Reference copy

SIAM J. CONTROL AND OPTIMIZATION
Vol. 16, No. 1, January 1978

THE OPTIMAL RECOURSE PROBLEM IN DISCRETE TIME: L^1-MULTIPLIERS FOR INEQUALITY CONSTRAINTS*

R. T. ROCKAFELLAR† AND R. J-B. WETS‡

Abstract. An optimal recourse problem is an optimization problem with both stochastic and dynamic aspects, involving the interplay of observations and responses. In discrete time (with a finite horizon), there are finitely many stages, at each of which a decision is selected on the basis of prior observations of random events and subject to costs and constraints affected by these observations as well as past decisions. The goal is to minimize expected cost, taking into account the known distribution of future random events. This paper is concerned with the derivation of necessary and sufficient conditions for optimality in the case of convex costs and constraints.

It is shown that if the recourse problem is strictly feasible and satisfies a new condition called *essentially complete recourse*, optimal solutions can be characterized by a "pointwise" Kuhn–Tucker property involving L^1-multipliers. Applications to multistage stochastic programs with special structures are developed in the last two sections of the paper. In particular, the relation between the general model and discrete-time stochastic control models is brought out by applying the basic results to a linear stochastic problem with state constraints.

1. Introduction. For $k = 1, \cdots, N$, let $\xi_k \in R^{\nu_k}$ and $u_k \in R^{n_k}$ represent the observation and decision (control) associated with stage k of a sequential decision process. The sequence of observations

$$\xi = (\xi_1, \xi_2, \cdots, \xi_N) \in R^{\nu_1} \times R^{\nu_2} \times \cdots \times R^{\nu_N} = R^{\nu}$$

and the sequence of decisions

$$u = (u_1, u_2, \cdots, u_N) \in R^{n_1} \times R^{n_2} \times \cdots \times R^{n_N} = R^n$$

determine a "cost" denoted $f_0(\xi, u)$. The objective is to find a *recourse function* (or *policy*, or *decision rule*, or *control law*) $\xi \mapsto u(\xi)$ which minimizes the expected value of this cost subject to certain constraints, including a kind of nonanticipativity, i.e. the property that $u_k(\xi)$ essentially depends only on $\xi_1, \cdots, \xi_k$. This is an *optimal recourse problem in discrete time*. Our aim here is to derive necessary and sufficient conditions for the optimality of a recourse function in the case of a problem satisfying convexity assumptions with respect to the decision variables.

To give a precise formulation, let $(\Xi, \mathscr{F}, \sigma)$ denote the sample space associated with the random elements of the problem; Ξ is a Borel subset of R^{ν}, $\mathscr{F}$ is the Borel field on Ξ, and σ is a Borel probability measure on $(\Xi, \mathscr{F})$. The corresponding expectation operator is denoted simply by E.

A function $u: \Xi \to R^n$ is said to be *nonanticipative* in the sequential framework described above if it is of the form

$$u(\xi) = (u_1(\xi_1), u_2(\xi_1, \xi_2), \cdots, u_N(\xi_1, \cdots, \xi_N));$$

it is *essentially nonanticipative* if it is measurable (with respect to $\mathscr{F}$) and differs

* Received by the editors September 9, 1976.

† Department of Mathematics, University of Washington, Seattle, Washington 98105. This research was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under AFOSR Grant 72-2269.

‡ Department of Mathematics, University of Kentucky, Lexington, Kentucky 40506, and Université Scientifique et Médicale de Grenoble, Grenoble, France. This research was sponsored by the National Science Foundation under Grant MPS 75-07028.

only on a set of measure zero (with respect to σ) from some measurable nonanticipative function.

It is useful, for purposes of comparison with other work in stochastic optimization, to recognize that this concept of essential nonanticipativity can also be formulated in terms of a nest of sigma-fields. Let $\mathcal{F}'$ denote the class of all sets in $\mathcal{F}$ of measure zero with respect to σ, and for $k = 1, \cdots, N$ let $\mathcal{F}_k$ be the sigma-field generated by $\xi_1, \cdots, \xi_k$ completed with respect to σ, i.e. the class of all sets of $\mathcal{F}$ of the form

$$((A \times [R^{\nu_{k+1}} \times \cdots \times R^{n}]) \cap \Xi) \Delta B,$$

where A is a Borel set in $R^{\nu_1} \times \cdots \times R^{\nu_k}$, B is a set in $\mathcal{F}'$, and Δ denotes symmetric difference. Then each $\mathcal{F}_k$ is a sigma-field.

$$\mathcal{F}' \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_N = \mathcal{F},$$

and *a function* $u: \Xi \mapsto R^n$ *is essentially nonanticipative if and only if for* $k = 1, \cdots, N$ *the function* $u_k: \Xi \mapsto R^{n_k}$ *is* $\mathcal{F}_k$*-measurable.*

In fact, everything that follows remains valid for an arbitrary choice of sigma-fields $\mathcal{F}_1, \cdots, \mathcal{F}_N$ nesting as indicated, if the latter property is adopted as the generalized definition of essential nonanticipativity. We therefore work mainly in this notational framework.

For the conditional expectation given $\mathcal{F}_k$, we write E^k. This is taken to be a *regular* conditional expection, i.e. representable as an indefinite integral with respect to a regular conditional probability. (Such regular conditional probabilities exist, even for a general choice of $\mathcal{F}_k$, because $\mathcal{F}$ is the Borel field on Ξ and σ is a regular Borel probability measure.)

The optimal recourse problem considered here consists of minimizing the expected cost

$$I_{f_0}(u) = E\{f_0(\xi, u(\xi))\} \tag{1.1}$$

over all essentially nonanticipative functions $u: \Xi \to R^n$ satisfying *almost surely* (a.s.)

$$f_i(\xi, u(\xi)) \leqq 0, \qquad i = 1, \cdots, m, \tag{1.2}$$

and the abstract constraint $u(\xi) \in U(\xi)$. It is assumed that for every $\xi \in \Xi$ the set $U(\xi)$ is closed and convex with nonempty interior, and the functions $u \mapsto f_i(\xi, u)$, $i = 0, 1, \cdots, m$, are defined for all $u \in U(\xi)$ (finite, i.e. real-valued), convex and lower semicontinuous. It is assumed further that for each $u \in R^n$ the set

$$U^{-1}(u) = \{\xi \in \Xi | u \in U(\xi)\}$$

is Borel measurable (i.e. belongs to $\mathcal{F}$) and the functions $\xi \mapsto f_i(\xi, u)$ are all Borel measurable relative to $U^{-1}(u)$. Setting

$$f_i(\xi, u) = +\infty \quad \text{if } u \notin U(\xi),$$

we obtain from these assumptions that each f_i is a normal convex integrand on $\Xi \times R^n$ [1, Lemma 2] and the multifunction $U: \Xi \to R^n$ is measurable [2, Cor. 3.1].

It follows that $f_i(\xi, u(\xi))$ is Borel measurable in $\xi \in \Xi$ when $u(\xi)$ is measurable

[1, Cor. to lemma 5]. Moreover, the multifunction

$$D: \xi \to D(\xi) = \{u \in U(\xi) | f_i(\xi, u) \leqq 0, i = 1, \cdots, m\} \tag{1.3}$$

is measurable [2, Cors. 4.1 and 4.3]. This multifunction with closed, convex values provides an abstract description of the constraint structure, and it is crucial in what follows.

We assume that the sets $D(\xi)$ are uniformly bounded (i.e. their union for all $\xi \in \Xi$ is a bounded subset of R^n). This enables us to restrict our attention in the recourse problem to functions u belonging to the space $L_n^\infty = L^\infty(\Xi, \mathcal{F}, \sigma; R^n)$. We suppose in addition that to each bounded set $K \subset R^n$ there corresponds a summable function $\alpha: \Xi \to R$ and a constant $\beta \in R$ such that

$$|f_0(\xi, u)| \leqq \alpha(\xi) \quad \text{for all } u \in U(\xi) \cap K, \tag{1.4}$$

$$|f_i(\xi, u)| \leqq \beta \quad \text{for all } u \in U(\xi) \cap K, \quad i = 1, \cdots, m. \tag{1.5}$$

These "growth" conditions imply that for every function u in the class

$$\mathcal{U} = \{u \in L_n^\infty | u(\xi) \in U(\xi) \text{ a.s.}\}$$

the functions $f_i(\cdot, u(\cdot))$, $i = 1, \cdots, m$, are essentially bounded, while $f_0(\cdot, u(\cdot))$ is summable.

With these assumptions the optimal recourse problem introduced above is well-defined and can be stated as:

P Minimize the functional (1.1) over all $u \in \mathcal{U} \cap \mathcal{N}_\infty$ satisfying (1.2) a.s.,

where $\mathcal{N}_\infty$ represents the constraint of nonanticipativity:

$$\begin{aligned} \mathcal{N}_\infty &= \{u = (u_1, \cdots, u_N) \in L_n^\infty | u_k \text{ is } \mathcal{F}_k\text{-measurable}, k = 1, \cdots, N\} \\ &= L_{n_1}^\infty(\Xi, \mathcal{F}_1, \sigma) \times L_{n_2}^\infty(\Xi, \mathcal{F}_2, \sigma) \times \cdots \times L_{n_N}^\infty(\Xi, \mathcal{F}_N, \sigma). \end{aligned}$$

Clearly $\mathcal{N}_\infty$ is a linear subspace of $\mathcal{L}_n^\infty$, while $\mathcal{U}$ is a convex set, as is the class of all $u \in \mathcal{U}$ satisfying (1.2) a.s. The functional (1.1) is convex and finite on $\mathcal{U}$. Thus we are dealing with a convex optimization problem. In such a setting, it is typical to find multiplier characterizations of optimality which are always sufficient but not necessary without some "constraint qualification."

