ON THE EQUIVALENCE OF MULTISTAGE RECOURSE MODELS IN STOCHASTIC OPTIMIZATION

R.T. ROCKAFELLAR*

The following abstract model covers many stochastic optimization problems requiring a sequence of decisions. In each of N stages, an element x_k is chosen from a space X_k. Ultimately a cost $f(\xi_1,\dots,\xi_N, x_1,\dots,x_N)$ must be paid, where the elements ξ_k, belonging to spaces Ξ_k, represent exterior factors beyond the control of the decision maker. The cost may be $+\infty$, as a representation of the fact that certain combinations of $\xi = (\xi_1,\dots,\xi_N)$ and $x = (x_1,\dots,x_N)$ are impossible or forbidden. It is assumed that only $(\xi_1,\dots,\xi_k)$ is known with certainty at the time when x_k must be chosen. The only other information available about the exterior factors is that the occurrences of ξ are governed by a known probability distribution. The problem is to determine "decision rules" such that the overall expected cost is minimized.

To make more precise, let us suppose that each Ξ_k is a Hausdorff topological space, and σ is a regular Borel probability measure on $\Xi_1 \times\dots\times \Xi_N$. The support of σ will be denoted by Ξ. A function

$$x : \Xi \to X_1 \times\dots\times X_N$$

is <u>nonanticipative</u> if it is of the form

$$(1) \qquad x(\xi) = (x_1(\xi_1),\ x_2(\xi_1,\xi_2),\dots,\ x_N(\xi_1,\dots,\xi_N)).$$

The problem may be formulated as that of minimizing the functional

$$(2) \qquad F(x) = \int_\Xi f(\xi, x(\xi))\ \sigma(d\xi)$$

over some class of nonanticipative functions x.

Of course, the integral (2) may not be well defined, unless further conditions are imposed. But these can be of a very general nature. Assume that the spaces X_k are topological, and that f is Borel measurable. If $x(\xi)$ is measurable in ξ, then the mapping $\xi \to (\xi, x(\xi))$ is Borel measurable, and hence $f(\xi, x(\xi))$ is measurable in ξ. The following convention is then adopted : if either the positive or negative part of the function $\xi \to f(\xi, x(\xi))$ is summable, the integral F(x) is assigned its classical

* Supported in part by grant AF-AFOSR-72-2269.

value (possibly $+\infty$ or $-\infty$), while otherwise $F(x)$ is taken to be $+\infty$.

Observe that under this convention the inequality $F(x) < +\infty$ implies

(3) $$x(\xi) \in D(\xi) \text{ for almost every } \xi \in \Xi,$$

where the set

(4) $$D(\xi) = \{x \mid f(\xi,x) < +\infty\}$$

is the <u>implicit</u> <u>feasible</u> <u>region</u> of the decision space $X_1 \times \ldots \times X_N$. Thus in minimizing F over a class of measurable functions one is, in effect, minimizing subject to the constraint (3).

This note is concerned with two fundamental questions that arise in justifying and analyzing the model. First, to what extent is (3) essentially equivalent to the stronger constraint

(5) $$x(\xi) \in D(\xi) \text{ for every } \xi \in \Xi ,$$

which in some contexts appears more natural ? In other words, what conditions are needed to insure that a measurable, nonanticipative function x satisfying (3) can be converted to one satisfying (5) by alteration on a set of measure zero ? Secondly, when is it true that the infimum in the problem can be approached by functions x which are actually continuous ? The "nonanticipative" property renders these questions quite difficult for $N > 1$, and no one has previously provided any answers.

Our purpose is to describe some results in this direction in the case where

(6) $$X_1 \times \ldots \times X_N = R^{n_1} \times \ldots \times R^{n_N} = R^n$$

and certain convexity, semicontinuity and summability conditions are satisfied by the cost function

$$f : \Xi \times R^n \to R \cup \{+\infty\}.$$

Most of the proofs will appear elsewhere [1].

The context is delimited by the following basic assumptions.

(A1) The support Ξ of the probability measure σ is compact.

(A2) For each $\xi \in \Xi$, $f(\xi,x)$ is convex and lower-semicontinuous as a function of $x \in R^n$.

(A3) For each $\xi \in \Xi$, the set $D(\xi)$ has a nonempty interior.

(A4) The multifunction $\xi \to c\ell\, D(\xi)$ is continuous from Ξ to R^n (i.e. lower-semicontinuous with closed graph).

