MATHEMATICS OF OPERATIONS RESEARCH
Vol. 16, No. 1, February 1991
Printed in U.S.A.

SCENARIOS AND POLICY AGGREGATION IN OPTIMIZATION UNDER UNCERTAINTY*

R. T. ROCKAFELLAR AND ROGER J.-B. WETS†

A common approach in coping with multiperiod optimization problems under uncertainty where statistical information is not really enough to support a stochastic programming model, has been to set up and analyze a number of scenarios. The aim then is to identify trends and essential features on which a robust decision policy can be based. This paper develops for the first time a rigorous algorithmic procedure for determining such a policy in response to any weighting of the scenarios. The scenarios are bundled at various levels to reflect the availability of information, and iterative adjustments are made to the decision policy to adapt to this structure and remove the dependence on hindsight.

1. Introduction. Most systems that need to be controlled or analyzed involve some level of uncertainty about the value to be assigned to various parameters or the actual layout of some of the system's components. Not much is lost by simply assigning "reasonable" values to the unknown elements, as long as their role is relatively insignificant. But in many situations the model builder cannot do this without running the risk of invalidating all the implications that are supposed to be drawn from the analysis.

When a probabilistic description of the unknown elements is at hand, either because a substantial statistical base is available or because a probabilistic law can be derived from conceptual considerations (measurement error, life and death processes, etc.), one is naturally led to consider stochastic models. When only partial information, or no information at all, is available, however, there is understandably a reluctance to rely on such models. In presuming that probability distributions exist they seem inherently misdirected. Besides, the problems of stochastic optimization that they lead to can be notoriously hard to solve.

A common approach in practice is to rely on *scenario analysis*. The uncertainty about parameters or components of the system is modeled by a small number of versions of subproblems derived from an underlying optimization problem. These correspond to different "scenarios" a word that is used to suggest some kind of limited representation of information on the uncertain elements or how such information may evolve. The idea is that by studying the different subproblems and their optimal solutions one may be able to discover similarities and trends and eventually come up with a "well hedged" solution to the underlying problem, something which can be expected to perform rather well under all scenarios, relative to some weighting of scenarios. As examples, see [1] and [2].

*Received October 29, 1987; revised August 23, 1989.
AMS 1980 subject classification. Primary: 90C42.
IAOR 1973 subject classification. Main: Programming: probabilistic.
OR/MS Index 1978 subject classification. Primary: 663 Programming/stochastic.
Key words. Optimization under certainty, scenario analysis, progressive hedging, information prices, multistage decision problems, stochastic programming, decomposition methods, splitting methods, proximal point algorithm, augmented Lagrangians.
†The work of both authors was supported in part by grants from the Air Force Office of Scientific Research and the National Science Foundation.

0364-765X/91/1601/0119/$01.25

To give this a mathematical formulation, let us write the scenario subproblems as

$$(\mathscr{P}_s) \qquad \text{minimize } f_s(x) \text{ over all } x \in C_s \subset \mathbf{R}^n$$

where the index s ranges over a relatively modest, finite set S: the set of scenarios. It is not our intention to address in this paper the question of how the scenario subproblems might be chosen or constructed. We take them for granted and suppose that we know how to solve them individually for each choice of a scenario s. The question we do raise is that of how to work with the different s-dependent solution vectors so obtained and consolidate them into an overall decision or decision policy. The essential difficulty obviously lies in the fact that actions in the real world must be taken without the hindsight that goes into solving the problems $(\mathscr{P}_s)$. In multistage models the actions could, however, respond in time to increasing degrees of information that becomes available about the particular scenario being followed. The expression of such information structure must be an important part of the formulation.

Let us suppose we are dealing with time periods $t = 1, \ldots, T$ and write

$$(1.1) \qquad x = (x_1, \ldots, x_T) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_T},$$

where $n_1 + \cdots + n_T = n$. The component x_t represents the decision that must be made at time t. More generally let X denote a function or mapping that assigns to each $s \in S$ a vector

$$(1.2) \qquad X(s) = (X_1(s), \ldots, X_T(s)),$$

where $X_t(s)$ denotes the decision to be made at time t if the scenario happens to be s. It is such a mapping—let us call it a *policy*—that we are really looking for, but it has to satisfy the crucial constraint that *if two different scenarios s and s' are indistinguishable at time t on the basis of information available about them at time t, then $X_t(s) = X_t(s')$*. A policy, if it is to make sense, cannot require different courses of action at time t relative to scenarios s and s' if there is no way to tell at time t which of the two scenarios one happens to be following.

A good way of modeling this constraint is to introduce an information structure by *scenario bundling*, i.e., by partitioning the scenario set S at each time t into finitely many disjoint subsets, which can be termed *scenario bundles*. The scenarios in any one bundle are regarded as observationally indistinguishable at time t. Denoting the collection of all scenario bundles at time t by $\mathscr{A}_t$, we impose the requirement that *$X_t(s)$ must be constant relative to $s \in A$ for each $A \in \mathscr{A}_t$*. Thus from the space of all mappings $X\colon S \to \mathbf{R}^n$ with components $X_t\colon S \to \mathbb{R}^{n_t}$ as in (1.2), a space we denote by $\mathscr{E}$, we single out the subspace

$$(1.3) \qquad \mathscr{N} := \{X \in \mathscr{E} \mid X_t \text{ is constant on each bundle } A \in \mathscr{A}_t \text{ for } t = 1, \ldots, T\}$$

as specifying the policies that meet our fundamental constraint of not being based on hindsight. The policies X belonging to $\mathscr{N}$ will be called *implementable* policies. We make a distinction here between implementable policies and *admissible* policies, which belong to the set

$$(1.4) \qquad \mathscr{C} := \{X \in \mathscr{E} \mid X_t(s) \in C_s \text{ for all } s \in S\}.$$

For most purposes it would be reasonable to suppose that the partition $\mathscr{A}_{t+1}$ is a refinement of the partition $\mathscr{A}_t$ in the sense that each bundle $A \in \mathscr{A}_t$ is a union of

bundles in $\mathscr{A}_{t+1}$. One could then think equivalently of a branching of possibilities at a tree, with each time t and therefore of a representation of the information structure in terms of a tree, with each scenario corresponding to a path from the root of the tree out to a leaf. This would be consistent with the idea that information increases in time and is never lost. None of what we say in this paper actually depends on such an assumption, though. Information about the scenario being followed could be allowed to vary quite generally.

The central question of scenario analysis can now be stated. Given the collection of scenario subproblems $(\mathscr{P}_s)$ and a license perhaps to modify them (perturb their objectives) so as to assist in adapting to the information structure, we have the means of generating various policies $X \in \mathscr{C}$ that may be called *contingent* policies: $X(s)$ is obtained by solving a possibly perturbed version of the scenario subproblem $(\mathscr{P}_s)$ for each $s \in S$. How can we use these means to determine an implementable policy $X^* \in \mathscr{N}$ that in some sense is good for the *underlying* problem of optimization under uncertainty?

Note that a contingent policy is at least always *admissible*: $X \in \mathscr{C}$. But this condition is not built into our use of the term "implementable". Obviously a policy that is both admissible and implementable is what we really want—this is what we shall mean by a *feasible* policy. But *implementability is a logically inescapable requirement*, whereas admissibility might be waived by the modeler in some situations that only risk the violation of $X(s) \in C_s$ for a few extreme or unlikely scenarios, or entail mild transgressions of certain nonkey constraints in more ordinary scenarios.

The simplest case of a one-stage model ($T = 1$) helps to illustrate these ideas. In this case we only know the present. We know nothing that would pin down a particular scenario or subclass of scenarios, but are forced to make a decision "here and now". A policy X, with just one time component, is implementable if for all $s \in S$ one has $X(s) = x$ for some (fixed) vector x. In other words, the space $\mathscr{N}$ consists of just the constant mappings from S to $\mathbb{R}^n$, in contrast to the space $\mathscr{C}$, which consists of all possible mappings from S to $\mathbb{R}^n$. (The partition $\mathscr{A}_1$ in this example is the "trivial partition" consisting of the set S by itself, no scenario being regarded as distinguishable from any other at the time the single decision has to be taken. All of S is a single bundle.) In this setting, the question is one of proceeding from a mapping X that is not constant to a mapping that is constant by some method making use of the insights gained by solving the individual scenario subproblems in various forms.

An attractive way of passing from a general policy X to a policy that is implementable is to assign to each scenario $s \in S$ weight p_s that reflects its relative importance in the uncertain environment, with

$$p_s > 0 \quad \text{for all } s \in S, \quad \text{and} \quad \sum_{s \in S} p_s = 1. \tag{1.5}$$

These weights are used in blending the response $X(s)$ of X so as to meet the requirement of not allowing the decision at time t to distinguish among the scenarios in a bundle $A \in \mathscr{A}_t$. Specifically one calculates for every time t and for every $A \in \mathscr{A}_t$ the vector

$$X_t(A) := \sum_{s \in A} p_s X_t(s) \Big/ \sum_{s \in A} p_s, \tag{1.6}$$

which represents a "weighted average" of all the responses $X_t(s)$ for scenarios in the

bundle A. Then one defines a new policy by taking

$$\hat{X}_t(s) = X_1(A) \quad \text{for all } s \in A. \tag{1.7}$$

Clearly $\hat{X}$ is implementable: $\hat{X} \in \mathcal{N}$. (In the one-stage model, $\hat{X}$ would simply be the constant mapping whose value is the vector $\sum_{s \in S} p_s X(s)$.) The transformation

$$J: X \mapsto \hat{X} \text{ defined by (1.6)–(1.7)} \tag{1.8}$$

is obviously linear and satisfies $J^2 = J$. It is a projection from $\mathcal{E}$ onto $\mathcal{N}$ which depends only on the weights p_s. We call it the *aggregation operator* relative to the given information structure and weights. It aggregates the possibly different references that a policy might make for the scenarios in any bundle into a single compromise response to that bundle.

If we were to start from a contingent policy X^0 in which $X^0(s)$ is for each s an optimal solution to the unmodified scenario subproblem $(\mathcal{P}_s)$, which is the typical beginning for all scenario analysis, the corresponding implementable policy $\hat{X}^0 = JX^0$ might be contemplated as a kind of solution to the underlying problem. There is no guarantee, however, that $\hat{X}^0$ will inherit from X^0 the property of admissibility. Even if $\hat{X}^0$ is admissible as well as implementable, therefore feasible, the sense in which it might be regarded as "optimal" needs to be clarified. As a matter of fact, $\hat{X}^0$ is an optimal solution to a certain "projected" problem, which will be described presently, but this is not at all the problem that one is interested in.

If instead of introducing the weights p_s in an *a posteriori* manner we were to do so at the outset, we would be led in our search for a well-hedged decision policy to the functional

$$F(X) := \sum_{s \in S} p_s f_s(X(s)) \tag{1.9}$$

and the problem

$$\text{minimize } F(X) \text{ over all } X \in \mathcal{C} \cap \mathcal{N}. \tag{1.10}$$

An optimal solution X^* to this problem would indeed be admissible and implementable. Among all admissible, implementable policies it would do the best job, in a certain specific sense, of responding to the relative importance of the scenarios as assessed through the weights p_s. It would provide a sound method of hedging against the unknowns.

The weights need not be regarded as "hard data" for this interpretation to be valid. The road is always open at another level to play with the values of the weights and see how sensitively the problem is affected by them, although we do not take that issue up here.

The trouble is that problem (1.10) may be much larger and therefore much harder to solve than the individual scenario subproblems $(\mathcal{P}_s)$, so that it cannot be tackled directly. There is little prospect, either, that the desired policy X^* is approximated at all closely by the policy $\hat{X}^0$ already described. This is seen from the elementary fact that $\hat{X}^0$ actually solves

$$\text{minimize } \hat{F}(V) \text{ over all } Y \in \hat{\mathcal{C}}, \tag{1.11}$$

where $\hat{\mathcal{C}} = \{Y \in \mathcal{N} \mid \exists X \in \mathcal{C} \text{ with } JX = Y\}$ and $\hat{F}(Y) = \min\{F(X) \mid X \in \mathcal{C}, JX = Y\}$. The projected problem (1.11) is utterly different from (1.10).