A natural constraint qualification to consider is that P be *strictly feasible*. This is taken to mean that there exist $\tilde{u} \in \mathcal{N}_\infty$ and $\varepsilon > 0$ such that

$$f_i(\xi, \tilde{u}(\xi)) \leqq -\varepsilon \quad \text{a.s.} \quad \text{for } i = 1, \cdots, m, \tag{1.6}$$

and

$$\tilde{u}(\xi) + \varepsilon B \subset D(\xi) \quad \text{a.s.}, \tag{1.7}$$

where B is the closed unit ball in R^n. However, strict feasibility is not enough in itself. What we need for our characterization of optimality, as it turns out, is for P also to have the property of *essentially complete recourse*, in the sense that for $k = 1, \cdots, N$ the multifunction

$$\begin{aligned} D^k: \xi \to D^k(\xi) &= \{(u_1, \cdots, u_k) | u \in D(\xi)\} \\ &= \text{projection of } D(\xi) \text{ on } R^{n_1} \times \cdots \times R^{n_k} \end{aligned} \tag{1.8}$$

is $\mathscr{F}_k$-measurable. (In this case, the constraint multifunction D is said to be *essentially nonanticipative.*) *Henceforth, we assume the problem* P *to be endowed with both strict feasibility and essentially complete recourse, as well as all other properties of* U, f_i *and* D *already mentioned.*

The optimality condition to be studied below involves the function

$$h : \Xi \times R^n \times R^m_+ \times R^n \to R$$

defined by

$$h(\xi, u, y, p) = f_0(\xi, u) + \sum_{i=1}^{m} y_i f_i(\xi, u) - u \cdot p. \tag{1.9}$$

This acts much like the Hamiltonian in control theory.

The *Lagrangian* associated with the problem P is defined to be the function

$$I_h(u, y, p) = E\{h(\xi, u(\xi), y(\xi), p(\xi))\} \quad \text{for } (u, y, p) \in \mathscr{U} \times \mathscr{Y} \times \mathscr{M}_1, \tag{1.10}$$

where

$$\mathscr{Y} = \{y = (y_1, \cdots, y_m) \in L^1_m | y_i(\xi) \geqq 0 \text{ a.s. for } i = 1, \cdots, m\},$$

$$\mathscr{M}_1 = \{p = (p_1, \cdots, p_N) \in L^1_n | E^k\{p_k(\xi)\} = 0 \text{ a.s. for } k = 1, \cdots, N\}.$$

(Here $p_k(\xi) \in R^{n_k}$.) The set $\mathscr{Y}$ is convex, while $\mathscr{M}_1$ is a linear subspace. In fact, as is easy to verify from the definitions, $\mathscr{M}_1$ and $\mathscr{N}_\infty$ are complementary to each other with respect to the natural pairing between L^1_n and L^∞_n.

$$\mathscr{M}_1 = \mathscr{N}_\infty^\perp \quad \text{and} \quad \mathscr{N}_\infty = \mathscr{M}_1^\perp.$$

Our growth conditions on the functions f_i imply that $I_h(u, y, p)$ is finite throughout $\mathscr{U} \times \mathscr{Y} \times \mathscr{M}_1$, and, of course, convex in u and affine in (y, p).

A *saddle point* of I_h with respect to minimization in u and maximization in (y, p) is an element $(\bar{u}, \bar{y}, \bar{p})$ of $\mathscr{U} \times \mathscr{Y} \times \mathscr{M}_1$ satisfying

$$I_h(\bar{u}, y, p) \leqq I_h(\bar{u}, \bar{y}, \bar{p}) \leqq I_h(u, \bar{y}, \bar{p}) \quad \text{for all } (u, y, p) \in \mathscr{U} \times \mathscr{Y} \times \mathscr{M}_1. \tag{1.11}$$

We shall prove in § 2 that the regularity conditions imposed on P ensure the existence of such a saddle point $(\bar{u}, \bar{y}, \bar{p})$, with $\bar{u}$ an optimal solution to P and $(\bar{y}, \bar{p})$ an optimal solution to an associated dual problem. (See [3] for a general exposition of the relation between the saddle points of a Lagrangian and the optimal solutions of the corresponding convex program and its dual.)

As is also shown in § 2, the saddle points $(\bar{u}, \bar{y}, \bar{p})$ of I_h are characterized by the following Kuhn–Tucker conditions, whose satisfaction for some $(\bar{y}, \bar{p})$ is therefore necessary and sufficient for the optimality of $\bar{u}$ in P:

(a) $\bar{u} \in \mathscr{N}_\infty$ *and*

$$\bar{u}(\xi) \in U(\xi) \text{ almost surely} \tag{1.12}$$

$$f_i(\xi, \bar{u}(\xi)) \leqq 0 \quad \text{for } i = 1, \cdots, m \quad \text{almost surely;} \tag{1.13}$$

(b) $\bar{y} = (\bar{y}_1, \cdots, \bar{y}_m) \in L^1_m$ *and*

$$\bar{y}_i(\xi) \geqq 0 \quad \text{for } i = 1, \cdots, m \quad \text{almost surely,} \tag{1.14}$$

$$\bar{y}_i(\xi) f_i(\xi, \bar{u}(\xi)) = 0 \quad \text{for } i = 1, \cdots, m \quad \text{almost surely;} \tag{1.15}$$

(c) $\bar{p} \in \mathcal{M}_1$ *and*

(1.16) $$h(\xi, \bar{u}(\xi), \bar{y}(\xi), \bar{p}(\xi)) = \lim_{u \in U(\xi)} h(\xi, u, \bar{y}(\xi), \bar{p}(\xi)) \quad \textit{almost surely.}$$

The Kuhn–Tucker conditions show that if $\bar{y}$ and $\bar{p}$, the multipliers associated with P, are known or can be generated by an algorithmic procedure, a function $\bar{u} \in L_n^\infty$ is optimal for P if and only if it is nonanticipative and $\bar{u}(\xi)$ satisfies certain constraints "pointwise" for each $\xi \in \Xi$, namely (1.12), (1.16), and (1.13) with equality holding when $\bar{y}_i(\xi) > 0$. *Moreover, if* P *is such that the pointwise minimum in* (1.16) *is almost surely unique, as is true for example if $f_0(\xi, \cdot)$ is almost surely convex on $U(\xi)$, then the function $\bar{u} \in L_n^\infty$ is optimal if it merely satisfies* (1.12) *and* (1.16), *without regard to nonanticipativity and the other constraints*. Indeed, these other properties must then hold automatically for $\bar{u}$, since according to the above there does exist at least one optimal recourse function characterized by the Kuhn–Tucker conditions. This is discussed further in a more specialized context in § 3.

Essentially complete recourse plays a vital role in the derivation of these results. The importance of this kind of property was first brought out in [4] in connection with our work on a special case of P. It was shown in [4] that if a stochastic program with a two-stage constraint structure has *relatively complete recourse*, the multipliers appearing in the Kuhn–Tucker conditions may be chosen to be L^1-functions; one has to rely on esoteric elements of $(L^\infty)^*$ when this condition is not satisfied. It can be shown that essentially complete recourse is implied by relatively complete recourse in that setting (see the remarks in § 3 following Theorem 6). Essentially complete recourse is a more general and abstract condition demanding that at each stage k the set from which the decision u_k must be chosen, namely

$$D_k(\xi, u_1, \cdots, u_{k-1}) = \{u_k \in R^{n_k} | (u_1, \cdots, u_{k-1}, u_k) \in D^k(\xi)\},$$

really depends only on past decisions and observations, and one therefore does not have to restrict further to an intersection relative to all possible future observations (an implicit constraint induced by the need to maintain availability of recourse under all circumstances).

In a companion paper [5], essentially complete recourse was used extensively, first in the justification of the dynamic programming technique for optimal recourse problems, but then also to obtain a system of L^1-multipliers, in fact a summable martingale, that can be associated with the nonanticipativity restriction on the recourse functions. However, our concern in [5] was only with such multipliers. The model was formulated directly in terms of the nonanticipative constraint multifunction D; no structure of D in terms of inequality constraints as in (1.3) was explicitly introduced, and hence there was no multiplier vector $y(\xi)$. The existence of multipliers associated with the nonanticipativity restriction was first pointed out in [6].

2. Basic results. Our first theorem shows that the regularity conditions imposed on the recourse problem P guarantee the existence of an optimal solution $\bar{u}$, and that such functions $\bar{u}$ correspond to saddle-points $(\bar{u}, \bar{y}, \bar{p})$ of l_h. We proceed by observing that the question can be settled through reducing P to an equivalent problem without explicit inequality constraints. We then utilize the key result of [5] to complete the proof. The second theorem demonstrates that the saddle points of l_h can be characterized by the Kuhn–Tucker conditions, and these therefore furnish necessary and sufficient conditions for optimality. The third theorem brings in the corresponding dual problem D.

