LARGE-SCALE EXTENDED LINEAR-QUADRATIC PROGRAMMING AND MULTISTAGE OPTIMIZATION

R.T.Rockafellar

Abstract. Optimization problems in discrete time can be modeled more flexibly by extended linear-quadratic programming than by traditional linear or quadratic programming, because penalties and other expressions that may substitute for constraints can readily be incorporated and dualized. At the same time, dynamics can be written with state vectors as in dynamic programming and optimal control. This suggests new primal-dual approaches to solving multistage problems. The special setting for such numerical methods is described. New results are presented on the calculation of gradients of the primal and dual objective functions and on the convergence effects of strict quadratic regularization.

* Dept. of Mathematics, University of Washington GN-50, Seattle, WA 98195. This work was supported by AFOSR grant 89-0081 and NSF grant DMS 881-9586.

1. Introduction.

Large-scale problems in numerical optimization often arise from multistage models where decisions must be taken over a finite sequence of time periods. Such models may be deterministic or stochastic, and they may involve a process that naturally takes place in discrete time or the discretization of a process in continuous time. The computational challenges can be quite serious because of dimensionality, so the need to take advantage of specific problem structure has long been recognized. Different tactics can be followed in representing such structure, however, and this may be as important a key to eventual computational success as the refinement of standard algorithms for more general problems.

In work having its roots in the traditions of mathematical programming, the principal theme has been that of setting up problems as high-dimensional linear or quadratic programs and looking for helpful patterns of sparsity in the resulting matrices—staircase structure, and so forth. In work driven by the concepts and applications in areas like dynamic programming and optimal control, however, the emphasis is more often placed on a system of dynamical equations and the utilization of well known optimality conditions like the "maximum principle." In addition to the decision variables, termed "controls," there are auxiliary "state" variables around which the numerical procedures typically revolve.

Neither of these lines of thinking appears at present to derive full benefit from the other. In control, not enough attention has been paid to developing a first-level methodology for linear and quadratic types of problems in which inequality constraints can be handled alongside of equations, and in which the dual variables that emerge can be exploited on the basis of convexity. In mathematical programming, on the other hand, the modeling framework developed years ago for small-scale problems has perhaps been too readily accepted as appropriate for large-scale applications. The paradigm of just identifying a linear or quadratic objective function and writing down a list of exact linear constraints, then invoking a computational scheme that adapts to their special properties, may sometimes be unsuitable and unduly limiting.

The purpose of this paper is to describe an approach to multistage optimization that tries to bridge between the two camps, and to indicate the potential for numerical development that is thereby opened up. Problems are formulated in terms of "extended linear-quadratic programming," which differs from ordinary linear and quadratic programming mainly in allowing more flexible treatment of putative constraints, for instance in the direct incorporation of penalty terms. The problem models we draw on are ones developed recently in Rockafellar [1], [2], [3], [4], and Rockafellar and Wets [5]. We aim especially at explaining how the use of piecewise linear-quadratic penalty expressions along with state variables, generated from dynamical equations, alters the computational environment away from the familiar one in large-scale quadratic programming and linear complementarity but at the same time paves the way for new techniques like envelope methods [6], [7], and splitting methods [8]–[17].

2. Monitoring functions and extended linear-quadratic programming.

A problem of quadratic programming, as customarily defined, consists of minimizing a (linear or) quadratic convex objective function subject to a system of linear constraints. By introducing Lagrange multipliers for some of the constraints, one obtains a Lagrangian function for the problem and a saddle point characterization of optimality, which leads in turn to a dual problem. Solution methods may be "primal," which is to say roughly that they are molded by ideas of direct descent in the objective subject to the constraints, "dual," which means they involve ascent in an associated dual problem, or "primal-dual," which refers to a focus on solving for the primal and dual variables jointly through the Lagrangian saddle point condition.