Nonetheless there turns out to be a relationship that can be exploited to trace a path from $\hat{X}^0$ to X^* by solving a sequence of projected problems in which the scenario subproblems are not the original ones but modified by the incorporation of tentative "information prices" and penalties. At iteration ν we take a contingent policy X^ν obtained by solving modified scenario subproblems ($\mathscr{P}_s^\nu$) and aggregate it into an implementable policy $\hat{X}^\nu$ whose robustness in the face of all eventualities is increasingly demanded. An advantage of this approach is that even if we do not pursue the search until $\hat{X}^\nu$ converges to X^*, we always have at hand a solution estimate that is better than just $\hat{X}^0$ or any other policy that could reliably be gleaned from scenario analysis as practiced until now. The word "better" is given specific meaning by our convergence theory. The very process of blending decision components iteratively in the manner we suggest is likely moreover to identify fairly early the trends and activities that will lead to the final solution.

The general principle that allows us to proceed in this manner in generating improving sequences of policies is what we call the *principle of progressive hedging* in optimization under uncertainty. It enables us by simple means to insist more and more on having our subproblems reflect the ultimate requirement that a policy, to be implementable, cannot distinguish between scenarios that at a particular time are regarded as indistinguishable from each other on the basis of information so far available. The realization of the principle that we give here is based mathematically on the theory of the proximal point algorithm in nonlinear programming, as developed in Rockafellar [3] and applied to problem decomposition and "splitting" in Spingarn [4]. It does not depend on convexity in its statement, although convexity provides a big boost in its analysis.

A notable by-product of our hedging algorithm is the generation of *information prices* relative to the chosen weights p_s. Potentially these might be used in some larger scheme for adjusting the weights or judiciously supplying more detail to the set of scenarios. In the limit the information prices solve a dual problem, which however is likely to have dimension at least as high as that of the primal problem. Because of this high dimensionality, approaches like Dantzig-Wolfe generalized programming (which in effect applies a cutting-plane method to the dual) are not suitable. On the other hand, augmented Lagrangian methods in their usual formulation are unable to maintain the crucial decomposability. The approach we take is not blocked by the dual dimensionality and yet does allow the separate scenario subproblems in each iteration to be solved in parallel, if desired.

2. General framework. There is no harm in interpreting the weights p_s mathematically as probabilities. They may indeed represent "subjective probabilities," but the reader should not conclude from the probabilistic language which follows that we necessarily regard them so. In passing to a probability framework we merely take advantage of the fact that it provides a convenient scheme for organizing ideas that mathematically fall into the same patterns as are found in dealing with probability. Much the same could be said about the use of geometric language in a nongeometric situation.

From now on, sums with the weights p_s will be written as *expectations* in the traditional notation:

$$\sum_{s\in S} p_s X(s) = E\{X(s)\},$$

for instance. Then in (1.6) we have

$$X_t(A) = E\{X(s)|A\}, \tag{2.1}$$

the *conditional expectation* of $X_t(s)$ given that $s \in A$, and we can interpret the projection J: $X \mapsto \hat{X}$ quite simply as the *conditional expectation* operator relative to the given information structure and values p_s.

The information structure can itself be furnished with a traditional interpretation in terms of fields of sets: $\mathscr{F}_t$ is for each t the collection of all subsets expressible as unions of the (disjoint) sets in $\mathscr{A}_t$. Then $\hat{X}_t$ is the conditional expectation of X_t relative to $\mathscr{F}_t$. Such terminology, bringing to mind all the subleties of measure theory, is not in any way needed, however, in the present context where S is just a finite set. It could just get in the way of a "user-friendly" explanation of ideas that are really quite elementary, so for the purposes at hand we avoid it.

An inner product on the vector space $\mathscr{E}$ of all mappings from S to $\mathbb{R}^n$ is defined by

$$\langle X, Y \rangle := E\{X(s) \cdot Y(s)\} = \sum_{s \in S} p_s X(s) \cdot Y(s). \tag{2.2}$$

We think of $\mathscr{E}$ as a Euclidean space in this sense, the norm being

$$\|X\| = \left[E\{|X(s)|^2\}\right]^{1/2}, \tag{2.3}$$

where $|\cdot|$ is the ordinary Euclidean norm on $\mathbb{R}^n$. The aggregation operator J is then actually the *orthogonal projection* on the subspace $\mathscr{N}$, as is well known. The operator

$$K = I - J \qquad (KX = X - \hat{X}) \tag{2.4}$$

is the orthogonal projection on the subspace of $\mathscr{E}$ complementary to $\mathscr{N}$, which we denote by $\mathscr{M}$:

$$\begin{aligned} \mathscr{M} = \mathscr{N}^{\perp} &= \{W \in \mathscr{E} \mid JW = 0\} \\ &= \{W \in \mathscr{E} \mid E\{W_t(s) \mid A\} = 0 \text{ for all } A \in \mathscr{A}_t, t = 1, \ldots, T\}. \end{aligned} \tag{2.5}$$

Clearly

$$\mathscr{N} = \{X \in \mathscr{E} \mid KX = 0\} = \{X \in \mathscr{E} \mid X = \hat{X}\}. \tag{2.6}$$

Thus a policy X is implementable if and only if it satisfies the linear constraint equation $KX = 0$.

The functional F in (1.9) can be written now as

$$F(X) = E\{f_s(X(s))\}. \tag{2.7}$$

The problem we wish to solve then has the formulation

$$(\mathscr{P}) \qquad \text{minimize } F(X) \text{ subject to } X \in \mathscr{C},\ KX = 0.$$

An optimal solution X^* to this problem is what we take to be the best response we can offer to the uncertain environment, relative to the given weighting of the scenarios. The challenge for us, in adopting this point of view as a practical expedient, is that of demonstrating how such an X^* can be determined without going beyond the tools that are available.

We see our capabilities as extending in two directions. First we can readily calculate for any X the corresponding $\hat{X} = JX$ and therefore also $X - \hat{X} = KX$. The projections J and K are thus computable and appropriate to use in the context of an

algorithm. Second, we can solve, at least approximately to any desired degree, the scenario subproblems ($\mathscr{P}_s$) and a certain class of modified versions of these subproblems. The specific form of modified scenario subproblem that we work with in this paper is

$$(\hat{\mathscr{P}}_s(\hat{x}, w, r)) \quad \text{minimize } f_s(x) + x \cdot w + \tfrac{1}{2}r|x - \hat{x}|^2 \text{ over all } x \in C_s.$$

The vector $\hat{x}$ will stand for an estimate of x from which we do not want to stray too far; $w \in \mathbb{R}^n$ will be a price vector and $r > 0$ a penalty parameter.

Motivation comes in part from Lagrangian representations for problem ($\mathscr{P}$). The ordinary Lagrangian for this problem could be defined as the expression

$$F(X) + \langle KX, Y \rangle \quad \text{for } X \in \mathscr{C}, Y \in \mathscr{E},$$

with multiplier Y, but since K is an orthogonal projection one has

$$\langle KX, Y \rangle = \langle X, KY \rangle = \langle KX, KY \rangle. \tag{2.8}$$

Only the component $W = KY \in \mathscr{M}$ can really matter. We therefore find it convenient to define

$$L(X, W) = F(X) + \langle X, W \rangle \quad \text{for } X \in \mathscr{C}, W \in \mathscr{M}, \tag{2.9}$$

as the Lagrangian. The multiplier element W will be called an *information price system* because of its role relative to the implementability constraint $KX = 0$. More will be said about this later.

The ordinary Lagrangian (2.9), important as it can be for instance in stating optimality conditions, is limited in its numerical usefulness. More powerful in many ways if one can work with it, and not limited to the problems where convexity is present, is the corresponding *augmented* Lagrangian

$$\begin{aligned} L_r(X, W) &= F(X) + \langle X, W \rangle + \tfrac{1}{2}r\|KX\| \\ &= F(X) + \langle X, W \rangle + \tfrac{1}{2}r\|X - \hat{X}\|^2 \end{aligned} \tag{2.10}$$

$$\text{for } x \in \mathscr{C}, W \in \mathscr{M}, r > 0.$$

There is no room here for a general discussion of augmented Lagrangians, except to say that they combine features of multipliers and penalties. Through a good choice of $W \in \mathscr{M}$ and $r > 0$ one can expect that the subproblem

$$\text{minimize } L_r(X, W) \text{ over } X \in \mathscr{C} \tag{2.11}$$

can be used as a close representation of ($\mathscr{P}$), in the sense that its nearly optimal solutions will be good approximates to an optimal solution X^* of ($\mathscr{P}$). This is true without any assumption of convexity and does not necessarily entail r getting too large for comfort; much of the work can be done by W. Even in the convex case the augmented Lagrangian can be advantageous by providing greater stability to solution methods. We refer the reader to Bertsekas [5] and Rockafellar [6] for more on this topic.

Unfortunately, the augmented Lagrangian (2.10) cannot serve directly in our scheme. To use it we would have to be able to solve subproblems of the form (2.11), which do *not* meet our prescription. The difficulty lies in the fact that the term

$\|KX\|^2$ is not decomposable into separate terms for each scenario. Nonetheless we are able to take an approach which looks quite similar and does achieve the required decomposition.

The approach can be described quite broadly in terms of the following algorithmic scheme. We shall subsequently make it more specific, in order to have results on convergence. A fixed parameter value $r > 0$ is considered throughout this paper for simplicity. In practice one might wish to make adjustments in the value of r. This is an issue for which the theoretical backing is incomplete, although some elucidation will be provided in Proposition 5.3 and the comment that follows it.

PROGRESSIVE HEDGING ALGORITHM. *In iteration ν (where $\nu = 0, 1, \ldots$) one has an admissible but not necessarily implementable policy $X^\nu \in \mathscr{C}$ and a price system $W^\nu \in \mathscr{M}$. (Initially one can take X^0 to be the policy obtained by letting $X^0(s)$ be for each scenario $s \in S$ an optimal solution to the given scenario subproblem $(\mathscr{P}_s)$. One can take $W^0 = 0$.)*

Step 1. *Calculate the policy $\hat{X}^\nu = JX^\nu$, which is implementable but not necessarily admissible. (If ever one wishes to stop, this policy $\hat{X}^\nu$ is to be offered as the best substitute yet available for a solution to $\mathscr{P}$.)*

Step 2. *Calculate as $X^{\nu+1}$ an (approximately) optimal solution to the subproblem*

$$(\mathscr{P}^\nu) \qquad \text{minimize } F(X) + \langle X, W^\nu \rangle + \tfrac{1}{2}r\|X - \hat{X}^\nu\|^2 \quad \text{over all } X \in \mathscr{C}.$$

This decomposes into solving (approximately) for each scenario $s \in \mathscr{S}$ the subproblem

$$(\mathscr{P}_s^\nu) \qquad \text{minimize } f_s(x) + x \cdot W^\nu(s) + \tfrac{1}{2}r|x - \hat{X}^\nu(s)|^2 \quad \text{over all } x \in C_s$$

in order to get $X^{\nu+1}(s)$. The policy $X^{\nu+1}$ will again be admissible but not necessarily implementable.

Step 3. *Update from W^ν to $W^{\nu+1}$ by the rule $W^{\nu+1} = W^\nu + rKX^{\nu+1}$. The price system $W^{\nu+1}$ will again be in $\mathscr{M}$. Return to Step* 1 *with ν replaced by $\nu + 1$.*

Left open in this statement is the sense in which the modified scenario subproblems in Step 2 need only be solved "approximately". These subproblems fit the general form mentioned earlier:

$$(\mathscr{P}_s^\nu) = \big(\hat{\mathscr{P}}_s(\hat{X}^\nu(s), W^\nu(s), r)\big).$$

In many applications they will turn out to be quadratic programming problems of reasonable dimension. Then one could well imagine solving them "exactly." The question of a stopping criterion for approximate solution is therefore not a *sine qua non*. A substantial answer will nevertheless be presented in §5.

The updating rule for the price system in Step 3 could in principle be replaced by something else without destroying the truly critical property of decomposability in Step 2. This rule is strongly motivated, though, by augmented Lagrangian theory (cf. [6]). It is essential not merely to the proofs of the convergence theorems but the very nature of the reformulation of the algorithm on which these proofs rely.

An obvious strength of the proposed procedure is that it involves at every iteration an *admissible* policy X^ν and an *implementable* policy $\hat{X}^\nu$. The distance expression

$$\|X^\nu - \hat{X}^\nu\| = \|KX^\nu\|^2 \tag{2.12}$$

can readily be computed and taken as a measure of how far one is from satisfying all the constraints. Note that (2.12) is a kind of *conditional variance* relative to the

weights p_s. In our convergence theorems for the convex case, at least, this quantity will tend to 0. At the same time, the price systems W^ν will tend to an optimal solution to the Lagrangian dual of problem $(\mathscr{P})$.