THEOREM 1. *The Lagrangian l_h has at least one saddle point $(\bar{u}, \bar{y}, \bar{p})$ relative to $\mathcal{U} \times \mathcal{Y} \times \mathcal{M}_1$. Moreover, the components $\bar{u}$ of such saddle points are precisely the optimal recourse functions in* P.

Proof. First observe that P consists of minimizing over $\mathcal{N}_\infty$ the functional

$$I_f(u) = E\{f(\xi, u(\xi))\} = E\{f_0(\xi, u(\xi)) + \psi_{D(\xi)}(u(\xi))\},$$

where $\psi_{D(\xi)}$ is the indicator of $D(\xi)$. Since D is a measurable multifunction and f_0 is a normal convex integrand, we know f is a normal convex integrand [2, Thm. 2 and Cor. 4.2].

According to our assumptions, $D(\xi)$ is uniformly bounded and there is a summable function $\alpha: \Xi \to R$ such that

$$u \in D(\xi) \Rightarrow |f(\xi, u)| \leqq \alpha(\xi).$$

Furthermore, by strict feasibility there exist $\tilde{u} \in \mathcal{N}_\infty$ and $\varepsilon > 0$ such that (1.7) holds.

These facts put us in the framework of [5, Thm. 2] and furnish not only the existence of an optimal solution $\bar{u}$ to P but also the characterization of such a function $\bar{u}$ as the first component of a saddle point $(\bar{u}, \bar{p})$ of the *reduced Lagrangian*

$$(2.1)\quad L(u, p) = E\{f(\xi, u(\xi)) - u(\xi) \cdot p(\xi)\} = I_f(u) - \langle u, p\rangle \quad \text{for } (u, p) \in L_n^1 \times \mathcal{M}_1.$$

The existence of an optimal solution is seen as follows. The subspace $\mathcal{N}_\infty$, being representable as

$$\mathcal{M}_1^\perp = \{u \in L_n^1 | \langle u, p\rangle = 0 \text{ for all } p \in \mathcal{M}_1\},$$

is closed in the weak topology $w(L_n^\infty, L_n^1)$. The functional I_f on L_n^∞ is lower semicontinuous in this topology, because it is representable as the conjugate of the functional I_{f^*} on L_n^1, where f^* is the conjugate integrand ([1, Thm. 2] and [7, Thm. 2]). The sets

$$\{u \in \mathcal{N}_\infty | I_f(u) \leqq \mu\}, \quad \mu \in R,$$

are therefore closed in this topology, in fact compact by the uniform boundedness of $D(\xi)$, since

$$(2.2)\qquad I_f(u) < +\infty \Rightarrow u(\xi) \in D(\xi) \quad \text{a.s.}$$

The nonempty sets in this nest of compact sets therefore have a nonempty intersection, and this consists obviously of optimal solutions to P.

The existence of the multiplier $\bar{p}$ in [5, Thm. 2] is obtained by a more subtle argument, the details of which will not be repeated here. By our hypothesis, the convex functional I_f is finite and norm-continuous at a certain point $\tilde{u}$ of $\mathcal{N}_\infty$, and this furnishes by Fenchel's duality theorem a norm-continuous linear functional φ

on L_n^∞ such that φ vanishes on $\mathcal{N}_\infty$ and

$$\inf_{u \in \mathcal{N}_\infty} I_f(u) = \inf_{u \in L_n^\infty} \{I_f(u) - \varphi(u)\}.$$

The property of essentially complete recourse enters in showing that φ can actually be taken to be of the form $\varphi(u) = \langle u, \bar{p}\rangle$ for some $\bar{p}$ in L_n^1 (and hence in $\mathcal{M}_1 = \mathcal{N}_\infty^\perp$). This yields the existence of at least one saddle point $(\bar{u}, \bar{p})$ of L in (2.1), and it follows then by the usual reasoning in minimax theory that such saddle points characterize the optimal solutions $\bar{u}$ to P.

To complete the proof of Theorem 1, we must show that a pair $(\bar{u}, \bar{p})$ is a saddle point of the reduced Lagrangian L if and only if there exists $\bar{y} \in \mathcal{Y}$ such that $(\bar{u}, \bar{y}, \bar{p})$ is a saddle point of the Lagrangian I_h. The sufficiency of this condition is obvious from the fact that

$$L(u, p) = \sup_{y \in \mathcal{Y}} I_h(u, y, p). \tag{2.3}$$

(In view of (2.2), there is no loss of generality in replacing L_n^∞ by $\mathcal{U}$ in discussing saddle points of I_h).

Now consider any saddle point $(\bar{u}, \bar{p})$ of L. We have $\bar{u} \in \mathcal{U}$ and

$$L(\bar{u}, \bar{p}) = \sup_{p \in \mathcal{M}_1} L(\bar{u}, p) = \sup_{p \in \mathcal{M}_1} \sup_{y \in \mathcal{Y}} I_h(\bar{u}, y, p), \tag{2.4}$$

while on the other hand, using the fact already noted that the conjugate of I_f is I_{f^*} on L_n^1, we have

$$I_f(\bar{u}) - \langle \bar{u}, \bar{p}\rangle = L(u, p) = \inf_{u \in L_n^\infty} L(u, \bar{p}) = \inf_{u \in L^\infty} \{I_f(u) - \langle u, \bar{p}\rangle\} = -I_{f^*}(\bar{p}), \tag{2.5}$$

where by definition

$$-f^*(\xi, \bar{p}(\xi)) = \inf_{u \in R^n} \{f(\xi, u) - u \cdot \bar{p}(\xi)\}. \tag{2.6}$$

In order to verify for some $\bar{y} \in \mathcal{Y}$ that $(\bar{u}, \bar{y}, \bar{p})$ is a saddle point of I_h, it suffices in view of (2.4) to establish that

$$I_h(u, \bar{y}, \bar{p}) \geqq L(\bar{u}, \bar{p}) \quad \text{for all } u \in \mathcal{U},$$

or in other words that

$$E\{h(\xi, u(\xi), \bar{y}(\xi), \bar{p}(\xi))\} \geqq E\{f(\xi, \bar{u}(\xi)) - \bar{u}(\xi) \cdot \bar{p}(\xi)\} \quad \text{for all } u \in \mathcal{U}. \tag{2.7}$$

We know from (2.5) and (2.6) that

$$f(\xi, \bar{u}(\xi)) - \bar{u}(\xi) \cdot \bar{p}(\xi) = \inf_{u \in R^n} \{f(\xi, u) - u \cdot \bar{p}(\xi)\} \quad \text{almost surely.}$$

Thus $\bar{u}(\xi)$ is almost surely an optimal solution to the convex programming problem

$$\begin{aligned} &\text{minimize } f_0(\xi, u) - u \cdot p(\xi) \quad \text{over all } u \in U(\xi) \\ &\text{satisfying } f_i(\xi, u) \leqq 0 \quad \text{for } i = 1, \cdots, m. \end{aligned} \tag{2.8}$$

However, this problem is strictly feasible almost surely, due to the assumed existence of $\tilde{u} \in \mathscr{U}$ and $\varepsilon > 0$ satisfying (1.6), and it therefore has almost surely a Kuhn–Tucker vector, i.e. a vector $y \in R^m_+$ such that (cf. (1.9)):

$$\inf_{u \in U(\xi)} h(\xi, u, y, \bar{p}(\xi)) = \inf \text{ in } (2.8) = f(\xi, \bar{u}(\xi)) - \bar{u}(\xi) \cdot \bar{p}(\xi).$$

Let $Y(\xi)$ denote the set of all vectors $y \in R^m_+$ such that

$$h(\xi, u, y, \bar{p}(\xi)) \geqq f(\xi, \bar{u}(\xi)) - \bar{u}(\xi) \cdot \bar{p}(\xi) \quad \text{for all } u \in U(\xi). \tag{2.9}$$

As we have just seen, $Y(\xi) \neq \varnothing$ almost surely. Let A denote a countable dense subset of R^n. Since for each $y \in R^m_+$ the function $h(\xi, \cdot, y, \bar{p}(\xi))$ is finite, lower semicontinuous (l.s.c.), and convex on $U(\xi)$ (a convex set with nonempty interior), it is continuous on the interior of $U(\xi)$ and relative to all line segments in $U(\xi)$, and hence

$$\inf_{u \in U(\xi)} h(\xi, u, y, \bar{p}(\xi)) = \inf_{u \in U(\xi) \cap A} h(\xi, u, y, \bar{p}(\xi)).$$

