) &= \{\omega \in \Omega \mid f'(x, \omega; y) + f'(x, \omega; -y) > 0 \text{ for some } y \in R^n\} \\ &= \{\omega \in \Omega \mid \max_{\|y\|=1} [f'(x, \omega; y) + f'(x, \omega; -y)] > 0\}. \end{aligned}$$

Since $[f'(x, \omega; y) + f'(x, \omega; -y)]$ is a continuous function of y for any fixed $x \in R^n, \omega \in \Omega$, we have

$$\begin{aligned} D_\omega(x) &= \{\omega \mid \max_{\|y\|=1} [f'(x, \omega; y) + f'(x, \omega; -y)] > 0\} \\ &= \{\omega \mid \sup_n [f'(x, \omega; y_n) + f'(x, \omega; -y_n)] > 0\}, \end{aligned} \tag{18}$$

where $\{y_n\}$ is a countable set, dense in $\{y \mid \|y\| = 1\}$. But the function $\sup_n [f'(x, \cdot; y_n) + f'(x, \cdot; -y_n)]$ is a measurable function of ω , since it is the pointwise supremum of measurable functions. Therefore, from (18), we have that $D_\omega(x)$ is measurable.

Assume now that F is differentiable at a point $x \in R^n$. Then, for every $y \in R^n$, we have

$$F'(x; y) + F'(x; -y) = \int_{\Omega} [f'(x, \omega; y) + f'(x, \omega; -y)] dP(\omega) = 0.$$

Since

$$f'(x, \omega; y) + f'(x, \omega; -y) \geq 0, \quad \forall \omega \in \Omega,$$

we conclude that, for each $y \in R^n$, the set

$$D_\omega(x, y) = \{\omega \in \Omega \mid f'(x, \omega; y) + f'(x, \omega; -y) > 0\}$$

has measure zero, i.e.,

$$P[D_\omega(x, y)] = 0, \quad \forall y \in R^n.$$

Now we have

$$\begin{aligned} D_\omega(x) &= \{\omega \in \Omega \mid \sup_n [f'(x, \omega; y_n) + f'(x, \omega; -y_n)] > 0\} \\ &= \bigcup_n \{\omega \in \Omega \mid f'(x, \omega; y_n) + f'(x, \omega; -y_n) > 0\} = \bigcup_n D_\omega(x, y_n). \end{aligned}$$

Hence,

$$P[D_\omega(x)] \leq \sum_{i=1}^{\infty} P[D_\omega(x, y_i)] = 0,$$

and

$$P[D_\omega(x)] = 0.$$

The fact that $P[D_\omega(x)] = 0$ implies that F is differentiable at the point x follows directly from Proposition 2.2. Equation (17) is an obvious consequence of Proposition 2.2.

From Proposition 2.3, we can conclude that, whenever we have $P[D_\omega(x)] = 0$ for every $x \in R^n$, the function F is everywhere differentiable; therefore, for the solution of the problem of minimizing the function F subject to constraints, we can use classical analytical or numerical methods. Furthermore, Eq. (17) gives an explicit characterization of the gradient of F . The calculation of this gradient requires a numerical integration which, for most cases, presents no more difficulty than the integration which would be required if the function $f(\cdot, \omega)$ were everywhere differentiable for each $\omega \in \Omega$. The condition $P[D_\omega(x)] = 0$ is satisfied in many practical situations for large classes of probability measures. The following lemma can be useful in some cases for verifying the condition $P[D_\omega(x)] = 0$. The proof of this lemma appears in Ref. 11, Theorems 25.4 and 25.5.

Lemma 2.2. If $g: R^n \rightarrow R$ is a convex function, the set of points where g is nondifferentiable is a countable union of closed sets and has Lebesgue measure zero.

Lemma 2.2 can be used to prove differentiability in several special cases of cost functionals. As an example, we prove the following proposition which will be useful in Section 3.

Proposition 2.4. Let $f: R^n \rightarrow R$ be a convex function, and consider the function

$$F(x) = E\{f(x - \omega)\},$$

where ω belongs to the probability space (R^n, B_n, P) , where B_n is the σ -algebra of Borel sets of R^n and P is a probability measure on B_n which is absolutely continuous with respect to Lebesgue measure restricted on B_n . Then, if $E\{|f(x - \omega)|\} < \infty$ for all $x \in R^n$, the function F is everywhere differentiable.

Proof. For a fixed vector $x \in R^n$, the set $D_\omega(x)$ is a translation of the set D of points where f is nondifferentiable. Therefore, by Lemma 2.2, $D_\omega(x)$ is a Borel set with Lebesgue measure zero. By the absolute continuity of P , we have that $P[D_\omega(x)] = 0$ for all $x \in R^n$. By using Proposition 2.3, the result follows.