The progressive hedging algorithm can be viewed as a particular instance of a method of Spingarn for minimizing a function over a subspace [4] (Algorithm 2), which in turn is a specialization via Spingarn [7] of the proximal point algorithm in Rockafellar [3] and derives its convergence theory from the latter. It can be placed more broadly in the context of splitting algorithms like those of Lions and Mercier [8] (covering the convex optimization methods of Gabay and Mercier [9] and Fortin and Glowinski [10], among others), which likewise are a specialization of the proximal point algorithm in [3], as demonstrated by Eckstein [11]. Recently in this vein also is the decomposition method of Han [12]. In our main convergence theorems in §5 we directly work with the theory in [3], however, in part for the purpose of bringing out more clearly the connection with augmented Lagrangian methods by portraying the hedging algorithm as a proximal step taken on a certain saddle function relative to the "variance" norm (2.3).

Some rudimentary analysis of the nonconvex case is provided in §6. We have much less to say about it at the present stage of development and try only to indicate a potential in this direction.

Our immediate task, in §3 and §4, is to lay the foundations for the rigorous handling of the scenario subproblems and the characterization of the dual elements W.

3. Basic assumption and properties. It will be assumed throughout the rest of this paper that for each $s \in S$ the feasible set C_s in the scenario subproblem $(\mathscr{P}_s)$ is nonempty and closed, and the objective function f_s is locally Lipschitz continuous on $\mathbb{R}^n$ with all level sets of the form

$$\{x \in C_s | f_s(x) \leqslant \alpha\}, \qquad \alpha \in \mathbb{R}, \tag{3.1}$$

bounded. This last condition is trivially satisfied, of course, if C_s itself is bounded. The closedness of C_s presumably comes from the constraint structure used to define C_s, but such explicit structure will not play any role here. The local Lipschitz continuity of f_s is present if f_s is smooth (i.e., of class $\mathscr{C}^1$ on $\mathbb{R}^n$) or, on the other hand, if f_s is convex.

We shall speak of the *convex case* of our problem $(\mathscr{P})$ when for every $s \in S$ the function f_s is convex and the set C_s is convex. The *linear-quadratic* case will refer to the more special situation where f_s is quadratic (convex) and C_s is polyhedral (convex). We regard linear and affine functions as included under the heading of "quadratic."

We proceed with some of the elementary consequences of these conditions. The first topic is their effect on the given scenario subproblems $(\mathscr{P}_s)$, whose solution is called for at the outset of our proposed algorithm.

PROPOSITION 3.1. *Each of the scenario subproblems $(\mathscr{P}_s)$ has finite optimal value and at least one optimal solution. Furthermore, the value*

$$\hat{\alpha} = \min_{X \in \mathscr{E}} F(X), \tag{3.2}$$

exists and is given by

$$\hat{\alpha} = E\{\alpha_s\}, \quad \textit{where } \alpha_s = \min(\mathscr{P}_s). \tag{3.3}$$

It is a lower bound for the optimal value in $(\mathscr{P})$.

PROOF. For the first part the argument is the standard one. The sets (3.1) for $\alpha > \inf(\mathscr{P}_s)$ are nonempty and compact under our assumptions, and since they are nested they must have a nonempty intersection. This intersection consists of the optimal solutions to $(\mathscr{P}_s)$. The existence of an optimal solution implies of course that the optimal value in $(\mathscr{P}_s)$ is finite. The second part of the proposition merely records that because of decomposability we are actually minimizing F over $\mathscr{C}$ when solving each of the problems $(\mathscr{P}_s)$. Indeed, $\mathscr{C}$ is just the direct product of the set $\mathscr{C}_s$ and F is by (2.7) separable, with components $p_s f_s$. The minimum value for $p_s f_s$ over C_s is $p_s \alpha_s$, and the sum of all these quantities $p_s \alpha_s$ is therefore $\hat{\alpha}$. This sum is $E\{\alpha_s\}$ in our probabilistic notation. Problem $(\mathscr{P})$ requires the minimization of F over $C \cap \mathscr{N}$, not just $\mathscr{C}$, so $\hat{\alpha}$ is merely a lower bound for the optimal value in $(\mathscr{P})$. □

Next we provide background for the solution of the subproblems $(\mathscr{P}^\nu)$ in our algorithm, which decompose into the modified scenario subproblems $(\mathscr{P}_s^\nu) = (\hat{\mathscr{P}}_s(\hat{X}^\nu(s), W^\nu(s), r))$.

PROPOSITION 3.2. *Every modified scenario subproblem of the form* $(\hat{\mathscr{P}}_s(\hat{x}, w, r))$ *(where* $r > 0$*) has finite optimal value and at least one optimal solution. In the convex case, this optimal solution is unique.*

PROOF. Let $\hat{f}_s$ denote the objective function in $(\hat{\mathscr{P}}_s(\hat{x}, w, r))$,

$$\hat{f}_s(x) = f_s(x) + x \cdot w + \tfrac{1}{2} r |x - \hat{x}|^2. \tag{3.4}$$

In the convex case, this is of course a strictly convex function on C_s and therefore has at most one minimizing point relative to C_s. To reach the desired conclusions it will suffice (in view of the existence argument used for the preceding proposition) to demonstrate that all level sets of the form $\{x \in C_s | \hat{f}_s(x) \leqslant \alpha\}$, $\alpha \in \mathbb{R}$, are closed and bounded. They are obviously closed, since C_s is closed and f_s is continuous. That they are bounded can be seen from the inequality

$$\hat{f}_s(x) \geqslant \alpha_s + x \cdot w + \tfrac{1}{2} r |x - \hat{x}|^2 \quad \text{for all } x \in C_s,$$

where α_s is the optimal value in $(\mathscr{P}_s)$ as in Proposition 2.2. This yields

$$\{x \in C_s | \hat{f}_s(x) \leqslant \alpha\} \subset \{x \in \mathbb{R}^n | x \cdot w + \tfrac{1}{2} r |x - \hat{x}|^2 \leqslant \alpha - \alpha_s\},$$

where the right side is a certain ball in $\mathbb{R}^n$. □

In the convex case, $(\hat{\mathscr{P}}_s(\hat{x}, w, r))$ is a convex programming problem. Thus, in executing our algorithm the critical step of solving all the modified subproblems $(\mathscr{P}_s^\nu)$ is open to the methods of convex programming. In the linear-quadratic case, these problems fall into the category of *quadratic* (convex) programming: a quadratic function with positive definite Hessian is minimized over a polyhedron. Special techniques such as pivoting algorithms can then produce an "*exact*" optimal solution to $(\mathscr{P}_s^\nu)$ as long as the dimension n and the number of linear constraints used in defining C_s are not too large.

In the important case where f_s is linear, i.e., where the original scenario subproblems $(\mathscr{P}_s)$ arise from a linear programming model, the nature of $(\hat{\mathscr{P}}_s(\hat{x}, w, r))$ and

$(\mathscr{P}_s^\nu)$ is even more special. Although the proximal term in $|x - \hat{x}|^2$ requires a quadratic programming technique rather than the simplex method, say, in solving such a subproblem the Hessian matrix is just rI. It is possible then by elementary algebra to reduce the subproblem to special one of finding the point of C_s *nearest* to a certain point in $\mathbb{R}^n$, namely in the case of $(\mathscr{P}_s^\nu)$ the point $\hat{X}^\nu(s) - r^{-1}W^\nu(s)$.

Another thing that should be noted about the modified scenario subproblems solved in our algorithm is the quite simple way they can be updated from one iteration to the next. In iteration ν we must (approximately) solve

$$(\mathscr{P}_s^\nu) \qquad \text{minimize } f_s(x) + \langle x, W^\nu(s)\rangle + \tfrac{1}{2}r|x - \hat{X}^\nu(s)|^2 \quad \text{over } C_s,$$

but in the preceding iteration we already solved

$$(\mathscr{P}_s^{\nu-1}) \quad \text{minimize } f_s(x) + \langle x, W^{\nu-1}(s)\rangle + \tfrac{1}{2}r|x - \hat{X}^{\nu-1}(s)|^2 \quad \text{over } C_s$$

in order to get $X^\nu(s)$, and we then set

$$W^\nu(s) = W^{\nu-1}(s) + r\big[X^\nu(s) - \hat{X}^\nu(s)\big].$$

By expanding the objectives in these two subproblems around the initial $\hat{X}^0(s)$ (as a suitable reference point), we can express the objective in $(\mathscr{P}_s^{\nu-1})$ as

$$f_s(x) - \alpha_s + \tfrac{1}{2}r|x - \hat{X}^0(s)|^2 + x \cdot \big(W^{\nu-1}(s) - r\big[\hat{X}^{\nu-1}(s) - \hat{X}^0(s)\big]\big) + \text{const.}$$

where $\alpha_s = \min(\mathscr{P}_s)$, and the objective in $(\mathscr{P}_s^\nu)$ is

$$f_s(x) - \alpha_s + \tfrac{1}{2}r|x - \hat{X}^0(s)|^2 + x \cdot \big(W^\nu(s) - r\big[\hat{X}^\nu(s) - \hat{X}^0(s)\big]\big) + \text{const.}$$

The value α_s has been introduced in these expressions because the common portion before the inner product is then nonnegative and vanishes when $x = \hat{X}^0(s)$. The important observation, since constant terms in an objective have no effect on the calculation of an optimal solution, is that *the objectives in* $(\mathscr{P}_s^{\nu-1})$ *and* $(\mathscr{P}_s^\nu)$ *differ only in a linear term*. As a matter of fact, the linear terms in the objectives differ in coefficient only by

$$\big(W^\nu(s) - r\big[\hat{X}^\nu(s) - \hat{X}^0(s)\big]\big) - \big(W^{\nu-1}(s) - r\big[\hat{X}^{\nu-1}(s) - \hat{X}^0(s)\big]\big)$$

$$= r\big(\big[X^\nu(s) - \hat{X}^\nu(s)\big] - \big[\hat{X}^\nu(s) - \hat{X}^{\nu-1}(s)\big]\big).$$

In passing from $(\mathscr{P}_s^{\nu-1})$ *to* $(\mathscr{P}_s^\nu)$ *we therefore need only add to the objective a linear term with this vector as its coefficient vector*, in order to move toward calculating the new elements $X^{\nu+1}(s)$.

The reason this observation can be useful is that it allows parametric techniques to come into play, particularly in the linear-quadratic case, in solving the modified scenario subproblems. The work involved can thereby be reduced very significantly. Other possibilities for reducing effort could lie in the information structure at hand. If scenarios s and s' are almost the same, for instance if they are indistinguishable to the decision maker until final time periods, then $(\mathscr{P}_s^\nu)$ and $(\mathscr{P}_{s'}^\nu)$ ought to have strong similarities. One might be able to take advantage of an overlap in form to increase efficiency in solving the two problems, or a "bunch" of such problems. This is an idea

that can only be developed in terms of greater detail about the scenario subproblems than we are ready to explore in the present paper.

Let us now look at problem $(\mathscr{P})$ itself.

PROPOSITION 3.3. *In problem $(\mathscr{P})$ the feasible set $\mathscr{E}$ is nonempty and closed, the objective F is locally Lipschitz continuous on $\mathscr{E}$, and all level sets of the form*

$$\{X \in \mathscr{E} \mid F(x) \leqslant \alpha\}, \qquad \alpha \in \mathbb{R}, \tag{3.5}$$

are compact.

PROOF. The assertions about $\mathscr{E}$ and F are obvious from the corresponding assumptions about C_s and f_s for each $s \in S$. They imply the closedness of the sets (3.5). The boundedness is verified by using the constant $\hat{\alpha}$ in Proposition 3.1 to express the inequality $F(X) \leqslant \alpha$ as

$$\alpha - \hat{\alpha} \geqslant F(X) - \hat{\alpha} = \sum_{s \in S} p_s [f_s(X(s)) - \alpha_s].$$

This inequality implies $f_s(X(s)) \leqslant \alpha_s + [\alpha - \hat{\alpha}]/p_s$ for each $s \in S$. Any set (3.5) is therefore included in a set of the form

$$\prod_{s \in S} \{x \in C_s \mid f_s(x) \leqslant \alpha_s + [\alpha - \hat{\alpha}]/p_s\},$$

where by one of our basic assumptions each factor is bounded. It follows that any set (3.5) is compact. All sets of the form $\{X \in \mathscr{E} \cap \mathscr{N} \mid F(X) \leqslant \alpha\}$ for $\alpha \in \mathbb{R}$ are then compact too. In $(\mathscr{P})$ we minimize F over $\mathscr{E} \cap \mathscr{N}$, so this compactness leads by the standard existence argument in the proof of Proposition 3.1 to the assurance that, when $\mathscr{E} \cap \mathscr{N} \neq \emptyset$, problem $(\mathscr{P})$ has an optimal solution and consequently finite optimal value. □

A further observation about the nature of $(\mathscr{P})$ will complete this section.

