LAGRANGE MULTIPLIERS FOR AN N-STAGE MODEL IN STOCHASTIC CONVEX PROGRAMMING

R.T. Rockafellar *

Many decision processes are of a sequential nature and make use of information which is revealed progressively through the observation, at various times, of random variables with known distributions. A simple model with discrete stages is the following.

First we make an observation, which singles out for us an element ξ_1 of R^{ν_1}. Based on the information thereby gained, we choose a response x_1, which is an element of R^{n_1}. Then we make a new observation, yielding $\xi_2 \in R^{\nu_2}$, and choose a response $x_2 \in R^{n_2}$. This continues until, at the N^{th} stage, we determine $\xi_N \in R^{\nu_N}$ and choose $x_N \in R^{n_N}$. The choices are subject to certain constraints of the form

(1) $\qquad x \in X$ and $f_i(\xi,x) \leq 0$ for $i=1,\ldots,m$,

where

(2) $$\xi = (\xi_1,\ldots,\xi_N) \in R^{\nu} := R^{\nu_1} \times \ldots \times R^{\nu_N},$$

$$x = (x_1,\ldots,x_N) \in R^{n} := R^{n_1} \times \ldots \times R^{n_N}.$$

Of course, some of the functions f_i might depend only on certain initial components of ξ and x. The result of the decision process is a cost $f_0(\xi,x)$. The distribution of ξ is assumed known : one is given a regular Borel probability measure σ on R^{ν} with support Ξ. The problem is to determine decision rules which minimize the expected cost, taking into account the fact that the decision x_k can depend on the past observations $\xi_1,\ldots,\xi_k$, but not on the future observations $\xi_{k+1},\ldots,\xi_N$.

This is an N-stage stochastic programming problem. Our interest here is in the convex case, where X is a nonempty closed convex set in R^n, and $f_i(\xi,.)$ is a finite convex function on R^n for every $\xi \in \Xi$ and $i=0,1,\ldots,m$. The feasible set

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

is then closed and convex for every $\xi \in \Xi$.

---

* Supported in part by grant AF-AFOSR-72-2269.
The results reported here were obtained in collaboration with R.J.B. Wets.

A "decision rule" is represented by a function $x : \Xi \to R^n$. Such a function is said to be nonanticipative if it is of the form

(4) $$x(\xi) = (x_1(\xi_1), x_2(\xi_1,\xi_2), \ldots, x_N(\xi_1,\ldots,\xi_N)) .$$

The objective in the problem is to minimize the expectation

(5) $$\Phi(x) = E\, f_o(\xi,x(\xi)) = \int_{\Xi} f_o(\xi,x(\xi))\sigma(d\xi)$$

over some class of nonanticipative functions x, not yet specified precisely. For simplicity here, we take the set X to be bounded, so that only bounded functions x need to be considered. (This is no serious loss of generality in practice). We assume that $f_o(\xi,x)$ is summable (Borel measurable) with respect to $\xi \in \Xi$, and that $f_i(\xi,x)$ is continuous with respect to $\xi \in \Xi$ for i=1,...,m. Then for every bounded, Borel measurable function $x : \Xi \to R^n$ one has $f_i(\xi,x(\xi))$ Borel measurable in $\xi \in \Xi$, summable for i=0 and bounded for i=1,...,m. (This is known from the theory of normal convex integrands). Note also that the graph of the multifunction $D: \xi \to D(\xi)$ is closed.

Let $\mathcal{N}$ be the linear space consisting of all the bounded, nonanticipative functions $x : \Xi \to R^n$. We adopt as our basic model the problem :

(P) minimize $\Phi(x)$ over all $x \in \mathcal{N}$ satisfying $x(\xi) \in D(\xi)$ for every $\xi \in \Xi$.

The aim of this paper is to outline a theory of Lagrange multipliers corresponding to the constraints defining $D(\xi)$. The proofs, based mainly on results in [1] and the general theory of dual optimization problems [3], will be presented elsewhere in a broader context. For a somewhat different approach in the two-stage case, see [2].