Common in the case of primal methods have been "active set strategies" for identifying in a tentative way, to be updated as the procedure goes on, the set of linear inequality constraints likely to be active at the solution, and then treating these as equations. Such strategies often do not fit well with large-scale applications, however, because the dual dimension may be very high along with the primal dimension. Especially prominent among primal-dual methods have been those based on formulation of the saddle point condition as a problem in *linear complementarity*. We refer to Pang [18], Lin and Pang [19], for surveys of techniques in quadratic programming and linear complementarity. Recent developments based on Karmarkar's algorithm in linear programming can be found in Ye and Tse [20], Monteiro and Adler [21], and Goldfarb and Liu [22].

A problem of *extended linear-quadratic programming* can be described abstractly as consisting of the minimization of a piecewise linear-quadratic convex function subject to a system of linear constraints. A function is *piecewise linear-quadratic* if its domain can be expressed in principle (if not necessarily as a matter of convenience) by a union of polyhedral sets, on each of which the function is given by a linear or quadratic formula. This description of extended linear-quadratic programming obscures a fundamental difference in modeling concept, however, because the objective in this case may arise from some kind of mixing of what would be taken as the objective and constraints in a traditional approach.

In typical applications of optimization, the values of a variety of functions are of interest. The standard paradigm dictates that a single one of these functions must be chosen for minimization while the others are merely kept within certain bounds. The reality, of course, is that the choice of what should be the objective function versus what should be the constraint functions may be difficult because of trade-offs. The bounds to be used in writing down constraints may pose difficulties as well. In many situations black and white bounds on a function are not given, but just a desirable range of values that gradually merges into an undesirable or risky range. This is all the more true when the underlying application explicitly involves uncertainty and hedging. An insistence on exact constraints, except of a definitional variety (for instance, the natural nonnegativity of certain variables), makes little sense in such cases and can lead to poor models and unreliable "solutions."

There is nothing new in these observations, but their importance has grown with the scope of the problems being tackled with optimization methodology. Developments in stochastic programming as well as optimal control have made the deficiencies of mathematical models based on classical linear and quadratic programming quite apparent. Extended linear-quadratic programming, although not a panacea, offers a much more flexible approach to modeling which nonetheless retains the simple algebraic character of classical linear and quadratic programming *when viewed in primal-dual terms*. The function being minimized in the primal problem takes a more complicated form, but the Lagrangian saddle point problem characterizing optimality turns out still to be one that concerns just a quadratic function on a product of polyhedral sets.

In setting up a model in extended linear-quadratic programming, the notion of a "monitoring function" is central in obtaining from the given functions, either singly or in groups, the terms that make up the composite objective. Monitoring functions include penalty functions as a special case. For a start, let us consider a single function $h(u)$ from among those given in some application, where $u \in \mathbb{R}^k$ is the vector of "decision variables." The point of view is that the values of this function must be incorporated in the model, but not necessarily according to the standard paradigm: we are at a stage of formulation where no firm choice has yet been made about what should be a hard constraint or a contribution to the objective, at least as far as $h(u)$ is concerned.

In general we can think of placing a term $\rho(h(u))$ in the objective function being constructed for minimization—for some choice of $\rho : \mathbb{R} \to \overline{\mathbb{R}}$. Then ρ is a *monitoring function* for $h(u)$. For instance, if we wish the problem to simply have the minimization of $h(u)$ as its objective, we can take $\rho(s) = s$. If instead we wish to minimize a weighted sum of expressions, of which $h(u)$ is just one, we can take $\rho(s) = \lambda s$ for some

$\lambda > 0$. On the other hand, if our aim turns out to be that of enforcing the inequality $h(u) \leq 0$ as a hard constraint, this corresponds to taking ρ to be the indicator of $\mathbb{R}_-$, i.e., $\rho(s) = 0$ when $s \leq 0$ but $\rho(s) = \infty$ when $s > 0$. In similar fashion, the equation $h(u) = 0$ corresponds to defining $\rho(0) = 0$, but $\rho(s) = \infty$ for all $s \neq 0$.