PROPOSITION 3.4. *In the convex case $(\mathscr{P})$ is a (large-scale) problem of convex programming: the feasible set $\mathscr{E}$ is convex and the objective F is convex. In the linear-quadratic case $(\mathscr{P})$ is a (large-scale) problem of linear or quadratic programming: $\mathscr{E}$ is a polyhedron in $\mathscr{E}$ and F is linear or (convex) quadratic.*

PROOF. In the first case $\mathscr{E}$ is a product of convex sets and F is a sum of convex functions. In the second case $\mathscr{E}$ is a product of polyhedral sets, hence polyhedral, and F is a sum of functions that are at most quadratic, hence itself is at most quadratic. □

The *large-scale* nature of $(\mathscr{P})$, mentioned in Proposition 3.4, stems partly from the very introduction of scenarios in the mathematical model. As soon as one attempts to cover a variety of occurrences that could influence the decision process, one almost inevitably becomes interested in a scenario set S as large as technically can be managed in the calculation of solutions. Then in addition there is the presence of multiple time periods. This could itself lead to large-scale structure. Each of the scenario subproblems $(\mathscr{P}_s)$ might itself be a challenge. The fact that we shall be able to decompose $(\mathscr{P})$ into solving modified versions $(\mathscr{P}_s^\nu)$ of such subproblems may in that situation seem to have only a muted effect, even if parallel processing or the like is available for the subproblems. However, the principle developed in this paper need only cover an outer layer. The problems $(\mathscr{P}_s^\nu)$, with their multiple time periods but fixed (not "uncertain") structure, could themselves be decomposed by other techniques. In particular we have in mind here the idea of algorithms based on the separable saddle point representations we have developed recently in [13].

4. Optimality and duality. The question of what conditions can be used to characterize optimal solutions to $(\mathscr{P})$ has to be addressed for its own reasons, but it is critical also in the formulation of a notion of "approximate" solution that can be used in implementing our algorithm. The interpretation of the multiplier elements W^ν in the algorithm is involved with this matter as well.

To cover with adequate generality the diverse instances of the scenario subproblems $(\mathscr{P}_s)$ that interest us, where f_s might be a smooth function but on the other hand might be convex and only piecewise smooth, due to the introduction of penalty terms, we use the notation of nonsmooth analysis. The symbol $\partial f_s(x)$ will denote the set of *generalized subgradients* of f_s at x, as defined by Clarke [14] for arbitrary Lipschitz continuous functions. The reader does not need to understand fully what this means in order to appreciate our results. The main facts are simply that if f_s happens to be smooth (continuously differentiable) the set $\partial f_s(x)$ consists of the single vector $\nabla f_s(x)$, whereas if f_s is convex $\partial f_s(x)$ is the usual subgradient set in convexity theory [16]. In all cases $\partial f_s(x)$ is a nonempty compact convex set that depends on x.

Similarly the symbol $N_{C_s}(x)$ will denote the *generalized normal cone* to C_s at x, as defined for any closed set C_s [14]. If C_s is convex, this is the normal cone of convex analysis [16]. Whether convex or not, if C_s is defined by a system of smooth constraints such that the Mangasarian-Fromovitz constraint qualification is satisfied at x, then $N_{C_s}(x)$ is the polyhedral cone generated by the gradients of the active constraints at x. (Nonnegative coefficients are used for the gradients of the active inequality constraints, of course, and arbitrary coefficients for the equality constraints.) The set $N_{C_s}(x)$ is always a closed convex cone containing the zero vector, and it reduces solely to the zero vector if and only if x is an interior point of C_s.

This notation and its interpretations can be carried over to $\mathscr{E}$ and F in problem $(\mathscr{P})$ as well. Many formulas for determining subgradient sets and normal cones in particular situations can be found in [14], [15] and [16].

THEOREM 4.1. *Let X^* be a feasible solution to $(\mathscr{P})$: one has $X^* \in \mathscr{N}$ and $X^* \in \mathscr{E}$, i.e.*

$$X^*(s) \in C_s \quad \text{for all } s \in S. \tag{4.1}$$

Suppose that X^* is locally optimal and satisfies the following constraint qualification:

$$\text{the only } W \in \mathscr{M} \text{ satisfying } -W(s) \in N_{C_s}(X^*(s)) \text{ for all } s \in S \text{ is } W = 0. \tag{4.2}$$

Then there exists $W^* \in \mathscr{M}$ satisfying

$$-W^* \in \partial F(X^*) + N_{\mathscr{E}}(X^*), \tag{4.3}$$

and this is equivalent to

$$-W^*(s) \in \partial f_s(X^*(s)) + N_{C_s}(X^*(s)) \quad \text{for all } s \in S. \tag{4.4}$$

In the convex case, the existence of such an element W^* implies conversely that X^* is an optimal solution to $(\mathscr{P})$ (in the global sense).

PROOF. The overall character of this result is not surprising, but its formulation in terms of conditions in (4.1), (4.2) and (4.4) that concern C_s and f_s for each $s \in S$ needs to be checked for correctness. The two crucial formulas which yield this

formulation are

$$\partial F(X) = \{Y \in \mathscr{E} | Y(s) \in \partial f_s(X(s)) \text{ for all } s \in S\}, \tag{4.5}$$

$$\partial N_{\mathscr{C}}(X) = \{Y \in \mathscr{E} | Y(s) \in N_{C_s}(X(s)) \text{ for all } s \in S\}. \tag{4.6}$$

These are perhaps more subtle than may first appear, because subgradients and normal vectors depend by definition on the *particular inner product* being used in the Euclidean space in question, and our inner product (2.2), given by an expectation, is a specially adapted one.

We can think of the Euclidean space $\mathscr{E}$ as the direct product of Euclidean space $\mathscr{E}_s$ for $s \in S$, where $\mathscr{E}_s$ is $\mathbb{R}^n$ under the rescaled inner product

$$\langle x, y \rangle_s = p_s[x \cdot y]. \tag{4.7}$$

Correspondingly F can be viewed in the separable form

$$F(X) = \sum_{s \in S} F_s(X(s)) \quad \text{with } F_s := p_s f_s \text{ on } \mathscr{E}_s, \tag{4.8}$$

and $\mathscr{C}$ can be viewed of course as the product of the sets C_s in the space $\mathscr{E}_s$. According to a general formula of nonsmooth analysis proved in Rockafellar [16] (Proposition 2.5 and Corollary 2.5.1), one then has

$$\partial F(X) = \prod_{s \in S} \tilde{\partial} F_s(X(s)), \tag{4.9}$$

$$N_{\mathscr{C}}(X) = \prod_{s \in S} \tilde{N}_{C_s}(X(s)), \tag{4.10}$$

where the tilde $\tilde{}$ is introduced to indicate that the subgradient set and normal cone are to be taken *relative to the inner product* (4.7) rather than the canonical one. In the case of the normal cones this modification makes no difference at all, because the nature of a cone is not affected by a positive rescaling. Thus (4.10) is equivalent to (4.6). On the other hand

$$\tilde{\partial} F_s(X(s)) = \{p_s^{-1} z | z \in \partial F_s(X(s))\}, \qquad \partial F_s(X(s)) = \{p_s y | y \in \partial f_s(X(s))\},$$

so in the end we have $\tilde{\partial} F_s(X(s)) = \partial f_s(X(s))$. Formula (4.9) therefore reduces to (4.5).

Armed with (4.5) and (4.6) we can apply the general theory of necessary conditions in nonsmooth analysis to problem $(\mathscr{P})$. Viewing $(\mathscr{P})$ in terms of minimizing F (which is Lipschitz continuous by Proposition 3.3) over $\mathscr{C} \cap \mathscr{N}$, we first invoke the basic result (Corollary 2.4.1 of [16]) to conclude that if X^* gives a local minimum then

$$0 \in \partial F(X^*) + N_{\mathscr{C} \cap \mathscr{N}}(X^*). \tag{4.11}$$

Next we recall from convex analysis (cf. Corollary 8.1.2 of [16]) that

$$N_{\mathscr{C} \cap \mathscr{N}}(X^*) \subset N_{\mathscr{C}}(X^*) + N_{\mathscr{N}}(X^*) \tag{4.12}$$

as long as there does not exist

$$W \in N_{\mathscr{N}}(X^*) \quad \text{with} \quad -W \in N_{\mathscr{C}}(X^*),\ W \neq 0. \tag{4.13}$$

Formula (4.6) gives us $N_{\mathscr{C}}(X^*)$, and since $\mathscr{N}$ is a *subspace* of $\mathscr{E}$, the normal cone $N_{\mathscr{N}}(X^*)$ is just the subspace orthogonally complementary to $\mathscr{N}$ (with respect to the specified inner product for $\mathscr{E}$), namely $\mathscr{M}$. The nonexistence of a vector W having the properties in (4.13) is thus the condition we have set up in (4.2) as the constraint qualification for $(\mathscr{P})$. The combination of (4.11) and (4.12) now comes down to the assertion that

$$-W^* \in \partial F(X^*) + N_{\mathscr{C}}(X^*) \quad \text{for some } W^* \in \mathscr{M}, \tag{4.14}$$

where the subgradient condition reduces by (4.5) and (4.6) to the relations claimed in (4.4).

In the convex case, of course, all these subgradient calculations can be carried out in the less demanding context of convex analysis rather than general nonsmooth analysis. The asserted conditions for optimality, which are equivalent to (4.14), are then sufficient because of the stronger meaning assigned to subgradients and normal vectors in that context. Specifically, (4.14) says that for some $Y \in \partial F(X^*)$, which means

$$F(X) \geqslant F(X^*) + \langle X - X^*, Y \rangle \quad \text{for all } X \in \mathscr{E}, \tag{4.15}$$

the vector $-W^* - Y$ belongs to $N_{\mathscr{C}}(X^*)$, which means

$$\langle X - X^*, -W^* - Y \rangle \leqslant 0 \quad \text{for all } X \in \mathscr{C}. \tag{4.16}$$

Taking arbitrary $X \in \mathscr{C} \cap \mathscr{N}$ and using the fact that $\langle X, W^* \rangle = 0$ and $\langle X^*, W \rangle = 0$ (because $W \in \mathscr{M}$ and $\mathscr{M} \perp \mathscr{N}$) we see in (4.15) that (4.16) implies $\langle X - X^*, Y \rangle \geqslant 0$ and therefore $F(X) \geqslant F(X^*)$. Thus X^* is globally optimal for $(\mathscr{P})$ in this case. □

THEOREM 4.2. *In the convex case, the decomposed conditions* (4.1) *and* (4.4) *on a pair* $(X^*, W^*) \in \mathscr{N} \times \mathscr{M}$ *are equivalent to* (X^*, W^*) *being a saddle point of the ordinary Lagrangian* $L(X, W) = F(X) + \langle X, W \rangle$ *relative to minimizing over* $X \in \mathscr{C}$ *and maximizing over* $W \in \mathscr{M}$.

PROOF. This is just a small extension of the argument with which we concluded the preceding proof. It fits the standard patterns of convex analysis, so we omit it. □

THEOREM 4.3. *In the linear quadratic case, the constraint qualification in Theorem* 4.1 *is superfluous. The condition given for optimality is always both necessary and sufficient.*

PROOF. In this case $(\mathscr{P})$ is just a linear or quadratic programming problem, albeit of large size; cf. Proposition 3.4. In particular $\mathscr{C}$ is a polyhedron and F is smooth, so no constraint qualification is needed for the general optimality condition (4.3) to be necessary. □

As support for our algorithm we must develop optimality conditions for the subproblems $(\mathscr{P}^\nu)$ and $(\mathscr{P}_s^\nu) = (\hat{\mathscr{P}}_s(\hat{X}^\nu(s), W^\nu(s), r))$ as well. Fortunately the circumstances in these problems are closely parallel to the one already treated, so there is no call for going through the arguments in detail. We simply state the results without writing out the proofs.