Thus $U(\xi)$ can be replaced by $U(\xi) \cap A$ in (2.9) without affecting the nature of the condition on y. This yields the representation

$$Y(\xi) = \bigcap_{a \in A} Y_a(\xi), \tag{2.10}$$

where $Y_a(\xi)$ denotes the set of all $y \in R^m_+$ satisfying

$$h(\xi, a, y, \bar{p}(\xi)) \geqq f(\xi, \bar{u}(\xi)) - \bar{u}(\xi) \cdot \bar{p}(\xi),$$

or more specifically, is given for each ξ in the (Borel measurable) set $U^{-1}(a)$ by

$$Y_a(\xi) = \left\{ y \in R^m_+ \,\middle|\, \sum_{i=1}^{m} y_i f_i(\xi, a) \geqq f_0(\xi, \bar{u}(\xi)) - f_0(\xi, a) \right\},$$

while for other $\xi \in \Xi$ simply $Y_a(\xi) = R^m_+$. Each of the multifunctions $Y_a: \xi \to Y_a(\xi)$ is close-valued and Borel measurable [2, Cor. 4.3], and hence so is Y as the intersection of a countable collection in (2.10) [2, Cor. 1.3]. It follows that Y has a Borel measurable selection where it is nonempty-valued [2, Cor. 1.1]. Since $Y(\xi) \neq \varnothing$ almost surely, we therefore have the existence of a Borel measurable function $\bar{y}: \Xi \to R^m_+$ such that almost surely $\bar{y}(\xi) \in Y(\xi)$, i.e.

$$h(\xi, u, \bar{y}(\xi), \bar{p}(\xi)) \geqq f(\xi, \bar{u}(\xi)) - \bar{u}(\xi) \cdot \bar{p}(\xi) \quad \text{for all } u \in U(\xi). \tag{2.11}$$

We claim (2.11) implies $\bar{y}(\xi)$ is summable in ξ, so that actually $\bar{y} \in \mathscr{Y}$. Indeed, for the function $\tilde{u}$ in our strict feasibility assumption we can set $u = \tilde{u}(\xi)$ in (2.11) to obtain (almost surely)

$$\begin{aligned} f_0(\xi, \tilde{u}(\xi)) - \varepsilon \sum_{i=1}^{m} \bar{y}_i(\xi) - \tilde{u}(\xi) \cdot \bar{p}(\xi) &\geqq f_0(\xi, \tilde{u}(\xi)) - \tilde{u}(\xi) \cdot \bar{p}(\xi) + \sum_{i=1}^{m} \bar{y}_i(\xi) f(\xi, \tilde{u}(\xi)) \\ &\geqq f(\xi, \bar{u}(\xi)) - \bar{u}(\xi) \cdot \bar{p}(\xi) \\ &= f_0(\xi, \bar{u}(\xi)) - \bar{u}(\xi) \cdot \bar{p}(\xi) \end{aligned}$$

and thus for $i=1,\cdots,m$ (almost surely)

$$0 \leqq \varepsilon \bar{y}_i(\xi) \leqq f_0(\xi, \tilde{u}(\xi)) - f_0(\xi, \bar{u}(\xi)) - (\tilde{u}(\xi) - \bar{u}(\xi)) \cdot \bar{p}(\xi). \tag{2.12}$$

The right side of (2.12) is, of course, summable in ξ, and hence so is $\bar{y}_i(\xi)$.

We have thus established the existence of $\bar{y} \in \mathcal{Y}$ satisfying (2.11). But (2.11) implies (2.7) and therefore, as already argued, that $(\bar{u}, \bar{y}, \bar{p})$ is a saddle point of I_h. This ends the proof of Theorem 1.

COROLLARY. *The restricted Lagrangian*

$$I_\ell(u, y) = E\{\ell(\xi, u(\xi), y(\xi))\} \quad \text{for } (u, y) \in (\mathcal{U} \cap \mathcal{N}_\infty) \times \mathcal{Y}, \tag{2.13}$$

where

$$\ell(\xi, u, y) = f_0(\xi, u) + \sum_{i=1}^{m} y_i f_i(\xi, u), \tag{2.14}$$

has at least one saddle point $(\bar{u}, \bar{y})$ *relative to* $(\mathcal{U} \cap \mathcal{N}_\infty) \times \mathcal{Y}$. *Moreover, the components* $\bar{u}$ *of such saddle points are precisely the optimal recourse functions in* P.

Proof. Let $(\bar{u}, \bar{y}, \bar{p})$ be a saddle point of I_h relative to $\mathcal{U} \times \mathcal{Y} \times \mathcal{M}_1$, as exists by Theorem 1. Since $\bar{p} \in \mathcal{M}_1 = \mathcal{N}_\infty^\perp$, we have

$$I_h(u, y, \bar{p}) = I_\ell(u, y) \quad \text{for } (u, y) \in (\mathcal{U} \cap \mathcal{N}_\infty) \times \mathcal{Y},$$

and hence $(\bar{u}, \bar{y})$ is a saddle point of I_ℓ relative to $(\mathcal{U} \cap \mathcal{N}_\infty) \times \mathcal{Y}$. The existence of at least one such saddle point, together with the fact that P is equivalent to minimizing the functional

$$I_f(u) = \sup_{y \in \mathcal{Y}} I_\ell(u, y) \quad \text{for } u \in \mathcal{U} \cap \mathcal{N}_\infty,$$

yields the characterization of solutions $\bar{u}$ by the usual minimax considerations.

THEOREM 2. *An element* $(\bar{u}, \bar{y}, \bar{p})$ *is a saddle point of the Lagrangian* I_h *relative to* $\mathcal{U} \times \mathcal{Y} \times \mathcal{M}_1$ *if and only if the Kuhn–Tucker conditions* (a), (b), (c) *are satisfied.*

Proof. In either case we have $\bar{u} \in \mathcal{U}$, $\bar{y} \in \mathcal{Y}$ and $\bar{p} \in \mathcal{M}_1$. If $(\bar{u}, \bar{y}, \bar{p})$ is a saddle point, then $\bar{u}$ is optimal for P by Theorem 1, and in particular $\bar{u} \in \mathcal{N}_\infty$. Thus in showing the equivalence we can limit attention to the case where also $\bar{u} \in \mathcal{N}_\infty$. Then $\langle \bar{u}, p \rangle = 0$ for all $p \in \mathcal{M}_1$, so that $I_h(\bar{u}, y, p) = I_h(\bar{u}, y, \bar{p})$, and the saddle point condition can just as well be expressed as

$$\sup_{y \in \mathcal{Y}} I_h(\bar{u}, y, \bar{p}) = I_h(\bar{u}, \bar{y}, \bar{p}) = \inf_{u \in \mathcal{U}} I_h(u, \bar{y}, \bar{p}). \tag{2.15}$$

The left half of (2.13) is trivially equivalent to

$$\sup_{y \in R_+^m} h(\xi, \bar{u}(\xi), y, \bar{p}(\xi)) = h(\xi, \bar{u}(\xi), \bar{y}(\xi), \bar{p}(\xi)) \quad \text{a.s.},$$

and this is identical to (1.13) plus (1.15).

It remains only to show that the second equality in (2.15) implies (1.16), the opposite implication being immediate. Define the integrand j on $\Xi \times R^n$ by

$$j(\xi, u) = h(\xi, u, \bar{y}(\xi), \bar{p}(\xi)),$$

this value being interpreted as $+\infty$ for $u \notin U(\xi)$, so that

$$U(\xi) = \{u \in R^n \mid j(\xi, u) < +\infty\}.$$

Our hypotheses say that $j(\xi, u)$ is l.s.c. convex in u and Borel measurable in ξ, hence (since int $U(\xi) \neq \emptyset$) j is a (Borel) normal convex integrand [1, Lemma 2]. Furthermore, the "growth" conditions on the functions f_i imply for each bounded set $K \subset R^n$ the existence of a summable function $\gamma: \Xi \to R$ such that

$$|j(\xi, u)| \leqq \gamma(\xi) \quad \text{for all } u \in U(\xi) \cap K.$$

The right half of (2.15) thus can be regarded as the assertion that

$$I_j(\bar{u}) = \inf_{u \in L_n^\infty} I_j(u), \tag{2.16}$$

where

$$I_j(u) = E\{j(\xi, u(\xi))\}.$$

On the other hand, (1.16) can be restated as

$$j(\xi, \bar{u}(\xi)) = \inf_{u \in R^n} j(\xi, u) \quad \text{a.s.} \tag{2.17}$$

The question is thus reduced to that of the equivalence of (2.16) and (2.17), which is answered affirmatively by the theory of normal integrands and integral functionals. (In particular, the two properties can be expressed in terms of $0 \in \partial I_j(\bar{u})$ and $0 \in \partial j(\xi, \bar{u}(\xi))$, and then [7, Cor. 1B and Thm. 2] can be invoked.) Theorem 2 is thereby established.

We have mentioned in § 1 that the multipliers $\bar{y}$ and $\bar{p}$ for P solve a certain dual problem. This will now be described. Define the function g on $\Xi \times R^m \times R^n$ by

$$g(\xi, y, p) = \begin{cases} \inf_{u \in U(\xi)} h(\xi, u, y, p) & \text{if } y \in R_+^m, \\ -\infty & \text{if } y \notin R_+^m. \end{cases} \tag{2.18}$$

It will be shown below that $-g$ is a normal convex integrand. Let

$$I_g(y, p) = E\{g(\xi, y(\xi), p(\xi))\} \quad \text{for } (y, p) \in L_m^1 \times L_n^1. \tag{2.19}$$

The *dual problem* associated with P is taken to be:

D $\qquad$ Maximize $I_g(y, p)$ over all $(y, p) \in \mathcal{Y} \times \mathcal{M}_1$.