(A5) For each $x \in R^n$, $f(\xi,x)$ is measurable as a function of $\xi \in \Xi$.
(A6) Whenever $U \subset \Xi$ is open (relative to Ξ), $V \subset R^n$ is open, and f is finite on $U \times V$, one has

$$\int_U |f(\xi,x)| \sigma(d\xi) < +\infty \quad \text{for each } x \in V.$$

Example 1.

Let $X \subset R^n$ be a closed convex set, and for $i = 0,1,\ldots,m$ let f_i be a real valued (finite) function on $\Xi \times R^n$ such that $f_i(\xi,x)$ is convex in x. Let

$$f(\xi,x) = \begin{cases} f_o(\xi,x) & \text{if } x \in X \text{ and } f_i(\xi,x) \le 0,\ i = 1,\ldots,m, \\ +\infty & \text{otherwise.} \end{cases}$$

Then f satisfies (A2), and we have

$$D(\xi) = \{x \in X \mid f_i(\xi,x) \le 0\ ,\ i = 1,\ldots,m\}\ .$$

Suppose that X has a nonempty interior and $f_i(\xi,x)$ is continuous in ξ for $i = 1,\ldots,m$. If for each $\xi \in \Xi$ the set

$$\{x \in X \mid f_i(\xi,x) < 0\ ,\ i = 1,\ldots,m\}$$

is nonempty, it follows by routine convexity arguments that (A3) and (A4) hold. If furthermore $f_o(\xi,x)$ is a summable function of $\xi \in \Xi$ for each $x \in X$, then (A6) and (A7) are obviously satisfied as well.

LEMMA. Assumptions (A2), (A3) and (A5) imply in particular that f is Borel measurable on $\Xi \times R^n$, so that $F(x)$ is well-defined in the above sense whenever $x(\xi)$ is measurable in ξ.

Proof.

These assumptions imply that f is a normal convex integrand in the sense of [2,Lemma 2]. On the other hand, every such integrand is Borel measurable [3, Theorem 5].

A function $x : \Xi \to R^n$ will be called essentially nonanticipative if it can be made into a measurable nonanticipative function by altering its values on a set of measure zero. Let $\mathcal{N}_\infty$ denote the set of all such functions which are essentially bounded. If we like, we can identify $\mathcal{N}_\infty$ with a certain closed linear subspace of the Banach space $\mathcal{L}^\infty_n = \mathcal{L}^\infty(\Xi,\sigma\ ;\ R^n)$ consisting of all essentially bounded functions $x : \Xi \to R^n$. In fact, $\mathcal{N}_\infty$ is then closed not only with respect to the norm topology, but also the weak topology induced on $\mathcal{L}^\infty_n$ by the natural pairing with $\mathcal{L}^1_n = \mathcal{L}^1(\Xi,\sigma\,;\,R^n)$.

The following preliminary result is derived in [1] from the theory of convex integral functionals.

PROPOSITION.

Under assumptions (A1)-(A6), F is a convex functional from $\mathcal{L}^\infty_n$ to $R \cup \{+\infty\}$ which is lower-semicontinuous, not only with respect to the norm topology, but also the weak topology induced on $\mathcal{L}^\infty_n$ by $\mathcal{L}^1_n$. Furthermore, F is (norm) continuous on

$$\mathcal{W} = \{x\ \mathcal{L}^\infty_n \mid \exists\ \varepsilon > 0 \text{ with } x(\xi) + \varepsilon B \subset D(\xi) \text{ a.e.}\}$$

(where B is the unit ball of R^n), and this set $\mathcal{W}$ is the nonempty (norm) interior of $\{x \in \mathcal{L}^\infty_n \mid F(x) < +\infty\}$.

COROLLARY.

Suppose there is a compact set $X \subset R^n$ such that $D(\xi) \subset X$ for all $\xi \in \Xi$. Then the infimum of F(x) over all $x \in \mathcal{N}_\infty$ is attained.

The corollary, which provides an existence theorem for solutions to the stochastic optimization problem, is obtained from the fact that the set of functions $x \in \mathcal{L}^\infty_n$ satisfying $x(\xi) \in X$ almost everywhere is compact in the weak topology induced by $\mathcal{L}^1_n$. The hypothesis is satisfied, of course, in the case of Example 1 if the set X introduced there is bounded.

At all events, note that under the hypothesis of the corollary there is no loss of generality in the basic problem when the minimization is restricted to nonanticipative functions x which are essentially bounded, or in other words, when the problem is identified with that of minimizing F over $\mathcal{N}_\infty$. This formulation appears the best suited for obtaining strong results, at least in terms of convex analysis and duality.