PROPOSITION 4.4. *If a policy $X^{\nu+1}$ is locally optimal for the subproblem*

$$(\mathscr{P}^\nu) \qquad \text{minimize } F(X) + \langle X, W^\nu \rangle + \tfrac{1}{2}r\|X - \hat{X}^\nu\|^2 \quad \text{over } \mathscr{C},$$

it satisfies

$$(4.17) \quad X^{\nu+1} \in \mathscr{C} \quad \text{and} \quad -W^\nu - r[X^{\nu+1} - \hat{X}^\nu] \in \partial F(X^{\nu+1}) + N_{\mathscr{C}}(X^{\nu+1}),$$

and this is equivalent to

$$(4.18) \qquad X^{\nu+1}(s) \in C_s \quad \text{for all } s \in S,$$

$$(4.19) \quad -W^\nu(s) - r[X^{\nu+1}(s) - \hat{X}^\nu(s)] \in \partial f_s(X^{\nu+1}(s)) + N_{C_s}(X^{\nu+1}(s)) \quad \text{for all } s \in S.$$

In the convex case, this property of $X^{\nu+1}$ implies conversely that $X^{\nu+1}$ is the unique (globally) optimal solution to $(\mathscr{P}^\nu)$. Conditions (4.18) *and* (4.19) *in fact characterize in the same pattern the optimality of $X^{\nu+1}(s)$ for the subproblem $(\mathscr{P}_s^\nu)$.*

The main point here is that problem $(\mathscr{P}^\nu)$ decomposes into the individual problems $(\mathscr{P}_s^\nu)$. The conditions in Proposition 4.4 are the ones obtained for each $(\mathscr{P}_s^\nu)$. No constraint qualification is needed, because the subspace $\mathscr{N}$ is not involved.

Finally, the connection with duality in the convex case must be noted. The problem *dual* to $(\mathscr{P})$ with respect to the ordinary Lagrangian L is

$$(\mathscr{D}) \qquad \text{maximize } G(W) \text{ over all } W \in \mathscr{D} \cap \mathscr{M},$$

where

$$(4.20) \qquad G(W) = \inf_{X \in \mathscr{C}} L(X, W) = \inf_{X \in \mathscr{C}} \{F(X) + \langle X, W \rangle\},$$

$$(4.21) \qquad \mathscr{D} = \{W | G(W) > -\infty\}.$$

The working out of the formula for the dual objective is not really relevant for our purposes. Instead we are interested in the relationship between $(\mathscr{D})$ and $(\mathscr{P})$ insofar as it reflects on the character and interpretation of the multipliers W. The facts can be derived from the general duality theory for convex programming problems in [16]. They focus most significantly on the function

$$(4.22) \qquad \Phi(U) = \min\{F(X) | X \in \mathscr{C}, KX = U\} \quad \text{for } U \in \mathscr{M}.$$

This expresses the optimal value in a *perturbed* form of $(\mathscr{P})$, where the implementability constraint $KX = 0$ is relaxed to $KX = U$. Note that the minimum in the formula is indeed attained as long as there does exist an $X \in \mathscr{C}$ satisfying $KX = U$. This is clear from the compactness in Proposition 3.3. When there does exist such an X, $\Phi(U)$ is regarded as ∞. Thus Φ is extended-real-valued but nowhere takes on $-\infty$ (because of the attainment of the minimum). Its domain of finiteness is the nonempty set $K\mathscr{C}$, the projection of $\mathscr{C}$ on $\mathscr{M}$.

PROPOSITION 4.5. *The functional* Φ *on* $\mathscr{M}$ *is lower semicontinuous, in fact its level sets* $\{U \in \mathscr{M} | \Phi(U) \leqslant \alpha\}$ *for* $\alpha \in \mathbb{R}$ *are all compact. Furthermore*

$$\Phi(0) = \min(\mathscr{P})$$

(where $\min(\mathscr{P})$ *is the optimal value in* $(\mathscr{P})$ *and is interpreted as* ∞ *if* $(\mathscr{P})$ *has no feasible solution, i.e., if* $\mathscr{C} \cap \mathscr{N} = \emptyset$*), and*

$$\min_{U \in \mathscr{M}} \Phi(U) = \hat{\alpha},$$

where $\hat{\alpha}$ *is the value in Proposition* 2.1. *In the convex case,* Φ *is convex on* $\mathscr{M}$.

PROOF. The level set $\{U \in \mathscr{M} | \Phi(U) \leqslant \alpha\}$ is simply the image under the projection K of the level set $\{X \in \mathscr{C} | F(X) \leqslant \alpha\}$. The latter is compact by Proposition 3.3., so the former is compact as well. This point of view also makes obvious the fact that the minimum value of Φ on $\mathscr{M}$ is the same as the minimum value of F on $\mathscr{C}$, which is $\hat{\alpha}$ by Proposition 3.1. The epigraph of Φ is seen in the same way to be the image of the epigraph of $F + \delta_C$ (with δ_C the indicator of C) under the extended projection $(X, \alpha) \mapsto (KX, \alpha)$ from $\mathscr{C} \times \mathbb{R}$ onto $\mathscr{M} \times \mathbb{R}$. In the convex case the epigraph of $F + \delta_C$ is a convex set, hence so is the epigraph of Φ. Thus Φ is a convex function. □

THEOREM 4.6. *In the convex case the relation*

$$-\infty < \min(\mathscr{P}) = \sup(\mathscr{D}) \leqslant \infty \tag{4.23}$$

holds, and moreover the set of all optimal solutions to $(\mathscr{D})$ *is given by*

$$\begin{aligned} \operatorname{argmax}(\mathscr{D}) &= -\partial\Phi(0) \\ &= \{W^* \in \mathscr{M} | \Phi(U) \geqslant \Phi(0) - \langle U, W^* \rangle \text{ for all } U \in \mathscr{M}\}. \end{aligned} \tag{4.24}$$

The elements W^* *in this set, if any, are precisely the ones associated with an optimal solution* X^* *to* $(\mathscr{P})$ *by the optimality conditions in Theorem* 4.1 *or Theorem* 4.2.

In particular the set (4.24) *is nonempty if* $(\mathscr{P})$ *has an optimal solution* S^* *and the constraint qualification in Theorem* 4.1 *is satisfied. In the linear-quadratic case it is sure to be nonempty from* $(\mathscr{P})$ *being feasible, i.e., having* $\mathscr{C} \cap \mathscr{N} \neq \emptyset$.

PROOF. This specializes the theory in §30 of [16] to the present case. The assertions about the set (4.24) being nonempty are justified by Theorem 4.1 and 4.2 (and to some extent the existence in Proposition 3.3 of an optimal solution to $(\mathscr{P})$ when $\mathscr{C} \cap \mathscr{N} \neq \emptyset$). □

The importance of formula (4.24) in Theorem 4.6 is that it identifies the optimal multipliers W^* in our framework with subdifferentiability properties of the convex functional Φ. The subgradient inequality

$$\Phi(U) \geqslant \Phi(0) - \langle U, W^* \rangle \quad \text{for all } U \in \mathscr{M}, \tag{4.25}$$

furnishes us a means of seeing what W^* represents.

Let us go back to the idea that W^* is an "information price system" and give it the following, more specific interpretation: $W^*(s)$ is a price vector that can be used, if the scenario finally turns out to be s, to take the decision $X(s)$, which had to be chosen as part of an implementable policy X, and change it *with hindsight* to a different decision $X'(s) = X(s) + U(s)$. The cost of this service is $U(s) \cdot W^*(s)$.

Taking all possible scenarios into account with their various weights, and imagining how one might want to alter decisions after the fact in all cases, we come up with the cost expression

$$\sum_{s \in S} p_s[U(s) \cdot W^*(s)] = \langle U, W^* \rangle.$$

Only deviations U that belong to $\mathcal{M}$ need to be considered, because all other aspects of the uncertain environment could already be taken care of in our model through the selection of X as an *implementable* policy.

The inequality (4.25) expresses W^* as a system of "*equilibrium*" prices in the sense that under this system is no incentive for such *a posteriori* change of decisions. A change represented by a deviation $U \in \mathcal{M}$ would achieve $\Phi(U)$ in place of $\Phi(0)$ as the optimal value in the problem, but the cost of the change, as perceived at the time of decision making, would be $\langle U, W^* \rangle$. The net result for the decision maker would be $\Phi(U) + \langle U, W^* \rangle$. Because of the inequality in (4.25), there is no advantage in this procedure as compared with just accepting the implementability constraint $X \in \mathcal{N}$ and the corresponding optimal value $\Phi(0)$.

In summary, the price systems W^* are the ones that would charge for hindsight everything it might be worth. They do therefore truly embody the value of information in the uncertain environment.

Tighter expressions than (4.25) can be derived under additional assumptions. For instance, if W^* is unique then $W^* = -\nabla\Phi(0)$. We refer to the theory of subgradients of convex functions in §§23–25 of [16].

5. Convergence in the convex case. This section contains our main results. So as not to overburden the reader with all the details at once, we begin with the form of the algorithm in which exact solutions are calculated for the subproblems in Step 2. This is referred to as the case of *exact minimization* in contrast to the case of *approximate minimization* that will be treated afterward. Exact minimization, as the reader will recall, makes sense when the scenario subproblems fall within the realm of linear or quadratic programming and are not themselves of large scale.

Our essential line of argument will be to show that the progressive hedging algorithm can be constructed in terms of the generation of the sequence $\{(\hat{X}^\nu, W^\nu)\}_{\nu=1}^\infty$ (or rather, a slightly rescaled version of it) as a certain instance of the *proximal point algorithm* developed in Rockafellar [3]. The convergence theorems in [3] then give us what we want. This same line of argument has been followed by Spingarn [4], [7], in obtaining a wider class of decomposition methods expressed through operations on subgradient mappings. Rather than pass through the intermediary of Spingarn's formulation, however, we appeal directly to the original theory and emphasize the connection with calculating a certain "proximal saddle point" in each iteration. This has the advantage of illuminating the relationship between our procedure and augmented Lagrangian methods, which similarly can be treated in the convex case via the proximal point algorithm, either in a dual or a saddle point (i.e., primal-dual) form [6].

THEOREM 5.1. *Consider the algorithm in the convex case with exact minimization. Let* $\{\hat{X}^\nu\}_{\nu=1}^\infty$ *and* $\{W^\nu\}_{\nu=1}^\infty$ *be the sequences it generates from an arbitrary initial choice of* $X^0 \in \mathcal{E}$ *and* $W^0 \in \mathcal{M}$. *(In particular* $X^0(s)$ *could be obtained by solving* $(\mathcal{P}_s)$, *but that is not presupposed here.)*

These sequences will be bounded if and only if optimal solutions exist for the subproblems $(\mathcal{P})$ *and* $(\mathcal{D})$, *i.e., there exist elements* X^* *and* W^* *satisfying the*

optimality conditions in Theorem 4.1, *or equivalently the saddle point condition in Theorem* 4.2. *In that case, for some particular pair of such elements* X^* *and* W^* (*even though optimal solutions to* $(\mathscr{P})$ *and* $(\mathscr{D})$ *might not be unique*), *it will be true that*

$$\hat{X}^\nu \to X^* \quad \text{and} \quad W^\nu \to W^*. \tag{5.1}$$

Furthermore, in terms of the norm expression

$$\|(X,W)\|_r = \left(\|X\|^2 + r^{-2}\|W\|^2\right)^{1/2} \tag{5.2}$$

one will have in every iteration $\nu = 0, 1, 2, \ldots$ *that*

$$\|(\hat{X}^{\nu+1}, W^{\nu+1}) - (X^*, W^*)\|_r \leq \|(\hat{X}^\nu, W^\nu) - (X^*, W^*)\|_r, \tag{5.3}$$

with strict inequality unless $(\hat{X}^\nu, W^\nu) = (X^*, W^*)$.