It should be noted that the Proposition 2.4 can form the basis for developing an algorithm for the numerical solution of convex (or concave) deterministic optimization problems with nondifferentiable cost functionals subject to constraints. The nondifferentiable cost functional $f(x)$ is approximated by the differentiable cost functional $F(x) = E\{f(x - \omega)\}$, where ω belongs to a probability space with a probability measure which is absolutely continuous with respect to Lebesgue measure. For example, the probability distribution of ω can be selected to be uniform over an n -dimensional cube with nonempty interior. The approximate problem can be solved using standard techniques, and its solution can approximate the solution of the original problem as closely as desired by a suitable choice of the probability distribution of ω . This procedure is beset by the fact that a numerical integration must be performed for evaluating both the gradient and the value of the function F , and its practical applicability remains a subject for further investigation. However, in view of the limited existing methodology for solving numerically optimization problems with nondifferentiable cost functionals (Refs. 3, 4), it may prove useful in some practical cases and particularly in those cases where the number of variables with respect to which the cost functional is nondifferentiable is small.

Proposition 2.4 is also useful for the solution of the following stochastic optimal control problem. Consider the linear system

$$x_{k+1} = A_k x_k + B_k u_k + w_k, \quad k = 0, 1, \dots, N - 1,$$

where $x_k \in R^n$ is the state vector, $u_k \in R^m$ is the control vector, $w_k \in R^n$ is an input disturbance vector, and the matrices A_k , B_k and the initial condition x_0 are given. The vectors w_0, w_1, \dots, w_{N-1} are random vectors with given joint probability distribution. It is required to find a control sequence u_0, u_1, \dots, u_{N-1} which minimizes, subject to constraints $u_k \in U_k \subset R^m$, $k = 0, 1, \dots, N-1$, the cost functional

$$J(u_0, u_1, \dots, u_{N-1}) = E \left\{ \sum_{i=1}^N [f_i(x_i) + g_{i-1}(u_{i-1})] \right\},$$

where $f_i: R^n \rightarrow R$ are given convex functions and $g_{i-1}: R^m \rightarrow R$ are given functions. Let us define the vectors y_k by the equation

$$\begin{aligned} y_k &= x_k - w_k, & k &= 0, 1, \dots, N, \\ \omega_k &= A_{k-1} \dots A_1 w_0 + \dots + A_{k-1} w_{k-2} + w_{k-1}, & k &= 1, \dots, N, \\ \omega_0 &= 0. \end{aligned}$$

Then, y_k satisfies the state equation

$$\begin{aligned} y_{k+1} &= A_k y_k + B_k u_k, & k &= 0, 1, \dots, N-1. \\ y_0 &= x_0, \end{aligned}$$

and the cost functional is written as

$$\begin{aligned} J(u_0, u_1, \dots, u_{N-1}) &= \sum_{i=1}^N [F_i(y_i) + g_{i-1}(u_{i-1})], \\ F_i(y_i) &= E\{f_i(y_i + \omega_i)\}. \end{aligned}$$

Thus, the problem has been converted to a deterministic optimal control problem. For the solution of this problem, one can use classical methods whenever the functions F_i are differentiable. By Proposition 2.4, this will occur whenever the given probability distribution is such that the resulting probability measure for each of ω_i is absolutely continuous with respect to Lebesgue measure, even when the functions f_i are non-differentiable at some points.

Finally, we would like to mention the clear fact that all the results of this section can be generalized to the case where the space (Ω, \mathcal{F}, P) is any positive measure space rather than a probability space.

3. Problem of Production Planning to Meet a Stochastic Demand

Stochastic optimization problems with a cost functional characterized by nondifferentiability occur frequently in economics in resource

allocation problems where the return or profit rate is discontinuous with respect to resource allocated. The problem that we describe in this section is a typical example of this situation.

Consider a firm which is planning to produce x_i , $i = 1, \dots, n$, quantities of n products to meet a demand ω_i , $i = 1, \dots, n$, where ω_i are random variables with given joint probability distribution and expected values $\bar{\omega}_i$. The production is subject to constraints, resulting from limited resources, which we shall denote by

$$g_j(x_1, \dots, x_n) \leq 0, \quad j = 1, 2, \dots, m. \tag{19}$$

Assume that the firm will make a profit a_i per unit of product i sold and will incur a loss b_i per unit of product i which will remain unsold. Then, it can be easily seen that the expected value of the profit of the firm is

$$E\{\text{profit}\} = \sum_{i=1}^n a_i \bar{\omega}_i + E \left\{ \sum_{i=1}^n f_i(x_i - \omega_i) \right\}, \tag{20}$$

where the functions f_i , $i = 1, \dots, n$, are defined by

$$f_i(z) = \begin{cases} -b_i z & \text{for } z \geq 0, \\ a_i z & \text{for } z \leq 0. \end{cases}$$

The problem is to find the quantities x_1, \dots, x_n which maximize the expected profit, subject to the production constraints (19).