THEOREM 3. *The functional I_g in* D *is well-defined and concave, with*

$$I_g(y, p) = \inf_{u \in \mathcal{U}} I_h(u, y, p) \quad \textit{for all } (y, p) \in \mathcal{Y} \times \mathcal{M}_1. \tag{2.20}$$

Thus optimal solutions to D *exist, and they are precisely the components $(\bar{y}, \bar{p})$ of the saddle points $(\bar{u}, \bar{y}, \bar{p})$ of the Lagrangian I_h. In particular,*

$$\min \text{P} = \max \text{D}.$$

Proof. We begin by proving that $-g$ is a (Borel-)normal convex integrand. There exists in $\mathcal{U}$ a countable subcollection $\mathcal{U}'$ such that $U(\xi)$ is almost surely the closure of the set $\{u(\xi)|u \in \mathcal{U}'\}$. (This follows from the measurability of the multifunction U via Castaing's theorem; cf. [2, Thm. 1].) Then by convexity

$$g(\xi, y, p) = \inf_{u \in \mathcal{U}'} h(\xi, u(\xi), y, p) \quad \text{a.s.} \quad \text{for } y \in R^m_+. \tag{2.21}$$

For each $u \in \mathcal{U}'$, define

$$g_u(\xi, y, p) = \begin{cases} h(\xi, u(\xi), y, p) & \text{if } y \in R^m_+, \\ -\infty & \text{if } y \notin R^m_+. \end{cases}$$

Then $-g$ is a normal convex integrand by virtue of our regularity assumptions, and we have from (2.21) the representation

$$g(\xi, y, p) = \inf_{u \in \mathcal{U}'} g_u(\xi, y, p) \quad \text{a.s.}$$

Since the collection is countable, this implies $-g$ is a normal convex integrand [2, Cor. 4.1].

Normality ensures that $g(\xi, y(\xi), p(\xi))$ is measurable in ξ whenever $y(\cdot)$ and $p(\cdot)$ are. On the other hand, fixing any $u \in \mathcal{U}$ we have for all $y \in L^1_m$ and $p \in L^1_n$ the bound

$$g(\xi, y(\xi), p(\xi)) \leqq f_0(\xi, u(\xi)) + \sum_{i=1}^{m} y_i(\xi) f_i(\xi, u(\xi)) - u(\xi) \cdot p(\xi),$$

where the right side is summable. Thus $I_g(y, p)$ is always unambiguously a real number or $-\infty$. The concavity of I_g is obvious.

We establish (2.20) by fixing any (y, p) in $\mathcal{Y} \times \mathcal{M}_1$ and considering the integrand

$$j(\xi, u) = \begin{cases} h(\xi, u, y(\xi), p(\xi)) & \text{if } u \in U(\xi), \\ +\infty & \text{if } u \notin U(\xi). \end{cases} \tag{2.22}$$

The situation is extremely close to the one at the end of the proof of Theorem 2; j is a normal convex integrand, and we get from the theory of integral functionals that

$$\sup_{u \in L^\infty_n} \{\langle q, u\rangle - I_j(u)\} = I_{j^*}(q) \quad \text{for all } q \in L^1_n, \tag{2.23}$$

where

$$j^*(\xi, q(\xi)) = \sup_{u \in R^n} \{q(\xi) \cdot u - j(\xi, u)\}.$$

Taking $q = 0$, we turn the latter into

$$-j^*(\xi, 0) = g(\xi, p(\xi), y(\xi))$$

by (2.22) and (2.18), and then (2.23) becomes the equation in (2.20).

The rest of Theorem 3 is evident from (2.20) and the existence of a saddle point of I_h in Theorem 1.

3. Special structures. So far, it has been convenient and useful to endow P with as little structure as possible. This level of generality is rarely, if ever, needed in practice. The main purpose of this section, and the next one, is to consider recourse problems that possess some of the structural characteristics most commonly encountered in applications.

An initial observation may be made about the differentiable case, i.e. where $U(\xi)=R^n$ and the functions $u\mapsto f_i(\xi,u)$ are all differentiable with gradients denoted by $\nabla f_i(\xi,u)$. Then (1.16) of the Kuhn–Tucker conditions becomes

$$\nabla f_0(\xi,\bar u(\xi))+\sum_{i=1}^m \bar y_i(\xi)\nabla f_i(\xi,\bar u(\xi))=\bar p(\xi)\quad \text{a.s.,}\tag{3.1}$$

and hence part (c) of the conditions asserts simply that

$$E^k\left\{\nabla f_0(\xi,\bar u(\xi))+\sum_{i=1}^m \bar y_i(\xi)\nabla f_i(\xi,\bar u(\xi))\right\}=0\quad \text{a.s.}\quad \text{for } k=1,\cdots,N.\tag{3.2}$$

A. *The separable case.* By SP we denote a version of P that satisfies all the regularity conditions laid out in § 1 and is also *separable*, by which we mean that

(i) $$U(\xi)=\mathop{\times}_{k=1}^{N} U_k(\xi),$$

(ii) $$f_i(\xi,u)=\sum_{k=1}^N f_{ik}(\xi,u_k)\quad \text{for } i=0,1,\cdots,m,$$

where the multifunctions $U_k:\xi\mapsto U_k(\xi)\subset R^{n_k}$ are $\mathscr{F}_k$-measurable, and the functions $\xi\mapsto f_{ik}(\xi,u_k)$ are $\mathscr{F}_k$-measurable relative to the set

$$U_k^{-1}(u_k)=\{\xi\in\Xi \mid u_k\in U_k(\xi)\}\in\mathscr{F}_k.$$

The function h (as defined by (1.9)) is also separable, in the sense that

$$h(\xi,u,y,p)=\sum_{k=1}^N [\ell_k(\xi,u_k,y)-u_k\cdot p_k],\tag{3.3}$$

where

$$\ell_k(\xi,u_k,y)=f_{0k}(\xi,u_k)+\sum_{i=1}^m y_i f_{ik}(\xi,u_k)\tag{3.4}$$

and the functions $\xi\mapsto\ell_k(\xi,u_k,y)$ are $\mathscr{F}_k$-measurable relative to $U_k^{-1}(u_k)$.

Since SP possesses all the properties of P, the problem is solvable and the Kuhn–Tucker conditions (a), (b), (c) are necessary and sufficient for optimality. We shall show that (c) can be replaced by:

(sc) *for* $k=1,\cdots,N$ *one has*

$$\ell_k(\xi,\bar u_k(\xi),E^k\bar y(\xi))=\min_{u_k\in U_k(\xi)}\ell_k(\xi,u_k,E^k\bar y(\xi))\quad \textit{a.s.}\tag{3.5}$$

where

$$(E^k\bar y)(\xi)=E^k\{\bar y(\xi)\}\quad (\textit{conditional expectation given } \mathscr{F}_k).\tag{3.6}$$

Of course $E^k\bar{y}$ is $\mathcal{F}_k$-measurable by definition, so *the process* $\{E^k\bar{y}, k = 1, \cdots, N\}$ *is nonanticipative*. Note that everything in the expression (3.5) is $\mathcal{F}_k$-measurable, and therefore the "almost surely" can be interpreted with respect to the restriction of the probability σ to $\mathcal{F}_k$. Thus the minimization is entirely in terms of information pertinent to stage k and independent of the future. In particular, for the nest of sigma-fields $\mathcal{F}_k$ corresponding to the sequential notation $\xi = (\xi_1, \cdots, \xi_N)$ at the beginning of § 1, ξ can be replaced essentially by $\xi^k = (\xi_1, \cdots, \xi_k)$ throughout (3.5). The decision taken at stage k is then represented as a solution $\bar{u}_k(\xi^k)$ to an optimization problem depending only on the past information ξ^k and a vector $E^k\bar{y}(\xi^k)$ of expected "prices."

THEOREM 4. *A function $\bar{u}$ solves the separable optimal recourse problem* SP *if and only if there is a multiplier function $\bar{y}$ such that $(\bar{u}, \bar{y})$ satisfies* (a) *and* (b) *of the general Kuhn–Tucker conditions and* (sc) *above.*

Proof. From the Corollary to Theorem 1, we know that $\bar{u}$ is optimal if and only if $\bar{u} \in \mathcal{U} \cap \mathcal{N}_\infty$ and there exists $\bar{y} \in \mathcal{Y}$ such that

$$\sup_{y \in \mathcal{Y}} I_\ell(\bar{u}, y) = I_\ell(\bar{u}, \bar{y}) = \inf_{u \in \mathcal{U} \cap \mathcal{N}_\infty} I_\ell(u, \bar{y}). \tag{3.7}$$