Thus every iteration of the algorithm from the start makes a definite improvement until solutions are attained (*if that occurs in finitely many steps*). *One will also have in every iteration* $\nu = 1, 2, \ldots$ *that*

$$\|(\hat{X}^{\nu+1}, W^{\nu+1}) - (\hat{X}^\nu, W^\nu)\|_r \leq \|(\hat{X}^\nu, W^\nu) - (\hat{X}^{\nu+1}, W^{\nu-1})\|_r. \tag{5.4}$$

PROOF. A slight shift of notation will be useful. Let

$$V^\nu = \hat{X}^\nu, \tag{5.5}$$

and rescale all multiplier vectors by

$$\overline{W} = r^{-1}W, \qquad \overline{W}^\nu = r^{-1}W^\nu, \qquad \overline{W}^* = r^{-1}W^* \quad \text{(etc.)}. \tag{5.6}$$

Then

$$\|(X,W)\|_r = \|(X,\overline{W})\| = \left(\|X\|^2 + \|\overline{W}\|^2\right)^{1/2}. \tag{5.7}$$

Consider now the "projected" saddle function

$$l_r(V,\overline{W}) = \inf\{r^{-1}F(X) + \langle X, \overline{W}\rangle \mid X \in \mathscr{C}, \hat{X} = V\} \quad \text{for } V \in \mathscr{N}, \overline{W} \in \mathscr{M}. \tag{5.8}$$

We shall show that the progressive algorithm in the convex case with exact minimization takes the form that

$$\begin{aligned} &(V^{\nu+1}, \overline{W}^{\nu+1}) \text{ is the (unique) saddle point of} \\ &l_r^\nu(V,\overline{W}) = l_r(V,\overline{W}) + \tfrac{1}{2}\|V - V^\nu\|^2 - \tfrac{1}{2}\|\overline{W} - \overline{W}^\nu\|^2 \\ &\text{subject to minimizing in } V \in \mathscr{N} \text{ and maximizing in } \overline{W} \in \mathscr{M}. \end{aligned} \tag{5.9}$$

The nature of the function l_r must be clarified first. We can write the formula for l_r in (5.8) alternatively as

$$l_r(V,\overline{W}) = \inf_{U \in \mathscr{M}} \{\varphi_r(V,U) + \langle U, \overline{W}\rangle\}, \tag{5.10}$$

where

$$\varphi_r(V,U) = r^{-1}F(X) + \delta_c(X) \text{ for the unique } X \in \mathscr{E} \text{ having } JX = V,\ KX = U. \tag{5.11}$$

This just amounts to a "change of coordinates" corresponding to the orthogonal decomposition $\mathscr{E} = \mathscr{N} \times \mathscr{M}$ which expresses the closed proper convex function $r^{-1}F + \delta_{\mathscr{C}}$ (where $\delta_{\mathscr{C}}$ is the indicator of $\mathscr{C}$) in terms of the components V and U of X. Obviously φ_r is a closed proper convex function on $\mathscr{N} \times \mathscr{M}$, and from this it follows by (5.10) that l_r is a closed proper saddle function on $\mathscr{N} \times \mathscr{M}$ in the terminology of convex analysis (see §34 of [16]).

Associated with l_r is the multifunction

$$T_r(V,\overline{W}) = \{(Y,-U) \mid (Y,U) \in \partial l_r(V,\overline{W})\}, \tag{5.12}$$

which is known to be *maximal monotone* and to be given also by

$$T_r(V,\overline{W}) = \{(Y,-U) \mid (Y,-\overline{W}) \in \partial \varphi_r(V,U)\}. \tag{5.13}$$

(See §35 of [16] for a review of subgradients of saddle functions.) Our claim in (5.9) is that

$$(0,0) \in \partial l_r^{\nu}(V^{\nu+1},\overline{W}^{\nu+1}) = \partial l_r(V^{\nu+1},\overline{W}^{\nu+1}) + (V^{\nu+1} - V^{\nu}, \overline{W}^{\nu} - \overline{W}^{\nu+1}),$$

or in terms of T_r that

$$(V^{\nu} - V^{\nu+1}, \overline{W}^{\nu} - \overline{W}^{\nu+1}) \in T_r(V^{\nu+1},\overline{W}^{\nu+1}). \tag{5.14}$$

In order to verify this claim we employ (5.13) and (5.11) to get an expression for T_r in the original context of F and $\mathscr{C}$. We have

$$(Y,-\overline{W}) \in \partial\varphi_r(V,U) \Leftrightarrow Y - \overline{W} \in \partial\big(r^{-1}F + \delta_{\mathscr{C}}\big)(X), \tag{5.15}$$

where $X = V + U$.

Moreover the subdifferentiation rules of convex analysis (§23 of [16]) yield

$$\partial\big(r^{-1}F + \delta_{\mathscr{C}}\big)(X) = r^{-1}\partial F(X) + N_{\mathscr{C}}(X). \tag{5.16}$$

Therefore by (5.13),

$$(Y,-U) \in T_r(V,\overline{W}) \Leftrightarrow Y - \overline{W} \in r^{-1}\partial F(X) + N_{\mathscr{C}}(X), \tag{5.17}$$

where $X = V + U, V \in \mathscr{N}, U \in \mathscr{M}, Y \in \mathscr{N}$

We can now transmute our claim from (5.14) into the equivalent form

$$V^{\nu} - V^{\nu+1} - \overline{W}^{\nu+1} \in r^{-1}\partial F(V^{\nu+1} + \overline{W}^{\nu+1} - \overline{W}^{\nu}) + N_{\mathscr{C}}(V^{\nu+1} + \overline{W}^{\nu+1} - \overline{W}^{\nu}). \tag{5.18}$$

At this point a return to our original notation is in order. From (5.5), (5.6) and the

updating formula

$$W^{\nu+1} = W^{\nu} + r[X^{\nu+1} - \hat{X}^{\nu+1}]$$

used in our algorithm, which says $\overline{W}^{\nu+1} - \overline{W}^{\nu} = X^{\nu+1} + \hat{X}^{\nu+1}$, we get

$$V^{\nu+1} + \overline{W}^{\nu+1} - \overline{W}^{\nu} = \hat{X}^{\nu+1} + X^{\nu+1} - \hat{X}^{\nu+1} = X^{\nu+1},$$

$$V^{\nu} - V^{\nu+1} - \overline{W}^{\nu+1} = \hat{X}^{\nu} - \hat{X}^{\nu+1} - r^{-1}W^{\nu+1} = \hat{X}^{\nu} - X^{\nu+1} - r^{-1}W^{\nu}.$$

Thus (5.18) as a version of our claim is the same as

$$0 \in \partial F(X^{\nu+1}) + N_{\mathscr{C}}(X^{\nu+1}) + W^{\nu} + r(X^{\nu+1} - \hat{X}^{\nu}),$$

or better still

$$0 \in \partial F^{\nu}(X^{\nu+1}) + N_{\mathscr{C}}(X^{\nu+1}), \tag{5.19}$$

where F^{ν} is the objective in the subproblem $(\mathscr{P}^{\nu})$ in Step 2 of our algorithmic procedure,

$$F^{\nu}(X) = F(X) + \langle X, W^{\nu}\rangle + \tfrac{1}{2}r\|X - \hat{X}^{\nu}\|^2. \tag{5.20}$$

But because we are working at present in a framework of convex programming, the subdifferential condition (5.19) is both necessary and sufficient for

$$X^{\nu+1} \in \operatorname*{argmin}_{x \in \mathscr{C}} F^{\nu}(X) = \operatorname{argmin}(\mathscr{P}^{\nu}). \tag{5.21}$$

Recapitulating up to this point, we have established that the three conditions (5.9), (5.14), and (5.21) are equivalent. The uniqueness mentioned in (5.9) is evident from the strict convexity-concavity induced on the function l_r^{ν} by the proximal terms in V and $\overline{W}$. It is equivalent also to the uniqueness of $X^{\nu+1}$ in (5.21), which comes from the proximal term in (5.20). In terms of the operator T_r, which will play an ever more important role in our analysis, the uniqueness property is expressed by writing (5.14) as

$$(V^{\nu+1}, \overline{W}^{\nu+1}) = M_r(V^{\nu}, \overline{W}^{\nu}), \tag{5.22}$$

where

$$M_r = (I + T_r)^{-1}. \tag{5.23}$$

Although T_r is itself generally multivalued, its maximal monotonicity ensures, as is well known (cf. Minty [17]), that the operator M_r is single-valued everywhere and actually *nonexpansive*. This means in the notation

$$Z = (V, \overline{W}), \qquad Z^{\nu} = (V^{\nu}, \overline{W}^{\nu}), \qquad Z^{*} = (V^{*}, \overline{W}^{*}), \tag{5.24}$$

that one has always

$$\|M_r(Z') - M_r(Z')\| \leqslant \|Z' - Z\|. \tag{5.25}$$

Indeed, (5.25) can be stated even more strongly (cf. Proposition 1 of [3]) as

$$\|M_r(Z') - M_r(Z)\|^2 + \|(Z' - M_r(Z')) - (Z - M_r(Z))\|^2 \leqslant \|Z' - Z\|^2. \tag{5.26}$$

The assertions of the theorem follow from the representation of our procedure in the form (5.22)–(5.23), which is the exact version of the proximal point algorithm associated with the maximal monotone operator T_r. As a special case of Theorem 1 of [3] (a result stated to allow for inexactness in (5.22)), the sequence $\{(V^\nu, \overline{W}^\nu)\}_{\nu=1}^\infty$ is bounded if and only if there is a pair $(V^*, \overline{W}^*)$ satisfying

$$(0,0) \in T_r(V^*, \overline{W}^*) \quad \text{with } V^* \in \mathscr{N}, \overline{W}^* \in \mathscr{M}, \tag{5.27}$$

in which case the sequence actually converges to some such pair. We must confirm that the pairs $(V^*, \overline{W}^*)$ satisfying (5.28) are precisely the pairs $(X^*, r^{-1}W^*)$ such that X^* solves the primal problem $(\mathscr{P})$ and W^* solves the dual problem $(\mathscr{D})$. We have from (5.17) that the pairs satisfying (5.27) are the ones with

$$-\overline{W}^* \in r^{-1}\partial F(X^*) + N_{\mathscr{C}}(X^*) \quad \text{for } X^* = V^*.$$

This relation can also be written as

$$-W^* \in \partial F(X^*) + N_{\mathscr{C}}(X^*) \quad \text{for } X^* = V^*$$

and is equivalent by Theorem 4.1 to X^* being optimal for $(\mathscr{P})$ and W^* being an associated multiplier. By Theorem 4.6, W^* is such a multiplier if and only if W^* solves $(\mathscr{D})$.

All that remains to the proof of Theorem 5.1 is the verification of the inequalities (5.3) and (5.4). In the notation (5.24) these take the form

$$\|Z^{\nu+1} - Z^*\| \leqslant \|Z^\nu - Z^*\| \quad \text{for all } \nu, \tag{5.28}$$

with strict inequality unless $Z^\nu = Z^*$,

$$\|Z^{\nu+1} - Z^\nu\| \leqslant \|Z^\nu - Z^{\nu-1}\| \quad \text{for all } \nu. \tag{5.29}$$

Noting that the optimality relation (5.27) can be written as $Z^* = M_r(Z^*)$, whereas $Z^{\nu+1} = M_r(Z^\nu)$ for every ν, we can get both these inequalities from (5.26), as we now show.

In the case of (5.28) we take $Z = Z^*$ and $Z' = Z^\nu$ to turn (5.26) into

$$\|Z^{\nu+1} - Z^*\|^2 + \|(Z^\nu - Z^{\nu+1}) - 0\|^2 \leqslant \|Z^\nu - Z^*\|^2.$$

This yields the inequality in (5.28) and the information that the inequality is strict unless $\|Z^\nu - Z^{\nu+1}\| = 0$. Of course $\|Z^\nu - Z^{\nu+1}\| = 0$ if and only if $Z^\nu = M_r(Z^\nu)$, in which event the sequence generated by our procedure must forever more remain fixed at Z^ν, and Z^ν must of course coincide with Z^*. The full assertion of (5.28) is thereby justified.

In the case of (5.29) we apply (5.26) with $Z = Z^{\nu-1}$ and $Z' = Z^{\nu}$. This gives

$$\|Z^{\nu+1} - Z^{\nu}\|^2 + \|(Z^{\nu} - Z^{\nu+1}) - (Z^{\nu-1} - Z^{\nu})\|^2 \leqslant \|Z^{\nu} - Z^{\nu-1}\|^2$$

and in particular proves (5.29). □

A stronger result about the rate of convergence will be obtained now for the linear-quadratic case of problem $(\mathscr{P})$.

THEOREM 5.2. *Consider the algorithm in the linear-quadratic case with exact minimization, and suppose that* $(\mathscr{P})$ *and* $(\mathscr{D})$ *have unique optimal solutions. Then the convergence in Theorem* 5.1 *is at a linear rate from the start: there is a value* $\theta_r \in [0, 1)$ *such that in every iteration* $\nu = 0, 1, 2, \ldots$ *one has*

$$(5.30) \qquad \|(\hat{X}^{\nu+1}, W^{\nu+1}) - (X^*, W^*)\|_r \leqslant \theta_r \|(\hat{X}^{\nu}, W^{\nu}) - (X^*, W^*)\|_r.$$

PROOF. To demonstrate this we return to the context of the proximal point algorithm in the proof of Theorem 5.1. The results will follow from Theorem 2 of [3] through verifying a Lipschitz property for T_r^{-1} at $(0, 0)$.