Penalty representations of a constraint like $h(u) \leq 0$ fit this picture very easily. A standard quadratic penalty term would arise in taking $\rho(s) = 0$ when $s \leq 0$ but $\rho(s) = \frac{1}{2}\mu s^2$ when $s > 0$ (where the penalty parameter μ is positive), while a so-called linear penalty would take the form $\rho(s) = 0$ when $s \leq 0$ but $\rho(s) = \mu s$ when $s > 0$. As is well known, a linear penalty term with μ high enough may suffice to represent a hard constraint $h(u) \leq 0$ exactly. Exact penalty expressions can also be obtained through augmented Lagrangian terms of the kind where $\rho(s) = \lambda s + \frac{1}{2}\mu s^2$ when $s \geq -\lambda/\mu$, but $\rho(s) = -\lambda^2/2\mu$ when $s \leq -\lambda/\mu$ (see [23]). Further, one can make use of mixed penalties which start out quadratic but end up linear, as have turned out to be desirable in stochastic programming, cf. Rockafellar and Wets [24].

In all these examples the monitoring function ρ is convex and piecewise linear-quadratic. This continues to be observed when several functions $h_i(u)$ are monitored together by a term $\rho(h(u))$ where $h(u)$ is a vector $(h_1(u), \ldots, h_l(u))$. For instance, an expression of the form $\max_i h_i(u)$ in the objective corresponds to the piecewise linear function $\rho(s) = \max_i s_i$, where $s = (s_1, \ldots, s_l)$. The hard constraint $h(u) \leq 0$ in vector terms corresponds to ρ being the indicator of the nonpositive orthant of $\mathbb{R}^l$ (which is a polyhedral set), and so forth. Many other illustrations could be given in which the monitoring function acts like a weighted sum or max within the main region in which the vector $h(u)$ is desired to lie, but takes on penalty characteristics outside.

A dual form of representation turns out to characterize a large class of monitoring functions which includes the many examples indicated and is especially convenient to work with. A function in this class, on $\mathbb{R}^l$, is fixed by choosing a nonempty polyhedral set $V \subset \mathbb{R}^l$ (possibly all of $\mathbb{R}^l$) and a positive *semi*definite, symmetric matrix $Q \in \mathbb{R}^{l \times l}$ (which could be the zero matrix):

$$\rho_{VQ}(s) = \max_{v \in V}\{s \cdot v - \tfrac{1}{2} v \cdot Qv\}.$$

In the one-dimensional case ($l = 1$), the set V is just a closed interval $[v^-, v^+]$ (perhaps unbounded) and Q is specified by a single number $\beta \geq 0$, so that

$$\rho_{VQ}(s) = \max_{v^- \leq v \leq v^+}\{sv - \tfrac{1}{2}\beta s^2\}.$$

The maximum can then be calculated explicitly to obtain a more direct formula. In the *box-diagonal* case, by which we mean the case where V is a box (a product of closed intervals $[v_i^-, v_i^+]$) and Q is a diagonal matrix ($\mathrm{diag}\{\beta_1, \ldots, \beta_l\}$ with $\beta_i \geq 0$), the calculation can be carried out similarly, and $\rho_{VQ}(s)$ is a sum of separate terms in the components s_i of s. On the other hand, the example of $\rho(s) = \max_i s_i$ corresponds to V being the unit simplex in $\mathbb{R}^l$ (the set of vectors $(s_1, \ldots, s_l) \geq (0, \ldots, 0)$ with $\sum_{i=1}^l s_i = 1$) and Q being the zero matrix. In Rockafellar and Wets [25] it is demonstrated that even the "recourse cost function" in a stochastic linear or quadratic programming problem can be written as $\rho_{VQ}(h(u))$ for some choice of V, Q, and a linear (or affine) mapping h.

Although composite objectives involving convex, possibly nonsmooth monitoring functions has been explored by many researchers in nonlinear programming in recent years, the fact that most of the monitoring expressions of interest fall in the ρ_{VQ} class has not been recognized. The properties of this class make possible a treatment with a cleaner depiction of the role of generalized Lagrange multipliers in the form of dual variables.