The left half of (3.7) is equivalent to

$$\sup_{y \in R^m_+} \ell(\xi, \bar{u}(\xi), y) = \ell(\xi, \bar{u}(\xi), \bar{y}(\xi)),$$

which means that (1.13) and (1.15) hold (and hence all of (a) and (b)). It remains only to show that the right half of (3.7) is equivalent to (sc). But separability implies

$$I_\ell(u, \bar{y}) = \sum_{k=1}^{N} I_{\ell_k}(u_k, \bar{y}) \quad \text{for all } u \in \mathcal{U} \cap \mathcal{N}_\infty, \tag{3.8}$$

where

$$I_{\ell_k}(u_k, \bar{y}) = E\{\ell_k(\xi, u_k(\xi), \bar{y}(\xi))\} = E\{\ell_k(\xi, u_k(\xi), E^k\bar{y}(\xi))\}, \tag{3.9}$$

the last equality being true because the function u_k is $\mathcal{F}_k$-measurable and ℓ_k is affine in the multiplier y. For $k = 1, \cdots, N$, define the integrand r_k on $\Xi \times R^{n_k}$ by

$$r_k(\xi, u_k) = \begin{cases} \ell_k(\xi, u_k, E^k\bar{y}(\xi)) & \text{if } u_k \in U_k(\xi), \\ +\infty & \text{if } u_k \notin U_k(\xi). \end{cases} \tag{3.10}$$

Then for functions $u_k \in L^\infty_{n_k}(\Xi, \mathcal{F}_k, \sigma)$ we have from (3.9)

$$E\{r_k(\xi, u_k(\xi))\} = \begin{cases} I_{\ell_k}(u_k, \bar{y}) & \text{if } u_k(\xi) \in U_k(\xi) \quad \text{a.s.}, \\ +\infty & \text{otherwise}. \end{cases}$$

The right half of (3.7) is therefore identical to the assertion that for $k = 1, \cdots, N$:

(3.11) the minimum of $I_{r_k}(u_k) = E\{r_k(\xi, u_k(\xi))\}$ over all $u_k \in L^\infty_{n_k}(\Xi, \mathcal{F}_k, \sigma)$ is attained at $\bar{u}_k$, while condition (3.5) is the same as

$$\text{(3.12)} \qquad \begin{array}{l}\text{the minimum of } r_k(\xi, u_k) \text{ over all}\\ u_k \in R^{n_k} \text{ is attained at } \bar{u}_k(\xi) \text{ almost surely.}\end{array}$$

The equivalence of (3.11) and (3.12) follows from our regularity assumptions exactly as did the equivalence of (2.16) and (2.17) in the proof of Theorem 2: each r_k is an $\mathscr{F}_k$-normal convex integrand. This completes the proof of Theorem 4.

The Kuhn–Tucker conditions in this "decomposed" form have a number of significant features that render them attractive from a computational viewpoint. Notably, if at stage k the multiplier function $\bar{y}^k$ is known and the minimum in (3.5) is uniquely attained almost surely, then the minimizing points must be the values $\bar{u}_k(\xi)$ of the unique optimal decision function $\bar{u}_k$ associated with this stage. In other words, the requirement of $\mathscr{F}_k$-measurability is automatically taken care of, and there is no need to worry about the ultimate satisfaction of the constraints $f_i(\xi, \bar{u}(\xi)) \leqq 0$.

We remark also that in the differentiable case, with $U_k(\xi) = R^{n_k}$ for all k, condition (3.5) takes on the form

$$\text{(3.13)} \qquad \nabla f_{0k}(\xi, \bar{u}_k(\xi)) + \sum_{i=1}^{m} E^k \bar{y}_i(\xi) \nabla f_{ik}(\xi, \bar{u}_k(\xi)) = 0 \quad \text{a.s. } (\mathscr{F}_k).$$

The structure of separability also leads to a special dual problem associated with SP. For $k = 1, \cdots, N$, define the function g_k on $\Xi \times R^m$ by

$$\text{(3.14)} \qquad g_k(\xi, y) = \inf_{u_k \in U_k(\xi)} \ell_k(\xi, u_k, y) \quad \text{if } y \in R^m_+.$$

Then $-g_k$ is an $\mathscr{F}_k$-normal convex integrand, and the functional

$$\text{(3.15)} \qquad I_{g_k}(y) = E\{g_k(\xi, y(\xi))\} \quad \text{for } y \in L^1_m$$

is well-defined, concave (with $-\infty$ as a possible value) and satisfies

$$\text{(3.16)} \qquad I_{g_k}(y) = \inf_{u_k \in \mathscr{U}_k} I_{\ell_k}(u_k, y) \quad \text{for all } \mathscr{F}_k\text{-measurable } y \in \mathscr{Y},$$

where

$$\text{(3.17)} \qquad \mathscr{U}_k = \{u_k \in L^\infty_{n_k}(\Xi, \mathscr{F}_k, \sigma) | u_k(\xi) \in U_k(\xi) \text{ a.s. } (\mathscr{F}_k)\}.$$

These facts are established almost exactly as they were for g and I_g in the proof of Theorem 3.

As the special dual problem for SP, we introduce:

$$\text{SD} \qquad \text{Maximize } \sum_{k=1}^{N} I_{g_k}(E^k y) \text{ over all } y \in \mathscr{Y}.$$

The following result is then immediate from the decomposition

$$\text{(3.18)} \qquad I_\ell(u, y) = \sum_{k=1}^{N} I_\ell(u_k, E^k y) \quad \text{for } (u, y) \in \mathscr{U} \times \mathscr{Y}$$

and the fact that the Kuhn–Tucker conditions (a), (b), (sc) in Theorem 4 characterize the saddle points of this expression.

THEOREM 5. *The dual problem* SD *has optimal solutions, and they are precisely the components* $\bar{y}$ *of the pairs* $(\bar{u}, \bar{y})$ *satisfying the Kuhn–Tucker conditions* (a), (b), (sc). *In particular,*

$$\min \text{SP} = \max \text{SD}.$$

B. *Linear recourse models.* By LP we denote a version of SP that can be formulated as follows:

$$\text{LP}\qquad \begin{aligned} &\text{Minimize } E\left\{\sum_{k=1}^{N} c_k \cdot u_k(\xi^k)\right\} \\ &\text{subject to } \sum_{k=1}^{j} A_{jk}u_k(\xi^k) \geqq b_j \quad \text{a.s.} \quad \text{for } j=1,\cdots,N, \end{aligned}$$

where $c_k \in R^{n_k}$, $b_j \in R^{m_k}$, $A_{jk} \in R^{m_k \times n_k}$ and $\xi^k = (\xi_1, \cdots, \xi_k)$ with $\xi_k = (c_k, A_{kk}, \cdots, A_{Nk}, b_k)$. Thus the vectors c_k and b_k and matrices A_{jk} are random variables whose values become known in stage k, and we are in the sequential notational setting at the beginning of § 1 with $\xi = (\xi_1, \cdots, \xi_N)$. It is required that

$$u_k \in L^{\infty}_{n_k}(\Xi^k, \mathscr{F}^k, \sigma^k), \tag{3.19}$$

where $(\Xi^k, \mathscr{F}^k, \sigma^k)$ is the marginal probability space of the random variable ξ^k, i.e. of the random elements observed in the first k stages.

This formulation differs slightly from the previous pattern in having (3.19) in place of the $\mathscr{F}_k$-measurability of u_k as a function of Ξ (with $\mathscr{F}_k$ the "cylindrical extension" of $\mathscr{F}^k$ relative to Ξ, as introduced in § 1 for the setting where $\xi = (\xi_1, \cdots, \xi_N)$). In simpler terms, the recourse function is taken to be nonanticipative, rather than just essentially nonanticipative. However, the two formulations are equivalent as long as we are not concerned with the multipliers $\bar{p}(\xi)$, and this is justified in the present context by Theorem 3. (In introducing $\bar{p} \in L^1_n$, we need to regard the recourse function u as an element of L^{∞}_n and therefore must admit, as negligible, alterations of $u_k(\xi^k)$ on a set of ξ-values of probability zero, even if these involve $\xi_{n+1}, \cdots, \xi_N$.) Incidentally, in contrast to this equivalence, one cannot change the "almost surely" in the constraints of LP without risking a disastrous effect on the problem. This is shown by counterexamples in [8], where a condition on the probability measure σ is also developed which ensures against the discrepancy.

As with SP, we assume that LP satisfies all the regularity conditions we have imposed on P. Actually, the convexity, lower semicontinuity and measurability conditions are trivially satisfied; note that $U_k(\xi) = R^{n_k}$, while each f_i is an affine function of u_k with random variables as coefficients. The uniform boundedness assumption requires that for all realizations of ξ the polyhedron generated by the constraints of LP lies within a fixed ball. For the case where the matrices A_{jk} are nonrandom—or equivalently, have a degenerate distribution—a sufficient condition for uniform boundedness is given by Olsen [9, Lemma 2.4]; cf. also [10]. Various sufficient conditions for strict feasibility can easily be found. For example, one such criterion can be derived from the results of Isofescu and Theodorescu [11] for systems of stochastic linear inequalities.

Problem LP has a block-triangular structure which makes it easy to see more specifically when the property of essentially complete recourse is present. Consider the following decision procedure. In the first stage (having observed $\xi_1 = (c_1, A_{11}, b_1)$) we choose u_1 satisfying $A_{11}u_1 \geqq b_1$. In the second stage (having observed ξ_2) we choose u_2 satisfying $A_{22}u_2 \geqq b_2^+$, where $b_2^+ = b_2 - A_{21}u_1$. And so forth: in the kth stage (having observed ξ_k) we choose u_k satisfying

$$A_{kk}u_k \geqq b_k^+, \quad \text{where } b_k^+ = b_k - \sum_{i=1}^{k-1} A_{ki}u_i. \tag{3.20}$$

One says that *relatively complete recourse* is present if this procedure can almost surely be continued to the end (i.e. to the choice of u_N), or in other words, if with probability one we will not encounter a stage where we are stymied by the emptiness of the u_k-polyhedron defined by the constraint system (3.20).