Recall that in the linear-quadratic case F is a quadratic (possibly affine) function and $\mathscr{E}$ is a polyhedral convex set. In this case the multifunction $X \mapsto \partial F(X)$ reduces to an affine transformation. At the same time the multifunction $N_{\mathscr{E}}: X \mapsto N_{\mathscr{E}}(X)$ is *polyhedral* in a sense defined and demonstrated by Robinson [18]: the graph of $N_{\mathscr{E}}$ is the union of a finite collection of polyhedral convex subsets of $\mathscr{E} \times \mathscr{E}$. The multifunction $r^{-1}\partial F + N_{\mathscr{E}}$ is then polyhedral too, and hence so is the multifunction $\partial \varphi_r$ for φ_r in (5.11), because of (5.16). The graph of T_r differs from that of $\partial \varphi_r$ only by some changes in sign and shifts in the roles of various components; this is seen in (5.13). Therefore T_r is polyhedral, and it follows now that T_r^{-1} is polyhedral.

We have seen that (5.27) characterizes the optimal solutions X^* to $(\mathscr{P})$ and W^* to $(\mathscr{D})$: For each such pair we get $(V^*, \overline{W}^*)$ satisfying (5.27) by taking $V^* = X^*$, $\overline{W}^* = r^{-1}W^*$, and conversely. Our uniqueness assumption implies therefore that $T_r^{-1}(0, 0)$ is a singleton. Of course T_r^{-1} is at the same time a maximal monotone operator, because T_r is. Any maximal monotone operator has the property of being single-valued (in fact differentiable) almost everywhere on the interior of its effective domain (the set of points where it is nonempty-valued); see Theorem 1.3 of Mignot [19]. Furthermore, the interior of the effective domain is characterized as the set of points where the operator is nonempty-compact-valued; Rockafellar [20]. Since $T_r^{-1}(0, 0)$ is a singleton (which in particular is a nonempty compact set), we may conclude from these facts that T_r^{-1} is single-valued almost everywhere in a neighborhood of $(0, 0)$. But the graph of T_r^{-1} is the union of finitely many polyhedral convex sets in $\mathscr{E} \times \mathscr{E}$. The conclusion must be drawn that T_r^{-1} is actually single-valued on an entire neighborhood of $(0, 0)$ and indeed must be *piecewise affine* there. In particular T_r^{-1} is Lipschitz continuous at $(0, 0)$.

The Lipschitz property guarantees by Theorem 2 of Rockafellar [3] that a value $\theta_r \in [0, 1)$ exists for which the desired inequality (5.30) holds when ν is sufficiently large. The strict inequality in property (5.3) in Theorem 5.1 makes it possible for us, by raising the value of θ_r somewhat if necessary, to get the inequality to hold for all ν. □

The modulus of convergence in Theorem 5.2 depends on the choice of the parameter $r > 0$, which can also influence the behavior of the algorithm in other ways. This is a matter that will require further exploration. Some preliminary insights can be gained from the saddle point representation of our algorithm that was

observed in (5.9) of the proof of Theorem 5.1. We can record that representation in a more useful form for the present as follows.

PROPOSITION 5.3. *When the algorithm is executed in the convex case with exact minimization, the basic iteration can be expressed in terms of the function.*

$$(5.31)\quad l(V,W) := \inf\{F(X) + \langle X, W\rangle | X \in \mathscr{C}, \hat{X} = V\} \quad \text{for } V \in \mathscr{N}, W \in \mathscr{M}$$

by saying that $(\hat{X}^{\nu+1}, W^{\nu+1})$ *is the unique saddle point of the convex-concave function*

$$(5.32)\qquad l(V,W) + \tfrac{r}{2}\|V - \hat{X}^{\nu}\|^2 - \tfrac{1}{2r}\|W - W^{\nu}\|^2$$

with respect to minimizing over $V \in \mathscr{N}$ *and maximizing over* $W \in \mathscr{M}$.

PROOF. This differs only in notation from the version of (5.9) that was established in the proof of Theorem 5.1. □

The formulation of our algorithm in terms of Proposition 5.3 reveals a trade-off which must be respected in choosing r. A low value of r is likely to encourage progress in the primal sequence $\{\hat{X}^{\nu}\}$, but it could hinder progress in the dual sequence $\{W^{\nu}\}$. A high value of r may be expected to have the opposite effects. The ultimate consequences for the numerical behavior of the algorithm will have to be seen in practice, but a deeper study of convergence properties of the underlying proximal point algorithm could also lead to a better understanding of this situation. We cannot pursue the matter further in the present paper.

We now take up the question of how the algorithm may be realized with only approximate minimization in the subproblems $(\mathscr{P}_s^{\nu})$. As see in Proposition 4.4, the condition that in the convex case is necessary and sufficient for the point $X^{\nu+1}(s) \in C_s$ to be the exact (unique) optimal solution to $(\mathscr{P}_s^{\nu})$ is (4.19), which can be written as

$$(5.33)\qquad 0 \in \partial f_s^{\nu}\big(X^{\nu+1}(s)\big) + N_{C_s}\big(X^{\nu+1}(s)\big)$$

for the function

$$(5.34)\qquad f_s^{\nu}(x) = f_s(x) + x \cdot W^{\nu}(s) + \tfrac{1}{2}r|x - \hat{X}^{\nu}(s)|^2.$$

Problem $(\mathscr{P}_s^{\nu})$ consists, of course, minimizing f_s^{ν} over C_s.

The criterion we shall use for approximate minimization in $(\mathscr{P}_s^{\nu})$ is, in place of (5.33), the inequality

$$(5.35)\qquad \operatorname{dist}(0, D_s^{\nu}) \leqslant (1-\epsilon)^{\nu+1}\min\{1, |X^{\nu+1}(s) - \hat{X}^{\nu}(s)|\} \quad \text{with}$$

$$D_s^{\nu} := \partial f_s^{\nu}\big(X^{\nu+1}(s)\big) + N_{C_s}\big(X^{\nu+1}(s)\big),$$

where $\mu > 0$ and $0 < \epsilon < 1$ (fixed values). The left side of this inequality involves the Euclidean distance of the set $\partial f_s^{\nu}(X^{\nu+1}(s)) + N_{C_s}(X^{\nu+1}(s))$ from the origin of $\mathbb{R}^n$, a distance we known to be 0 in the case of exact minimization in (5.33). Note that unless $\hat{X}^{\nu}(s)$ itself is the unique solution to $(\mathscr{P}_s^{\nu})$, a special possibility that could be tested for at the outset of any procedure for solving $(\mathscr{P}_s^{\nu})$, the right side of the inequality in (5.35) will tend to some positive value. Then there will be leeway for determining in finitely many iterations an $X^{\nu+1}(s)$ for which the condition is met.

This form of stopping criterion is easier to implement than might be guessed. The exact condition in (5.33) corresponds in a standard linear or smooth nonlinear

programming formulation of $(\mathscr{P}_s^\nu)$ to the Kuhn-Tucker conditions. Then the set $\partial f_s^\nu(x)$ reduces just to the gradient vector of $\nabla f_s^\nu(x)$, which is $\nabla f_s(x) + W^\nu(s) + r(x - \hat{X}^\nu(s))$, and $N_{C_s}(x)$ is a cone generated by the gradients of the active constraint vectors at x in the constraint representation for C_s, as was explained at the beginning of §4. The fulfillment of the Kuhn-Tucker conditions at $X^{\nu+1}(s)$ means then that a vector

$$Y^\nu(s) \in \partial f_s^\nu\big(X^{\nu+1}(s)\big) + N_{C_s}\big(X^{\nu+1}(s)\big) \tag{5.36}$$

has been determined and shown to equal 0.

In the approximate minimization we are proposing, one has (5.36) holding but the vector $Y^\nu(s)$ is not quite 0. Nonetheless

$$\operatorname{dist}\big(0, \partial f_s^\nu\big(X^{\nu+1}(s)\big)\big) + N_{C_s}\big(X^{\nu+1}(s)\big) \leqslant |Y^\nu(s)|. \tag{5.37}$$

Our stopping criterion (5.35) is satisfied if

$$|Y^\nu(s)| \leqslant \mu(1-\epsilon)^{\nu+1}\min\{1, |X^{\nu+1}(s) - \hat{X}^\nu(s)|\}, \tag{5.38}$$

and then $X^{\nu+1}(s)$ is appropriately deemed an approximate solution to $(\mathscr{P}_s^\nu)$.

THEOREM 5.4. *If the algorithm in the convex case is implemented with approximate minimization in the sense of criterion* (5.35) *for the modified scenario subproblems* $(\mathscr{P}_s^\nu)$, *then all the convergence results in Theorems* 5.1 *and* 5.2 *remain valid.*

PROOF. Going back again to the formulation of the algorithm in terms of T_r and M_r in the proof of Theorem 5.1, we aim at applying the criterion given in Rockafellar [3] for approximate implementation of the proximal point algorithm. This criterion in present notation takes the form that

$$\|(V^{\nu+1}, \overline{W}^{\nu+1}) - M_r(V^\nu, \overline{W}^\nu)\| \leqslant \epsilon_\nu \quad \text{with } \epsilon_\nu > 0,\ \sum_{\nu=1}^{\infty} \epsilon_\nu < \infty, \tag{5.39}$$

$$\|(V^{\nu+1}, \overline{W}^{\nu+1}) - M_r(V^\nu, \overline{W}^\nu)\| \leqslant \delta_\nu \|(V^{\nu+1}, \overline{W}^{\nu+1}) - (V^\nu, \overline{W}^\nu)\| \tag{5.40}$$

$$\text{with } \delta_\nu > 0,\ \sum_{\nu=0}^{\infty} \delta_\nu < \infty.$$

We can subsume the two inequalities into the single, somewhat simpler conditions that

$$\|(V^{\nu+1}, \overline{W}^{\nu+1}) - M_r(V^\nu, \overline{W}^\nu)\| \tag{5.41}$$

$$< \tfrac{1}{2} r\mu(1-\epsilon)^{\nu+1}\min\{1, \|(V^{\nu+1}, \overline{W}^{\nu+1}) - (V^\nu, \overline{W}^\nu)\|\}$$

for fixed $\mu > 0$ and $\epsilon \in (0,1)$. (The factor $\frac{1}{2}r$ anticipates a relationship that will subsequently emerge.) We set out now to demonstrate that (5.41) is implied by (5.35).

To say that (5.35) is satisfied is to say that for some vector satisfying (5.36) one has (5.38). (The set on the right side of (5.36) is nonempty and closed.) Let us suppose we have such a vector $Y^\nu(s)$ for every scenario $s \in S$ and observe that then

$$Y^\nu \in \partial F^\nu(X^{\nu+1}) + N_{\mathscr{C}}(X^{\nu+1}) \tag{5.42}$$

in the notation (5.20), or what is the same thing,

$$Y^\nu \in \partial F(X^{\nu+1}) + N_{\mathscr{C}}(X^{\nu 1}) + W^\nu + r(X^{\nu+1} - \hat{X}^\nu).$$

We can relate this expression more closely to the formula in (5.17) for T_r by writing it as

$$r^{-1}Y^\nu - \overline{W}^\nu - (X^{\nu+1} - \hat{X}^\nu) \in r^{-1}\partial F(X^{\nu+1}) + N_{\mathscr{C}}(X^{\nu+1}), \tag{5.43}$$

where $\overline{W}^\nu = r^{-1}W^\nu$ as in the notation (5.6). The vector on the left in (5.43) can be decomposed into a component

$$J\big[r^{-1}Y^\nu - \overline{W}^\nu - (X^{\nu+1} - \hat{X}^\nu)\big] = r^{-1}JY^\nu - V^{\nu+1} + V^\nu \quad \text{in } \mathscr{N} \tag{5.44}$$

and a component

$$\begin{aligned} K\big[r^{-1}Y^\nu - \overline{W}^\nu - (X^{\nu+1} - \hat{X}^\nu)\big] &= r^{-1}KY^\nu - \overline{W}^\nu - KX^{\nu+1} \\ &= r^{-1}KY^\nu - \overline{W}^{\nu+1} \quad \text{in } \mathscr{M}. \end{aligned} \tag{5.45}$$

Here we are also using the notation (5.5) and the fact that $\overline{W}^{\nu+1} = \overline{W}^\nu + KX^{\nu+1}$. Decomposing $X^{\nu+1}$ itself by

$$X^{\nu+1} = JX^{\nu+1} + KX^{\nu+1} = V^{\nu+1} + (\overline{W}^{\nu+1} - \overline{W}^\nu),$$

we can now invoke (5.17) and write (5.43) in terms of T_r as

$$(r^{-1}JY^\nu - V^{\nu+1} + V^\nu, -\overline{W}^{\nu+1} + \overline{W}^\nu) \in T_r(V^{\nu+1}, \overline{W}^{\nu+1} - r^{-1}KY^\nu).$$