Let $\mathcal{X}$ be the set of all $x \in \mathcal{N}$ satisfying $x(\xi) \in X$ for every $\xi \in \Xi$, and let $\mathcal{Y}$ be the set of all m-tuples $y = (y_1,\ldots,y_m)$, where y_i is a nonnegative, regular Borel measure on Ξ. For $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, the function

(6) $$L(x,y) = \int_{\Xi} f_o(\xi,x(\xi))\sigma(d\xi) + \Sigma_{i=1}^m \int_{\Xi} f_i(\xi,x(\xi))y_i(d\xi)$$

is well-defined and finite, convex in x and affine in y.
Moreover, setting

(7) $$f(x) = \sup_{y \in \mathcal{Y}} L(x,y) \quad \text{for} \quad x \in \mathcal{X}$$

we have

(8) $$f(x) = \begin{cases} E_\xi f_o(\xi,x(\xi)) \text{ if } f_i(\xi,x(\xi)) \le 0 \text{ for every } \xi \in \Xi,\ i=1,\ldots,m , \\ +\infty \text{ otherwise} . \end{cases}$$

Therefore (P) is equivalent to minimizing f(x) over all $x \in \mathcal{X}$. Accordingly, we define

(9) $g(y) = \inf_{x\in\mathcal{X}} L(x,y)$ for $y \in \mathcal{Y}$

and take the problem dual to (P) to be

(D) maximize g(y) over all $y \in \mathcal{Y}$.

Then by definition

(10) $$\inf(P) = \inf_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} L(x,y) \geq \sup_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} L(x,y) = \sup (D) .$$

It is elementary in this framework that a pair $(\bar{x},\bar{y})$ is a saddle-point of L on $\mathcal{X} \times \mathcal{Y}$ if and only if min (P) = max (D) , $\bar{x}$ solves (P), and $\bar{y}$ solves (D). An important task therefore is to establish, under reasonable conditions, a duality theorem of the type min (P) = max (D), since this not only furnishes the existence of solutions to (P) and (D), but also their characterization.

To derive such a theorem from the general perturbational theory of duality, the Lagrangian L must first be shown to correspond to some system of perturbations of (P). While this is possible, certain technicalities arise which make it difficult to apply the theory in a direct way, and one needs to use further arguments based on other formulations of the primal problem (P).

Let $\mathcal{B}_m$ be the space of all bounded measurable functions $u : \Xi \to R^m$, and similarly $\mathcal{B}_n$, so that $\mathcal{X}$ and $\mathcal{N}$ are subsets of $\mathcal{B}_n$. For $x \in \mathcal{B}_n$ and $u \in \mathcal{B}_m$, define :

(11) $$F(x,u) = \begin{cases} E_\xi f_o(\xi,x(\xi)) \text{ if } x \in \mathcal{N} ,\ x(\xi) \in X \text{ and } f_i(\xi,x(\xi)) \leq u_i(\xi) \\ \qquad\qquad \text{for every } \xi \in \Xi ,\ i=1,\dots,m , \\ +\infty \quad \text{otherwise.} \end{cases}$$

Then (P) can be identified with the problem of minimizing F(x,0) over all $x \in \mathcal{B}_n$; the perturbed problem corresponding to $u \in \mathcal{B}_m$ consists instead of minimizing F(x,u) over all $x \in \mathcal{B}_n$. We can pair $\mathcal{B}_m$ with the space $\mathcal{M}_m$ consisting of all R^m-valued regular Borel measures y on Ξ :

(12) $$\langle u,y \rangle = \Sigma_{i=1}^m \int_\Xi u_i(\xi)y_i(d\xi) .$$

The Lagrangian function K corresponding to F is then

(13) $$K(x,y) = \inf_{u\in\mathcal{B}_m} \{F(x,u)+\langle u,y \rangle\} = \begin{cases} L(x,y) & \text{if } x \in \mathcal{X} ,\ y \in \mathcal{Y} \\ -\infty & \text{if } x \in \mathcal{X} ,\ y \notin \mathcal{Y} \\ +\infty & \text{if } x \notin \mathcal{X} . \end{cases}$$

Saddle points of K on $\mathcal{B}_n \times \mathcal{M}_m$ are, of course, the same as saddle-points of L on $\mathcal{X} \times \mathcal{Y}$.