The questions raised earlier concern the "almost everywhere" aspects of this formulation of the problem, as well as the relationship between minimizing over $\mathcal{N}_\infty$ and minimizing over $\mathcal{N}_\mathcal{C}$, the subspace of $\mathcal{N}_\infty$ consisting of the continuous nonanticipative functions. Our main result is the following (see [1]).

THEOREM.

Suppose in addition to (A1)-(A6) that the measure σ is "laminary", as defined below. Then every $x \in \mathcal{N}_\infty$ with $F(x) < +\infty$ can be converted, by alteration on a set of measure zero, to a measurable, (truly) bounded, (truly) nonanticipative function satisfying $x(\xi) \in cl\, D(\xi)$ for every $\xi \in \Xi$.

If furthermore $\mathcal{W} \cap \mathcal{N}_\infty \neq \emptyset$, then $\mathcal{W} \cap \mathcal{N}_e \neq \emptyset$ and

$$\inf \{F(x) \mid x \in \mathcal{N}_\infty\} = \inf \{F(x) \mid x \in \mathcal{N}_e\} .$$

To define what is meant by "laminary", we need some notation. For any set $S \subset \Xi$ and index k, $1 \leq k < N$, let

$$(7) \qquad \Lambda_k^S(\xi_1,\ldots,\xi_k) = \{(\xi_{k+1},\ldots,\xi_N) \mid (\xi_1,\ldots,\xi_k, \xi_{k+1},\ldots,\xi_N) \in S\} ,$$

$$(8) \qquad S^k = \{(\xi_1,\ldots,\xi_k) \mid \Lambda_k^S(\xi_1,\ldots,\xi_k) \neq \emptyset\} .$$

We say that the measure σ is laminary if it satisfies :

(i) The multifunction Λ_k^Ξ is lower-semicontinuous relative to Ξ^k, and

(ii) Whenever S is a Borel subset of Ξ with $\sigma(S) = \sigma(\Xi)$ such that S^k is a Borel set, then $\Lambda_k^S(\xi_1,\ldots,\xi_k)$ is dense in $\Lambda_k^\Xi(\xi_1,\ldots,\xi_k)$ for almost every $(\xi_1,\ldots,\xi_k)$ in S^k (with respect to the "projection" of σ on Ξ^k).

It is not hard to show, for instance, that σ is laminary if

$$(9) \qquad \sigma(d\xi) = \rho(\xi_1,\ldots,\xi_N)\, \pi_1(d\xi_1) \ldots \pi_N(d\xi_N) ,$$

where π_k is a (nonnegative regular Borel) measure on Ξ_k for $k = 1,\ldots,N$, and the density function ρ is positive on the support of the product measure $\pi_1 \times\ldots\times \pi_N$. (This follows from Fubini's theorem and the fact that in this case the multifunctions Λ_k^Ξ are constant-valued).

Even the first conclusion in the theorem can fail, without the presence of the two properties in the definition of "laminary". This can be demonstrated by counterexamples.

Example 2.

This is a two-stage example where $\Xi_1 = \Xi_2 = R$ and $R^{n_1} = R^{n_2} = R$. Let the interval $[0,1]$ be expressed as the union of two disjoint subsets A and A' of positive measure, such that A is dense and A' is closed, and let

$$T = (A \times [0,2]) \cup (A' \times [0,1]) .$$

Define the Borel measure σ on R^2 by

$$\sigma(S) = \text{mes}\,(S \cap T)\, /\, \text{mes}\, T ,$$

where "mes" denotes Lebesgue measure. Then σ is a probability measure whose support is

$$\Xi = c\ell\ T = [0,1] \times [0,2] .$$

But σ does not satisfy property (ii) in the definition of "laminary" (take $S = T$), even though σ is absolutely continuous with respect to a product measure. Define f on $\Xi \times R$ by

$$f(\xi_1,\xi_2,x_1,x_2) = \begin{cases} x_1 & \text{if } 0 \leq x_2 \leq x_1 - \xi_2, \\ +\infty & \text{otherwise}, \end{cases}$$

so that

$$D(\xi_1,\xi_2) = c\ell\ D(\xi_1,\xi_2) = \{(x_1,x_2) \in R^2 \mid 0 \leq x_2 \leq x_1 - \xi_2\}.$$

It may be verified that assumptions (A1)-(A6) are satisfied and

$$\min \{F(x) \mid x \in \mathcal{N}_\infty\} = F(\bar{x}) < 2 ,$$

where

$$\bar{x}(\xi) = (\bar{x}(\xi_1), \bar{x}_2(\xi_1,\xi_2)) = \begin{cases} (2,0) & \text{if } \xi_1 \in A , \\ (1,0) & \text{if } \xi_1 \in A' . \end{cases}$$

But if the stronger condition (5) is imposed, the minimum is instead $F(\bar{\bar{x}}) = 2$, where

$$\bar{\bar{x}}(\xi) = (\bar{\bar{x}}_1(\xi_1) , \bar{\bar{x}}_2(\xi_1,\xi_2)) \equiv (2,0).$$

Thus the constraint conditions (3) and (5) are <u>not</u> "equivalent" in this case, and in particular the first assertion of the theorem is false for $x = \bar{x}$.