Let us denote by $x \in R^n$ the n -tuple $[x_1, \dots, x_n]$ and by $\omega \in R^n$ the n -tuple $[\omega_1, \dots, \omega_n]$, and consider the convex function $f: R^n \rightarrow R$ given by

$$f(z) = -\sum_{i=1}^n f_i(z_i). \tag{21}$$

Then, maximizing the expected profit (20) is equivalent to minimizing

$$F(x) = E\{f(x - \omega)\}, \tag{22}$$

subject to the production constraints (19). It can be seen that the function f is nondifferentiable and that the cost functional F is of the form considered in Proposition 2.4. By using the results of the proposition, we obtain that, if the probability distribution of ω is such that the resulting probability measure on the Borel sets of R^n is absolutely continuous with respect to Lebesgue measure, then the function F is, in fact, differentiable. This is equivalent to requiring that the random variables $\omega_1, \dots, \omega_n$ have a joint probability density function $p(\omega_1, \dots, \omega_n)$ which is such that

$$\int_N p(\omega_1, \dots, \omega_n) d\omega_1, \dots, d\omega_n = 0$$

for every Borel set N with measure zero. In a practical situation, one can always modify the given probability measure to satisfy this condition, while approximating the function F arbitrarily closely. The gradient of the function F at any point x is given by

$$\nabla F(x) = \int_{\bar{D}_\omega(x)} \nabla f(x - \omega) dP(\omega),$$

where $\bar{D}_\omega(x)$ is the subset of R^n

$$\bar{D}_\omega(x) = \{\omega \in R^n \mid \omega_1 \neq x_1, \dots, \omega_n \neq x_n\}. \quad (23)$$

It should be noted that the condition that the probability measure be absolutely continuous with respect to Lebesgue measure is only sufficient for the differentiability of the function F of Eq. (22). By Proposition 2.3, a necessary and sufficient condition for differentiability of F at x is that $P[\bar{D}_\omega(x)] = 1$.

4. Conclusions

In this paper, we have examined the differentiability properties of functionals of the form

$$F(x) = E\{f(x, \omega)\},$$

where $x \in R^n$, ω is an element of a probability space, and f is a convex but not necessarily differentiable function of x for some ω . Such functionals arise often in connection with stochastic optimization problems, and their differentiability properties are vital to the solution of these problems. Necessary and sufficient conditions were given for the function F to be differentiable even when the function f is not. It was demonstrated that, for a large class of optimization problems, these conditions are satisfied; therefore, for such problems, classical solution methods can be used.

References

1. BAZARAA, M. S., GOODE, J. J., and SHETTY, C. M., *Optimality Criteria in Nonlinear Programming without Differentiability*, Operations Research, Vol. 19, No. 1, 1971.
2. BAZARAA, M. S., *Nonlinear Programming: Nondifferentiable Functions*, Georgia Institute of Technology, Ph.D. Thesis, 1971.

3. BERTSEKAS, D. P., and MITTER, S. K., *Steepest Descent for Optimization Problems with Nondifferentiable Cost Functionals*, Paper Presented at the 5th Annual Princeton Conference on Information Sciences and Systems, Princeton, New Jersey, 1971.
4. BERTSEKAS, D. P., and MITTER, S. K., *Descent Numerical Methods for Optimization Problems with Nondifferentiable Cost Functionals*, SIAM Journal on Control, Vol. 11, No. 4, 1973.
5. DEM'YANOV, V. F., and RUBINOV, A. M., *Minimization of Functionals in Normed Spaces*, SIAM Journal on Control, Vol. 6, No. 1, 1968.
6. HEINS, W., and MITTER, S. K., *Conjugate Convex Functions, Duality, and Optimal Control Problems, I. Systems Governed by Ordinary Differential Equations*, Information Sciences, Vol. 2, No. 2, 1970.
7. CHANEM, M. Z. E., *Optimal Control Problems with Nondifferentiable Cost Functionals*, Stanford University, Department of Engineering-Economic Systems, Ph.D. Thesis, 1970.
8. LUENBERGER, D. G., *Control Problems with Kinks*, IEEE Transactions on Automatic Control, Vol. AC-15, No. 5, 1970.
9. NEUSTADT, L. W., *A General Theory of Extremals*, Journal of Computational and System Science, Vol. 3, No. 1, 1969.
10. ROCKAFELLAR, R. T., *Conjugate Convex Functions in Optimal Control and the Calculus of Variations*, Journal of Mathematical Analysis and Applications, Vol. 32, No. 1, 1970.
11. ROCKAFELLAR, R. T., *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.
12. HALMOS, P. R., *Measure Theory*, Van Nostrand Reinhold Company, New York, New York, 1950.
13. DUNFORD, N., and SCHWARTZ, J. T., *Linear Operators, Part I*, John Wiley and Sons (Interscience Publishers), New York, New York, 1957.
14. ROYDEN, H. L., *Real Analysis*, The Macmillan Company, New York, New York, 1968.
15. ROCKAFELLAR, R. T., *Measurable Dependence of Convex Sets and Functions on Parameters*, Journal of Mathematical Analysis and Applications, Vol. 28, No. 1, 1969.
16. AUMANN, R. J., *Integrals of Set Valued Functions*, Journal of Mathematical Analysis and Applications, Vol. 12, No. 1, 1965.
17. BRINGLAND, T. F., *Trajectory Integrals of Set Valued Functions*, Pacific Journal of Mathematics, Vol. 33, No. 1, 1970.