These considerations lead to the adoption of the following *standard form* for a problem in extended linear-quadratic programming:

$$(\mathcal{P}_0) \qquad \text{minimize } p \cdot u + \tfrac{1}{2} u \cdot Pu + \rho_{VQ}(q - Ru) \text{ over } u \in U,$$

where the sets $U \subset \mathbb{R}^k$ and $V \subset \mathbb{R}^l$ are polyhedral and the matrices $P \in \mathbb{R}^{k \times k}$ and $Q \in \mathbb{R}^{l \times l}$ are symmetric and positive *semi*definite. In this, the set U expresses fixed linear constraints not handled through "monitoring." The use of this form is discussed more fully in [1, §3]. While in many situations it would

be right to think of the terms $p{\cdot}u + \frac{1}{2}u{\cdot}Pu$ as the basic objective function, modified by a penalty function applied to the vector $q - Ru$, the form can also be used in other ways. For instance, in taking V to be all of $\mathbb{R}^l$ and Q to be the $l \times l$ identity matrix, one obtains $\frac{1}{2}|q - Ru|^2$ for the monitoring term. If $q = 0$, this yields $p{\cdot}u + \frac{1}{2}u{\cdot}(P + R^*R)u$ as the objective expression to be minimized, where $*$ denotes the transpose of a matrix. The problem then falls into the classical category of quadratic programming, but with the representation of additional structure that could aid in its solution.

An important advantage of the standard form is that it points specifically to the Lagrangian function to be used in characterizing optimality. The Lagrangian is

$$L(u, v) = p{\cdot}u + \tfrac{1}{2}u{\cdot}Pu + q{\cdot}v - \tfrac{1}{2}v{\cdot}Qv - v{\cdot}Ru \text{ on } U \times V. \tag{1}$$

We have demonstrated in [25] that *a vector $\bar{u}$ solves $(\mathcal{P}_0)$ if and only if there is a vector $\bar{v}$ such that the pair $(\bar{u}, \bar{v})$ furnishes a saddle point of L relative to $U \times V$.* Especially to be noted here is the fact that the function L, unlike the objective function in $(\mathcal{P}_0)$, is truly linear-quadratic—not merely in a *piecewise* sense. The Lagrangian representation of $(\mathcal{P}_0)$ thus reveals an underlying simplicity in the problem format which might not immediately be evident. It also identifies the corresponding dual problem as

$$(\mathcal{Q}_0) \qquad \text{maximize } q{\cdot}v - \tfrac{1}{2}v{\cdot}Qv - \rho_{UP}(R^*v - p) \text{ over } v \in V.$$

The objective function in $(\mathcal{Q}_0)$ equals $\inf_{u \in U} L(u, v)$, while the objective function in $(\mathcal{P}_0)$ equals $\sup_{v \in V} L(u, v)$. The solutions to $(\mathcal{Q}_0)$ are the $\bar{v}$ components of the saddle points $(\bar{u}, \bar{v})$ for the minimax problem relative to (1) and may be interpreted as the optimal multiplier vectors associated with the monitoring expression in $(\mathcal{P}_0)$. When the monitoring expression represents hard constraints through infinite penalties, they are Lagrange multiplier vectors in the usual sense.

3. Problem Models in Multistage Optimization.

With these modeling ideas, there are other ways of dealing with dynamical structure than just through the sparsity pattern in some constraint matrix. In problems involving a linear process operating in discrete time, the dynamics can generally be expressed by a system

$$x_t = A_t x_{t-1} + B_t u_t + b_t \text{ for } t = 1, \ldots, T, \text{ with } x_0 = B_0 u_0 + b_0, \tag{2}$$

where u_t is a so-called *control* vector in $\mathbb{R}^{k_t}$ and x_t is a *state* vector in $\mathbb{R}^{n_t}$. The control vectors give the true decision variables, while the state vectors may just be artificial constructs, but useful nevertheless . The state vectors stand for certain affine expressions in the control vectors as derived from (2):

$$\begin{aligned} x_1 &= A_1(B_0u_0 + b_0) + B_1u_1 + b_1, \\ x_2 &= A_2(A_1(B_0u_0 + b_0) + B_1u_1 + b_1) + B_2u_2 + b_2, \ldots \end{aligned}$$

and they could therefore be eliminated in principle from the formulation of any optimization problem with respect to $(u_0, u_1, \ldots, u_T)$. For a number of reasons, however, it seems best to keep them. Because the dimension k_t of u_t and the dimension n_t of x_t are allowed to vary with t, there is no loss of generality in taking (2) as the form for the dynamics: even if a problem may not seem to involve state vectors, they can always be introduced so as to get this type of expression. (This is a device of long standing; see [5] for details.)