THEOREM 6. *Relatively complete recourse implies essentially complete recourse.*

Proof. Let us denote by $\Delta^k(\xi^k)$ the set of all $(u_1, \cdots, u_k)$ which can be generated by the first k stages of this procedure. Relatively complete recourse means that each element of $\Delta^k(\xi^k)$ is contained almost surely (with respect to the conditional distribution of $(\xi_{k+1}, \cdots, \xi_N)$ given $(\xi_1, \cdots, \xi_k)$) in the set $D^k(\xi)$ in (1.8), which consists of all $(u_1, \cdots, u_k)$ such that the procedure can be continued to the end when the total outcome of the random variable is $\xi = (\xi_1, \cdots, \xi_k, \xi_{k+1}, \cdots, \xi_N)$. Representing $\Delta^k(\xi^k)$ as the closure of a countable set, to each element of which this fact can be applied, we see from the closedness of $D^k(\xi)$ that

$$\Delta^k(\xi^k) \subset D^k(\xi)$$

almost surely (conditionally, given ξ^k). But trivially, the opposite inclusion is universally valid by the definition of $D^k(\xi)$. Therefore, relatively complete recourse is equivalent to the property that

$$D^k(\xi) = \Delta^k(\xi^k) \quad \text{a.s.} \tag{3.21}$$

(in the sense of the overall distribution of ξ). Of course, (3.21) implies that $D^k(\xi)$ essentially depends only on ξ^k, which is the property of essentially complete recourse.

Remark. The concept of relatively complete recourse, and with it Theorem 6, can easily be extended to SP and even to the general context of P, thereby also covering our use of the term in [4]. The multifunction U is itself assumed nonanticipative (as is true for instance in SP): the projection $U^k(\xi)$ consisting of all components $(u_1, \cdots, u_k)$ of elements u of $U(\xi)$ is thus assumed $\mathscr{F}_k$-measurable. The index set $\{1, \cdots, m\}$ is partitioned into subsets J_k such that, for $i \in J_k$, $f_i(\xi, u)$ is $\mathscr{F}_k$-measurable in ξ and depends only on the components $(u_1, \cdots, u_k)$ of u. Let $\Delta^k(\xi)$ consist of all elements $(u_1, \cdots, u_k)$ of $U^k(\xi)$ satisfying the constraints $f_i(\xi, u) \leqq 0$ for all indices $i \in J_1 \cup \cdots \cup J_k$. Then $\Delta^k(\xi)$ is $\mathscr{F}_k$-measurable in ξ. Relatively complete recourse is the property that each element of $\Delta^k(\xi)$ belongs to $D^k(\xi)$ almost surely (conditional probability given $\mathscr{F}_k$). This can also be expressed as above in terms of the almost sure feasibility of a "block-triangular" procedure for generating $u_1, \cdots, u_N$ sequentially. The proof of Theorem 6 remains valid in this case.

Our assumption of strict feasibility appears needed for the validity of the Kuhn–Tucker conditions (a), (b), (sc) of Theorem 4 in the case of LP, despite the linearities. This may be attributed to the (infinite-dimensional) constraint of nonanticipativity, even though the corresponding multipliers are suppressed in (sc).

The optimal recourse functions for LP, which exist according to Theorem 1 under the regularity conditions which have been imposed, are characterized as follows.

THEOREM 7. *In order that the function* $\bar{u} = (\bar{u}_1, \cdots, \bar{u}_N)$ *with* $\bar{u}_k \in L^\infty_{n_k}(\Xi^k, \mathcal{F}^k, \sigma^k)$ *be an optimal solution to* LP, *it is necessary and sufficient that the following conditions be satisfied for some function* $\bar{y} = (\bar{y}_1, \cdots, \bar{y}_N)$ *with* $\bar{y}_k \in L^1_{m_k}(\Xi^k, \mathcal{F}^k, \sigma^k)$, $k = 1, \cdots, N$:

$$A_{kk}\bar{u}_k(\xi^k) \geqq b^\dagger_k(\xi^k) \quad a.s., \tag{3.22}$$

$$\bar{y}_k(\xi^k) \geqq 0 \quad a.s., \tag{3.23}$$

$$\bar{y}_k(\xi^k) \cdot [b^\dagger_k(\xi^k) - A_{kk}\bar{u}_k(\xi^k)] = 0 \quad a.s., \tag{3.24}$$

$$\bar{y}_k(\xi^k) A_{kk} = c^\dagger_k(\xi^k) \quad a.s., \tag{3.25}$$

where

$$b^\dagger_k(\xi^k) = b_k - \sum_{j=1}^{k-1} A_{kj}\bar{u}_j(\xi^j), \tag{3.26}$$

$$c^\dagger_k(\xi^k) = c_k - \sum_{j=k+1}^{N} (E^k \bar{y}_j)(\xi^k) A_{jk}. \tag{3.27}$$

Proof. When conditions (a), (b) and (sc) of Theorem 4 are specialized to the present context, we get something slightly different. Namely, each $\bar{y}_k$ would appear in (3.23) and (3.24) as a function of all of ξ, while the expression in (3.25) would instead be $(E^k\bar{y}_k)(\xi^k)$. However, these conditions on $\bar{u}$ really involve only the latter expressions (and their expectations in earlier stages). Therefore, we can just as well apply E^k to (3.23) and (3.24), so that only $E^k\bar{y}_k$ is relevant throughout; it is a mere change of notation to then call this function $\bar{y}_k$, instead of the original function.

The dual problem in this context may be stated as:

LD

$$\text{Maximize } E\left\{\sum_{k=1}^{N} b_k \cdot y_k(\xi^k)\right\} \text{ over all summable}$$

$$y_k(\xi^k) \geqq 0,\ k = 1, \cdots, N, \text{ satisfying}$$

$$\sum_{k=j}^{N} (E^j y_k)(\xi^k) A_{kj} = c_j \quad \text{a.s.} \quad \text{for } j = 1, \cdots, N. \tag{3.28}$$

Note that the function $y = (y_1, \cdots, y_N)$ may be called a *nonanticipative* element of L^1_m. However, LD does not fit the same mold as LP, since in determining the component y_k for stage k we need consider the conditional expectations of the *future* components y_j, $k < j \leqq N$. Looked at another way, LD involves certain special *chance constraints*, in contrast to LP, because if the expected values of the multipliers y_j associated with future stages are treated as variables to be determined at stage k, then the decision which is taken poses a subsequent constraint on expectations that y_j must live up to.

THEOREM 8. *The dual problem* LD *has optimal solutions, and they are precisely the components* $\bar{y} = (\bar{y}_1, \cdots, \bar{y}_N)$ *of the pairs* $(\bar{u}, \bar{y})$ *satisfying the Kuhn–Tucker conditions in Theorem* 7. *In particular,*

$$\min \text{LP} = \max \text{LD}.$$

Proof. This follows as a specialization of Theorem 5 via a slight change in notation as in the proof of Theorem 7.

Problem LD resembles the dual obtained by Eisner and Olsen [12] for linear recourse models formulated in L^p-spaces, $1 < p < \infty$. The approach developed here, however, yields a min = max duality theorem with corresponding Kuhn–Tucker conditions, whereas [12] only allows for min = sup duality results.

4. A discrete time stochastic control problem. The purpose of this section is to illustrate, by an example, the relations between the recourse model and certain types of stochastic control problems in discrete time. The optimality conditions developed here can then be used to characterize optimal solutions to these stochastic control problems. The goal is not to give a description of the most general stochastic control problem that can be handled in the framework of the recourse model; it is easy to see how the problem described below can be generalized in many directions and still fit our pattern.

While there are a number of substantial contributions to the theory of necessary and sufficient conditions for stochastic control problems in discrete time, e.g. [13] and [14], there does not seem to be a treatment that allows for the inclusion of state-space constraints when seeking *pointwise optimality conditions*. Several papers do deal with state-space constraints in the continuous case; see [15], [16], [17] and [18]. The difference between the present approach and the one taken by Kushner [15], Haussmann [16] and Ichikawa [17] is that they seek an "expected maximum principle," in which case the multipliers associated with the state-space constraints (at a finite number of time periods) turn out to be elements of R. It is when seeking pointwise optimality conditions that the difficulties do arise, as illustrated in [18] where Bismut must rely on an $(L^\infty)^*$-multiplier rather than L^1-multiplier. Even for continuous-time deterministic problems with state-space constraints these exotic multipliers cannot always be avoided [19].