The difficulty in using this scheme straightforwardly is that it is not clear how to get the desired continuity and compactness properties out of topologies compatible with the pairing (12) and the analogous pairing between the x-space $\mathcal{B}_n$ and $\mathcal{M}_n$. We circumvent this by working simultaneously with two other versions of the problem :

($P_{\mathcal{L}}$) minimize $\Phi(x)$ over all $x \in \mathcal{N}$ satisfying $x(\xi) \in D(\xi)$ for almost every $\xi \in \Xi$,

($P_{\mathcal{C}}$) minimize $\Phi(x)$ over all continuous $x \in \mathcal{N}$ satisfying $x(\xi) \in D(\xi)$ for every $\xi \in \Xi$ (assuming Ξ compact).

It is clear that in general

(14) $$\inf (P_{\mathcal{L}}) \leq \inf (P) \leq \inf (P_{\mathcal{C}}) .$$

Both ($P_{\mathcal{L}}$) and ($P_{\mathcal{C}}$) are more open to attack by ordinary methods than is (P). In the case of ($P_{\mathcal{L}}$), we use the same system of perturbations as above, except that "every ξ " becomes "almost every ξ" in the definition (11) of F, and the spaces $\mathcal{B}_n$ and $\mathcal{B}_m$ are replaced by the Lebesgue spaces $\mathcal{L}^\infty_n(\Xi,\sigma)$ and $\mathcal{L}^\infty_m(\Xi,\sigma)$. One thus has a certain function $\bar{F}$ on $\mathcal{L}^\infty_n \times \mathcal{L}^\infty_m$; the problem ($P_{\mathcal{L}}$) is equivalent to minimizing $\bar{F}(x,0)$ over all $x \in \mathcal{L}^\infty_n$. Instead of pairing $\mathcal{L}^\infty_m(\Xi,\sigma)$ with $\mathcal{L}^1_m(\Xi,\sigma)$, it is better notationally in the present context to pair it via (12) with the subspace $\mathcal{M}^{\mathcal{L}}_m$ of $\mathcal{M}_m$ consisting of the measures which are absolutely continuous with respect to σ. (Obviously, $\mathcal{M}^{\mathcal{L}}_m$ is canonically isomorphic to $\mathcal{L}^1_m$). Similarly with $\mathcal{L}^\infty_n$ and $\mathcal{M}^{\mathcal{L}}_n$. The corresponding Lagrangian is then

(15) $$\bar{K}(x,y) = \inf_{u \in \mathcal{L}^\infty_m} \{\bar{F}(x,u)+\langle u,y\rangle\} = \begin{cases} L(x,y) \text{ if } x \in \mathcal{X}_{\mathcal{L}} & , y \in \mathcal{Y}_{\mathcal{L}} , \\ -\infty \quad \text{if } x \in \mathcal{X}_{\mathcal{L}} & , y \notin \mathcal{Y}_{\mathcal{L}} , \\ +\infty \quad \text{if } x \notin \mathcal{X}_{\mathcal{L}} & , \end{cases}$$

where $\mathcal{X}_{\mathcal{L}}$ consists of the (equivalence classes generated by the) functions in $\mathcal{N}$ satisfying $x(\xi) \in D(\xi)$ for almost every $\xi \in \Xi$ (with respect to σ), and $\mathcal{Y}_{\mathcal{L}}$ consists of the measures in $\mathcal{Y}$ which are absolutely continuous with respect to σ. Observe that for $(x,y) \in \mathcal{X}_{\mathcal{L}} \times \mathcal{Y}_{\mathcal{L}}$ we can also express the Lagrangian by

(16) $$L(x,y) = E_\xi[f_0(\xi,x(\xi)) + \Sigma^m_{i=1} \frac{dy_i}{d\sigma} (\xi) f_i(\xi,x(\xi))],$$

where $dy_i/d\sigma$ is the Radon-Nikodym derivative of the ith component of y with respect to σ .