This in turn becomes

$$\begin{aligned} &(V^\nu + r^{-1}JY^\nu, \overline{W}^\nu - r^{-1}KY^\nu) - (V^{\nu+1}, \overline{W}^{\nu+1} - r^{-1}KY^\nu) \\ &\quad \in T_r(V^{\nu+1}, \overline{W}^{\nu+1} - r^{-1}KY^\nu), \end{aligned}$$

which for the mapping $M_r = (I + T_r)^{-1}$ can be expressed as

$$(V^{\nu+1}, \overline{W}^{\nu+1}) - (0, r^{-1}KY^\nu) = M_r(V^\nu + r^{-1}JY^\nu, \overline{W}^\nu - r^{-1}KY^\nu). \tag{5.46}$$

The nonexpansive property of M_r in (5.25) lets us estimate

$$\begin{aligned} &\|M_r(V^\nu + r^{-1}JY^\nu, \overline{W}^\nu - r^{-1}KY^\nu) - M_r(V^\nu, \overline{W}^\nu)\| \\ &\quad \leqslant \|(r^{-1}JY^\nu, -r^{-1}KY^\nu)\| = r^{-1}\|Y^\nu\| \end{aligned} \tag{5.47}$$

(where we use the fact that JY^ν and KY^ν are the components of Y^ν with respect to the orthogonal decomposition $\mathscr{E} = \mathscr{N} \oplus \mathscr{M}$). But (5.46) gives

$$\begin{aligned} &\|(V^{\nu+1}, \overline{W}^{\nu+1}) - M_r(V^\nu + r^{-1}JY^\nu, \overline{W}^\nu - r^{-1}KY^\nu)\| \\ &\quad \leqslant \|(0, -r^{-1}KY^\nu)\| \leqslant r^{-1}\|Y^\nu\|. \end{aligned} \tag{5.48}$$

In combining (5.47) with (5.48) we achieve

$$\|(V^{\nu+1}, \overline{W}^{\nu+1}) - M_r(V^\nu, \overline{W}^\nu)\| \leqslant 2r^{-1}\|Y^\nu\|. \tag{5.49}$$

We originally chose Y^ν to have (5.38) satisfied for every $s \in S$. This condition gives

$$|Y^\nu(s)|^2 \leqslant \left[\mu(1-\epsilon)^{\nu+1}\right]^2, \qquad |Y^\nu(s)|^2 \leqslant \left[\mu(1-\epsilon)^{\nu+1}\right]^2 |X^{\nu+1}(s) - \hat{X}^\nu(s)|^2.$$

Calculating $\|Y^\nu\|^2 = E\{|Y^\nu(s)|^2\}$, we therefore have

$$\|Y^\nu\| \leqslant \mu(1-\epsilon)^{\nu+1}\min\{1, \|X^{\nu+1} - \hat{X}^\nu\|\}. \tag{5.50}$$

This brings us to the finish: (5.49) plus (5.50) implies (5.41). Thus the proposed stopping criterion (5.35) does ensure that (5.41) will be respected at every iteration, and the convergence facts in Theorems 1 and 2 of [3] can be applied to get the results claimed. □

6. Convergence in the nonconvex case. Outside of the convex case of problem $(\mathscr{P})$ we really have no substantial results of convergence of the algorithm along the lines of the ones in §5, at least at present. This territory has not been well investigated, however. We do think there are possibilities for using the algorithm effectively in the nonconvex case as well. So that the reader is not left with too narrow an impression, we wish to provide in this section some evidence supporting that opinion.

In the nonconvex case it is probably futile to count on being able to solve $(\mathscr{P})$ globally. The same could well to be true for the subproblems $(\mathscr{P}_s^\nu)$. The analog of "exact minimization" (we do not try to deal with "approximate minimization" here) is the calculation of $X^{\nu+1}(s)$ as a *locally* optimal solution to $(\mathscr{P}_s^\nu)$ at each iteration. A slippery quality of local minimization, however, is the variability of what "local" might mean from one iteration to the next.

Let us speak of a *δ-locally optimal* solution as an optimal solution relative to a δ-neighborhood of the point in question. While the calculation of a globally optimal solution may be out of the question, an idea not very farfetched is that the technology of optimization will allow us to calculate for fixed (possibly small) $\delta > 0$ specified in advance a δ-locally optimal solution $X^{\nu+1}(s)$ to $(\mathscr{P}_s^\nu)$. Properties such as Lipschitz continuity of the objective and constraint functions could support this capability.

Proceeding anyway on such a basis, we are able to show that the algorithm, if it does converge to something (a big assumption?), produces in the nonconvex case of $(\mathscr{P})$ about as good a "solution" as could be hoped for.

THEOREM 6.1. *Suppose that the algorithm is implemented in the nonconvex case in such a way that in each iteration the calculated vector $X^{\nu+1}(s)$ is δ-locally optimal for $(\mathscr{P}_s^\nu)$, where $\delta > 0$ is fixed. If the generated sequences $\{X^\nu\}_{\nu=1}^\infty$ and $\{W^\nu\}_{\nu=1}^\infty$ do converge to elements X^* and W^* respectively, then X^* and W^* satisfy the optimality conditions in Theorem* 4.1. *In this sense, X^* is a stationary point for $(\mathscr{P})$.*

Under these circumstances, in fact, X^* is a locally optimal solution to, and W^* a corresponding multiplier vector for, the problem $(\tilde{\mathscr{P}})$ obtained from $(\mathscr{P})$ by replacing each function f_s by

$$\tilde{f}_s(x) = f_s(x) + \tfrac{1}{2}r|x - X^*(s)|^2. \tag{6.1}$$

PROOF. The assertion about problem $(\tilde{\mathscr{P}})$ implies the one about X^* being a stationary point for $(\mathscr{P})$, because $\partial \tilde{f}_s(x) = \partial f_s(x) + r(x - X^*(s))$ (see Corollary 2.4.2 of [1]) and consequently

$$\partial \hat{f}_s(X^*(s)) = \partial f_s(X^*(s)). \tag{6.2}$$

The only distinction between $(\tilde{\mathscr{P}})$ and $(\mathscr{P})$ in the optimality conditions of Theorem 4.1 would lie in a possible discrepancy between the two sets shown in (6.2) to coincide. We may concentrate therefore on the assertion about $(\tilde{\mathscr{P}})$.

The condition that $X^{\nu+1}(s)$ is δ-locally optimal in $(\mathscr{P}_s^\nu)$ for every $s \in S$ implies that $X^{\nu+1}$ is locally optimal in the problem $(\mathscr{P}^\nu)$ where

$$F^\nu(X) = F(X) + \langle X, W^\nu \rangle + \tfrac{1}{2} r \|X - \hat{X}^\nu\|^2 \tag{6.3}$$

is minimized over $\mathscr{C}$. Indeed, $X^{\nu+1}$ is δ'-locally optimal in $(\mathscr{P}^\nu)$, where

$$\delta' = \delta \min_{s \in S} p_s^{1/2}, \tag{6.4}$$

because

$$\|X - X^{\nu+1}\|^2 \leqslant (\delta')^2 \Rightarrow \sum_{s \in S} p_s |X(s) - X^{\nu+1}(s)|^2 \leqslant (\delta')^2$$

$$\Rightarrow |X(s) - X^{\nu+1}(s)| \leqslant (\delta')^2 / p_s \quad \text{for all } s$$

and consequently

$$\|X - X^{\nu+1}\| \leqslant \delta' \Rightarrow |X(s) - x^{\nu+1}(s)| \leqslant \delta \quad \text{for all } s. \tag{6.5}$$

Working now with our assumption that $X^\nu \to X^*$ and $W^\nu \to W^*$ we see, because $F^\nu(X) \geqslant F^\nu(X^{\nu+1})$ when $X \in \mathscr{C}$ and $\|X - X^{\nu+1}\| \leqslant \delta'$, that in the limit one has

$$F(X) + \langle X, W^* \rangle + \tfrac{1}{2} r \|X - X'\|^2 \geqslant F(X^*) + \langle X^*, W^* \rangle + \tfrac{1}{2} r \|X^* - X^*\|^2 \text{ when } X \in \mathscr{C}, \|X - X^*\| \leqslant \delta', \tag{6.6}$$

or in other words, X^* is δ'-locally optimal for the problem

$$(\mathscr{P}^*) \qquad \text{minimize } F(X) + \langle X, W^* \rangle + \tfrac{1}{2} r \|X - X^*\|^2 \text{ over } X \in \mathscr{C}.$$

The convergence of W^ν to W^* implies, because $W^\nu \in \mathscr{M}$ and $W^{\nu+1} - W^\nu = rKX^{\nu+1}$, that $W^* \in \mathscr{M}$ and $KX^\nu \to 0$. The latter means $X^\nu - \hat{X}^\nu \to 0$. Hence $X^* - \hat{X}^* = 0$, i.e., $X^* \in \mathscr{N}$. The fact that X^* is δ'-locally optimal in $(\mathscr{P}^*)$ gives us then in particular, because $\langle X, W^* \rangle = 0$ when $X \in \mathscr{N}$, that X^* is δ'-locally optimal for minimizing the expression

$$F(X) + \tfrac{1}{2} r \|X - X^*\|^2 = E\{\tilde{f}_s(X(s))\}$$

over $X \in \mathscr{C} \cap \mathscr{N}$. This is all we needed to prove. □

References

[1] T. Higgins and H. Jenkins-Smith, "Analysis of the Economic Effect of the Alaskan Oil Export Ban," *J. Oper. Res.* **33** (1985), 1173–1202.

[2] B. Avi-Itzhak, T. J. Connelly, G. B. Dantzig, P. H. McAllister and W. D. Winkler, A Dynamic Equilibrium Model Model for Energy-Economic Planning, Tech. Report SOL 82-2, Stanford University, Stanford, CA, 1982.

[3] R. T. Rockafellar, "Monotone Operators and the Proximal Point Algorithm," *SIAM J. Control Optim.* **14** (1976), 877–898.
[4] J. E. Spingarn, "Applications of the Method of Partial Inverses to Convex Programming," *Math. Prog.* **32** (1985), 199–223.
[5] D. P. Bertsekas, *Constrained Optimzation and Lagrange Multiplier Methods*, Academic Press, New York, 1982.
[6] R. T. Rockafellar, "Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming," *Math. Oper. Res.* **1** (1976), 97–116.
[7] J. E. Spingarn, "Partial Inverse of a Montone Operator," *Appl. Math. Optim.* **10** (1983), 247–265.
[8] P.-L. Lions and B. Mercier, "Splitting Algorithms for the Sum of Two Nonlinear Operators," *SIAM J. Numer. Anal.* **16** (1979), 964–979.
[9] D. Gabay and B. Mercier, "A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximations," *Computers and Math. Appl.* **2** (1976), 17–40.
[10] M. Fortin and R. Glowinski, "On Decomposition-Coordination Methods using an Augmented Lagrangian," in *Augmented Lagrangian Methods: Applications to the Solution of Boundary Problems*, M. Fortin and R. Glowinski, Eds., North-Holland, Amsterdam, 1983.
[11] J. Eckstein, "The Lions-Mercier Splitting Algorithm and the Alternating Direction Method are Instances of the Proximal Point Algorithm," technical report LIDS-P-1769, Lab. for Info. and Decision systems, MIT, Cambridge, MA, 1988.
[12] S.-P. Han, "A Decomposition Method and its Applications to Convex Programming," *Math. Oper. Res.* **14** (1989), 2337–248.
[13] R. T. Rockafellar and R. J.-B Wets, "Generalized Linear-Quadratic Problems of Deterministic and Stochastic Optimimal Control in Discrete Time, *SIAM J. Control Optim.* (to appear).
[14] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
[15] R. T. Rockafellar, "Extensions of Subgradient Calculus with Applications to Optimization," *Nonlinear Anal. Theor. Meth. Appl.* **9** (1985), 665–698.
[16] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, NJ, 1970.
[17] G. J. Minty, "Montone (Nonlinear) Operators in Hilbert Space, *Duke Math. J.* **29** (1962), 341–346.
[18] S. M. Robinson, "Generalized Equations and their Solutions, Part I: Basic Theory," *Math. Prog. Studies* **10** (1979), 128–141.
[19] F. Mignot, "Contrôle dans les Inéquations Variationelles Elliptiques," *J. Functional Anal.* **22** (1976), 130–185.
[20] R. T. Rockafellar, "Local Boundedness of Nonlinear, Monotone Operators," *Michigan Math. J.* **16** (1969), 397–407.

ROCKAFELLAR: DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WASHINGTON GN-50, SEATTLE, WASHINGTON 98195
WETS: DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, DAVIS, CALIFORNIA 95616