Let $(\boldsymbol{\xi}_k,\ k = 1, \cdots, N)$ denote a vector-valued (discrete time) stochastic process; for $k = 1, \cdots, N$, the realizations of $\boldsymbol{\xi}_k$ are elements of $R^{\nu'}$ denoted by ξ_k. The state of the system at time k is denoted by x_k, also an element of $R^{\nu'}$. The dynamics are given by the relations

$$x_1 = \xi_1 \tag{4.1}$$

and for $k = 1, \cdots, N-1$,

$$x_{k+1} = Ax_k + Bu_k + \xi_{k+1}, \tag{4.2}$$

where A is a $(\nu' \times \nu')$-matrix, B is a $(\nu' \times n')$-matrix, and $u_k \in R^{n'}$ is the recourse (or control) selected at time k. To be consistent with our earlier notation, we set $\nu = N\nu'$ and $n = Nn'$. The recourse is selected on the basis of *complete information* and *total recall*, by which we mean that the recourse decision u_k is selected in complete knowledge of the past history of the system, i.e. up to and including x_k, the state of the system at time k. (Note that a number of problems with incomplete observation and partial recall can actually be cast as problems with complete information and total recall, see for example [20], [21].) In this set-up, it is equivalent to assert that the decision maker observes ξ_k and recalls past observations and past decisions, since clearly from (4.2) it follows that knowledge of earlier states and decisions, and observation of x_{k+1} uniquely determines ξ_{k+1}, the "noise" of the system.

Moreover, the particular form of the dynamics of this system (4.2) allows us to short-circuit the state component in the description of the model. Indeed, combining (4.1) and (4.2) we obtain

$$x_{k+1} = \sum_{q=0}^{k} A^q \xi_{k+1-q} + \sum_{q=0}^{k-1} B u_{k-q}, \tag{4.3}$$

i.e. the state of the system is a linear function of the past recourse decisions and realizations.

For performance criterion we take a real-valued functional φ_0 defined on $R^\nu \times R^n \times R^\nu$ such that for all $\xi \in \Xi$ the function $(u, x) \mapsto \varphi_0(\xi, u, x)$ is convex, and for all $(u, x) \in R^n \times R^\nu$ the function $\xi \mapsto \varphi_0(\xi, u, x)$ is (Borel) measurable. The problem in rough form is to minimize $E\{\varphi_0(\xi, u(\xi), x(\xi))\}$ subject to (4.3) and the further constraints that

$$u(\xi) \in U(\xi) \quad \text{a.s.} \tag{4.4}$$

and for $i = 1, \cdots, m$

$$\varphi_i(\xi, u(\xi), x(\xi)) \leqq 0 \quad \text{a.s.} \tag{4.5}$$

The multifunction U is assumed to have closed, convex values with nonempty interior. For $i = 1, \cdots, m$, the functions φ_i on $R^\nu \times R^n \times R^\nu$, are required to satisfy the same assumptions as φ_0.

Relation (4.3) allows us to formulate this stochastic control problem as a problem of the type P. Let us write (4.3) (regarded as including (4.1)) in the form

$$x = S\xi + Tu$$

and define

$$f_i(\xi, u) = \varphi_i(\xi, u, S\xi + Tu) \quad \text{for all } u \in U(\xi) \quad i = 0, 1, \cdots, m.$$

Suppose in fact that $\varphi_i(\xi, u, S\xi + Tu)$ is (for each fixed u) summable in ξ when $i = 0$ and essentially bounded in ξ when $i = 1, \cdots, m$. It can be verified that $f_0, \cdots, f_m$ and U satisfy all the conditions imposed on them in § 1, and the corresponding optimal recourse problem P represents the present situation. If in addition P is strictly feasible and the abstract constraint multifunction D corresponding to $f_1, \cdots, f_m$ and U is uniformly bounded and nonanticipative, all our general assumptions are satisfied and the above results can be applied. In this way we obtain necessary and sufficient conditions for optimality from the basic Kuhn–Tucker conditions (a), (b), (c).

Many of the regularity conditions in question are "standard" for control problems. For example, it is common to assume uniform boundedness of the set $\{U(\xi), \xi \in \Xi\}$, and this ensures the uniform boundedness of the multifunction D. As far as nonanticipative feasibility is concerned, we have already explained its relation to the notion of relatively complete recourse that has played an important role in the literature devoted to stochastic programming [4]. This concept has also recently surfaced in stochastic control theory [22], [23]. For a system without state constraints, Striebel [22] introduced the concept of *optimality from time t onward*, requiring essentially that for each control satisfying (4.4) and each time t—whatever be the resulting state—there is a control which is optimal from time t onward. Striebel and Rishel [23] use this condition in their study of optimality criteria for continuous time stochastic control problems. Their motivation for introducing "optimality from time t onward" is quite different from ours but seems to be required by technical considerations that are akin to those that lead us to essentially complete recourse. In particular, Rishel shows that this condition allows him to obtain an explicit form for the generator applied to the value function.

Finally, we note that certain classes of stochastic control problems yield separable recourse problems. This is certainly the case if

(i) $U(\xi) = \mathop{\times}\limits_{k=1}^{N} U_k(\xi)$ where U_k is $\mathcal{F}_k$-measurable,

(ii) for $i = 0, 1, \cdots, m$, $\varphi_i(\xi, u, x) = \sum_{k=1}^{N} \varphi_{ik}(\xi, u_k) + x \cdot r_i$,

where the functions $\xi \mapsto \varphi_{ik}(\xi, u_k)$ are $\mathcal{F}_k$-measurable and $r_i \in \mathbb{R}^\nu$. In this case we can rely on the sharper results of § 3A, if not in fact 3B, in deriving optimality conditions.

REFERENCES

[1] R. T. ROCKAFELLAR, *Integrals which are convex functionals*, Pacific J. Math., 24 (1968), pp. 525–539.

[2] ———, *Measurable dependence of convex sets and functions on parameters*, J. Math. Anal. Appl., 28 (1969), pp. 4–25.

[3] ———, *Conjugate Duality and Optimization*, Monograph Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1974.

[4] R. T. ROCKAFELLAR AND R. WETS, *Stochastic convex programming: relatively complete recourse and induced feasibility*, this Journal, 14 (1976), pp. 574–589.

[5] ———, *Nonanticipativity and L^1-martingales in stochastic optimization problems*, Mathematical programming Studies, 6 (1976); also in Stochastic Systems: Modeling, Identification and Optimization, R. Wets, ed., North-Holland, 1976, pp. 170–187.

[6] ———, *Stochastic convex programming: Kuhn–Tucker conditions*, J. Math. Economics, 2 (1975), pp. 349–370.

[7] R. T. ROCKAFELLAR, *Integrals which are convex functionals, II*, Pacific J. Math., 39 (1971), pp. 439–469.

[8] R. T. ROCKAFELLAR AND R. WETS, *Continuous versus measurable recourse in N-stage stochastic programming*, J. Math. Anal. Appl., 48 (1974), pp. 836–859.

[9] P. Olsen, *Discretization of multistage stochastic programming problems*, Mathematical Programming Studies, 6 (1976); also in Stochastic Systems: Modeling, Identification and Optimization, R. Wets, ed., North-Holland, Amsterdam, 1976, pp. 111–124.

[10] R. Grinold, *Symmetric duality for continuous linear programs*, SIAM J. Appl. Math., 18 (1970), pp. 84–96.

[11] M. Isofescu and R. Theodorescu, *Linear programming under uncertainty*, Colloquium on Applications of Mathematics to Economics, A. Prekopa, ed., Publishing House of the Hungarian Academy of Sciences, Budapest, 1965, pp. 133–140.

[12] M. Eisner and P. Olsen, *Duality for stochastic programming interpreted as L.P. in L_p-space*, SIAM J. Appl. Math., 28 (1974), pp. 779–792.

[13] C. Striebel, *Optimal control of discrete time stochastic systems*, Tech. Report, Univ. Minnesota, Minneapolis, 1974.

[14] K. Hinderer, *Foundations of Non-stationary Dynamic Programming with Discrete Time Parameter*, Lecture Notes in Operations Research and Mathematical Systems, 33, Springer-Verlag, Berlin, 1970.

[15] H. Kushner, *Necessary conditions for continuous parameter stochastic optimization problems*, this Journal, 10 (1972), pp. 550–565.

[16] U. Haussmann, *General necessary conditions for optimal control of stochastic systems*, Mathematical Programming Studies, 6, 1976; also in Stochastic Systems: Modeling, Identification and Optimization, R. Wets, ed., North-Holland, Amsterdam, 1976, pp. 30–48.

[17] A. Ichikawa, *Notes on a maximum principle of Haussmann*, Tech. Report, Univ. British Columbia, Vancouver, 1975.

[18] J.-M. Bismut, *An example of optimal stochastic control with constraints*, this Journal, 12 (1974), pp. 401–418.

[19] R. T. Rockafellar, *State constraints in convex problems of Bolza*, this Journal, 10 (1972), pp. 691–715.

[20] Y. C. Ho and K. C. Chu, *Information structure in dynamic multi-person control problems*, Automatica, 10 (1974), pp. 341–351.

[21] J.-M. Bismut, *An introductory approach to duality in optimal stochastic control*, SIAM Rev., to appear.

[22] C. Striebel, *Martingale conditions for optimal stochastic control of continuous time stochastic systems*, Tech. Report, Univ. of Minnesota, Minneapolis, 1974.

[23] R. Rishel, *Controls optimal from time t onward and dynamic programming for systems of controlled jump processes*, Mathematical Programming Studies, 6, 1976; also in Stochastic Systems: Modeling, Identification and Optimization, R. Wets, ed., North-Holland, Amsterdam, 1976, pp. 125–144.