The corresponding dual of $(P_{\mathcal{L}})$ is

$$(D_{\mathcal{L}}) \qquad \text{maximize } g_{\mathcal{L}}(y) = \inf_{x \in \mathcal{X}_{\mathcal{L}}} L(x,y) \text{ over all } y \in \mathcal{Y}_{\mathcal{L}} \ .$$

A duality theorem relating $(P_{\mathcal{L}})$ and $(D_{\mathcal{L}})$ can be derived, making use of the compactness we have assumed for the set X in (3). We can identify $\mathcal{N}$ with a subspace of $\mathcal{L}^{\infty}_{n}$ which is closed in the weak topology induced by the pairing with $\mathcal{M}^{\mathcal{L}}_{n}$; the compactness of X then implies that $\mathcal{X}_{\mathcal{L}}$ is compact in the same weak topology. On the other hand, it can be shown from the theory of convex integral functionals that L(x,y) is lower semicontinuous in x with respect to this topology for each $y \in \mathcal{Y}_{\mathcal{L}}$. Applying a standard minimax theorem (or a corresponding result in duality theory), we are able to prove :

THEOREM 1. $\min (P_{\mathcal{L}}) = \sup (D_{\mathcal{L}}) > -\infty$.

(Here min $(P_{\mathcal{L}})$ is interpreted as $+\infty$ if $(P_{\mathcal{L}})$ has no feasible solutions ; otherwise, the use of "min" indicates that an optimal solution exists).

On the other hand, in the case of $(P_{\mathcal{C}})$ we can obtain results from the perturbation scheme (11) with the spaces $\mathcal{B}_n$ and $\mathcal{B}_m$ replaced by the corresponding spaces $\mathcal{C}_n$ and $\mathcal{C}_m$ of <u>continuous</u> functions on Ξ .
(The boundedness of Ξ is needed for this to make sense). With $\mathcal{C}_m$ paired with $\mathcal{M}_m$ via (12), we obtain the same Lagrangian K as in (13), but with $\mathcal{X}$ replaced by $\mathcal{X}_{\mathcal{C}}$, consisting of the functions in $\mathcal{X}$ which are continuous. The dual problem is then

$$(D_{\mathcal{C}}) \qquad \text{maximize } g_{\mathcal{C}}(y) = \inf_{x \in \mathcal{X}_{\mathcal{C}}} L(x,y) \text{ over all } y \in \mathcal{Y} \ .$$

This time a duality theorem can be derived in terms of the norm topology on the perturbation space $\mathcal{C}_m$, since this is compatible with the pairing with the multiplier space $\mathcal{M}_m$. (The norm topology on $\mathcal{L}^{\infty}_{m}$ was not, of course, compatible with the pairing with $\mathcal{M}^{\mathcal{L}}_{m}$). Let us call $(P_{\mathcal{C}})$ <u>strictly feasible</u> if there exists a continuous function $x \in \mathcal{N}$ satisfying $x(\xi) \in X$ and $f_i(\xi, x(\xi)) < 0$, $i=1,\dots,m$, for every $\xi \in \Xi$. For such an x, the function $u \to F(x,u)$ is bounded above in a neighborhood of the origin in $\mathcal{C}_m$. Fundamental duality theory therefore gives us the following.

THEOREM 2. <u>Assuming</u> $(P_{\mathcal{C}})$ <u>is strictly feasible</u> (<u>with</u> Ξ <u>bounded</u>), <u>one has</u> $\inf (P_{\mathcal{C}}) = \max (D_{\mathcal{C}}) < +\infty$.

It is interesting to contrast Theorems 1 and 2.
In Theorem 1, we have the existence of primal solutions (assuming the constraints are feasible), but the existence of dual solutions is not assured. It might be hoped

that dual solutions would exist under the assumption that $(P_{\mathcal{L}})$ is strictly feasible i.e. that there is a function $x \in \mathcal{N}$ satisfying for some $\varepsilon > 0$ the conditions $x(\xi) \in X$ and $f_i(\xi, x(\xi)) < -\varepsilon$, $i=1,\ldots,m$, for almost every $\xi \in \Xi$. But examples are known, even in the two-stage case, which show this is generally false [2]. In Theorem 2 we have the existence of a dual solution, and hence Lagrange multipliers characterizing the solutions to the primal problem. However, the continuity requirement makes it very unlikely that the infimum in the primal is actually attained, so this characterization is rather vacuous. Of course, neither Theorem 1 nor Theorem 2 is applicable directly to the problems we really want to analyze, namely (P) and (D).