<u>Example 3</u>.

Again we consider a two-stage case with $\Xi_1 = \Xi_2 = R$ and $R^{n_1} = R^{n_2} = R$, but this time it is only property (i) of the definition of "laminary" which is lacking, and still the first assertion of theorem is false. The probability measure is

$$\sigma(S) = \frac{1}{2} \text{ mes } (S \cap T) ,$$

where

$$T = ([0,1] \times [0,1]) \cup ([-1,0] \times [-1,0]) .$$

We have $\Xi = T$, so Λ_1^{Ξ} is not lower-semicontinuous at $\xi_1 = 0$. Let

$$f(\xi_1,\xi_2,x_1,x_2) = \begin{cases} 0 \text{ if } \xi_2 \geq 0 \text{ and } -2+3\xi_2 \leq x_1 \leq 2 , \\ 0 \text{ if } \xi_2 \leq 0 \text{ and } -2 \leq x_1 \leq 2+3\xi_2 , \\ +\infty \text{ otherwise.} \end{cases}$$

Assumptions (A1)-(A6) are satisfied, and

$$\min \{F(x) \mid x \in \mathcal{N}_\infty\} = F(\bar{x}) = 0$$

for the function

$$\bar{x}(\xi) = (\bar{x}_1(\xi_1), \bar{x}_2(\xi_1,\xi_2)) = \begin{cases} (3/2,0) \text{ if } \xi_1 > 0 , \\ (0,0) \text{ if } \xi_1 = 0 , \\ (-3/2,0) \text{ if } \xi_1 < 0. \end{cases}$$

In fact $\bar{x}(\xi) + \frac{1}{2} B \subset D(\xi)$ almost everywhere. But it is easy to see there does not exist any nonanticipative function x whatsoever which satisfies $x(\xi) \in D(\xi)$ for all $\xi \in \Xi$.

The proof of the first assertion of the theorem makes use of convexity mainly just as a matter of convenience in terms of the formulation. However, for the rest of the theorem, concerning approximation of the infimum via continuous recourse functions x, convexity seems to be essential. The basic tool is a theorem of E. MICHAEL [4] on the existence of continuous selections, and convexity is already an important hypothesis in this result, as is well understood.

Of course, the proof is not effected by means of a single continuous selection, but by a certain sequence of N selections, each from a multifunction dependant on the preceding selections. This is the great complication caused by the requirement of nonanticipativity. The multifunctions must be constructed in such a way that Michael's theorem is applicable at each step, and here convexity seems to play a crucial role over and over again.

The argument establishing the first part of the theorem is similarly complicated, but a sequence of measurable, rather than continuous, selections is involved.

These results are motivated especially by applications to Example 1. A theory of Lagrange multipliers and duality for this case, based heavily on them, is outlined in [5]. The multipliers are certain measures, not necessarily absolutely continuous with respect to the underlying probability measure σ. This theory provides an alternative to the approach of R. WETS and the author in [6], where the multipliers in general can take the form of elements of the dual of an $\mathcal{L}^\infty$ space.

REFERENCES

[1] E. MICHAEL,
"Continuous selections, I",
Ann. of Math. 63 (1956), 361-382.

[2] R.T. ROCKAFELLAR,
"Integrals which are convex functionals",
Pacific J. Math. 24 (1968), 525-539.

[3] R.T. ROCKAFELLAR,
"Measurable dependance of convex sets and functions on parameters",
J. Math. Analysis Appl. 28 (1969), 4-25.

[4] R.T. ROCKAFELLAR,
"Lagrange multipliers for an N-stage model in stochastic programming",
Colloque d'Analyse Convexe (St-Pierre-de-Chartreuse, 1974),
J.P. AUBIN (editor), Springer-Verlag, to appear.

[5] R.T. ROCKAFELLAR and R.J.B. WETS,
"Continuous versus measurable recourse in N-stage stochastic programming",
J. Math. Analysis Appl., to appear.

[6] R.T. ROCKAFELLAR and R.J.B. WETS,
"Stochastic convex programming : basic duality" and "Stochastic convex programming : extended duality and singular multipliers",
Pacific J. Math., to appear.