The route we now follow in obtaining results about (P) and (D) is to impose conditions under which, to a certain extent, the various primals and duals are "equivalent", so that Theorems 1 and 2 can be joined into a single statement. The extra conditions concern the probability measure σ. Without such conditions, the inequalities in (14) can be strict, as seen from examples in [1].

For each $S \subset R^{\nu} = R^{\nu_1} \times \ldots \times R^{\nu_N}$ and index k, $1 \leq k < N$, let S^k denote the projection of S on $R^{\nu_1} \times \ldots \times R^{\nu_k}$, and let

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

As in [1], we shall say that σ is laminary if the support Ξ of σ is compact and the following two conditions are satisfied :

a) If S is any Borel subset of Ξ with $\sigma(\Xi \setminus S) = 0$, and if S^k is a Borel set, then for almost every $(\xi_1,\ldots,\xi_k) \in S^k$ (with respect to the "projection" of σ on S^k) one has

$$\mathrm{cl}\Lambda_k^S(\xi_1,\ldots,\xi_k) = \Lambda_k^{\Xi}(\xi_1,\ldots,\xi_k) .$$

b) The multifunction Λ_k^{Ξ} is continuous relative to Ξ^k .

It is not hard to see that this condition is satisfied in particular if Ξ is a product of compact sets $\Xi_k \subset R^{\nu_k}$, and

$$\sigma(d\xi) = \rho(\xi_1,\ldots,\xi_N)\sigma_1(d\xi_1)\ldots\sigma_N(d\xi_N) ,$$

where σ_k is a regular Borel measure on Ξ_k . It is also satisfied trivially if Ξ is a finite set.

THEOREM 3. Suppose the probability measure σ is laminary (with bounded support Ξ), and let $x \in \mathcal{N}$ be such that $x(\xi) \in D(\xi)$ for almost every ξ (with respect to σ). Then there exists $\bar{x} \in \mathcal{N}$, agreeing with x almost everywhere, such that $\bar{x}(\xi) \in D(\xi)$ for every $\xi \in \Xi$.

If in addition int $X \neq \emptyset$ and $(P_{\mathcal{L}})$ is strictly feasible, then for arbitrary $\varepsilon > 0$ there exists a continuous function $\bar{\bar{x}} \in$ such that

$$\sigma\{\xi \in \Xi \mid \varepsilon < |\bar{x}(\xi) - \bar{\bar{x}}(\xi)|\} < \varepsilon .$$

This result can be derived from [1, Prop.7 and the proof of Theorem 2]. From it, we obtain, with a few manipulations our main result :

THEOREM 4. Suppose that σ is laminary (with compact support), $(P_{\mathcal{L}})$ is strictly feasible, and int $X \neq \emptyset$. Then

$$\begin{aligned} +\infty > \min (P) = \min (P_{\mathcal{L}}) &= \inf (P_{\mathcal{C}}) \\ &= \max (D) = \max (D_{\mathcal{C}}) = \sup (D_{\mathcal{L}}) > -\infty . \end{aligned}$$

COROLLARY. Under these assumptions (and the basic assumptions made earlier), (P) and (D) have optimal solutions, and the pairs of such solutions are characterized as the saddle points of the Lagrangian L on $\mathcal{X} \times \mathcal{Y}$.

REFERENCES

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

[2] R.T. Rockafellar and R.J.B. Wets, "Stochastic convex programming : basic duality" and "Stochastic convex programming : singular functionals and extended duality", both to appear in Pacific J. Math.

[3] R.T. Rockafellar, "Conjugate Duality and Optimization", SIAM/CBMS monograph series n°16, SIAM Publications, 1974.