The model problem we propose to consider for such a dynamical system is

$$\begin{aligned} &\text{minimize subject to (2) with } u_t \in U_t \text{ for } t = 0, 1 \ldots, T \text{ the expression} \\ &f(u_0, \ldots, u_T) := p_0{\cdot}u_0 + \tfrac{1}{2}u_0{\cdot}P_0u_0 \\ &\qquad + \sum\nolimits_{t=1}^{T} \big[p_t{\cdot}u_t + \tfrac{1}{2}u_t{\cdot}P_tu_t + \rho_{V_tQ_t}(q_t - C_tx_{t-1} - D_tu_t) - c_t{\cdot}x_{t-1}\big] \\ &\qquad + \rho_{V_{T+1}Q_{T+1}}(q_{T+1} - C_{T+1}x_T) - c_{T+1}{\cdot}x_T, \end{aligned} \tag{$\mathcal{P}$}$$

where the sets $U_t \subset \mathbb{R}^{k_t}$ and $V_t \subset \mathbb{R}^{l_t}$ are polyhedral and the matrices $P_t \in \mathbb{R}^{k_t \times k_t}$ and $Q_t \in \mathbb{R}^{l_t \times l_t}$ are positive semidefinite. (The elements V_t, Q_t, q_t, C_t and c_t are indexed from $t = 1$ to $t = T + 1$ instead of

from $t = 0$ to $t = T$ for the sake of achieving a symmetric formulation of duality, as will soon be seen.) Note that the objective is correctly written as $f(u_0, \ldots, u_T)$, because the vectors x_t stand for affine expressions in $u_0, \ldots, u_T$ determined by the dynamics (2), as already mentioned:

$$(x_0, \ldots, x_T) = X(u_0, \ldots, u_T) \text{ for an affine mapping } X : \mathbb{R}^{k_0+\cdots+k_T} \to \mathbb{R}^{n_0+\cdots+n_T}.$$

Problem $(\mathcal{P})$ was introduced in Rockafellar and Wets [5]. It was shown in that paper that $(\mathcal{P})$ is the primal problem associated with the Lagrangian $L(u, v)$ on $U \times V$, where

$$\begin{aligned} &U := U_0 \times \cdots \times U_T, \qquad u = (u_0, \ldots, u_T), \\ &V := V_1 \times \cdots \times V_{T+1}, \qquad v = (v_1, \ldots, v_{T+1}), \\ &L(u, v) := [p_0 \cdot u_0 + \tfrac{1}{2} u_0 \cdot P_0 u_0] + [q_{T+1} \cdot v_{T+1} - \tfrac{1}{2} v_{T+1} \cdot Q_{T+1} v_{T+1}] \\ &\qquad + \sum\nolimits_{t=1}^{T} \big[p_t \cdot u_t + \tfrac{1}{2} u_t \cdot P_t u_t + q_t \cdot v_t - \tfrac{1}{2} v_t \cdot Q_t v_t - v_t \cdot D_t u_t\big] - l(u, v) \end{aligned} \tag{3}$$

and the expression $l(u, v)$ is affine separately in u and v and is defined from the dynamical equations (2) by

$$l(u, v) = \sum\nolimits_{t=1}^{T+1} x_{t-1} \cdot (C_t^* v_t + c_t). \tag{4}$$

(As earlier, $*$ denotes the transpose of a matrix.)

THEOREM 1 [5]**.** *A control sequence $\bar{u} = (\bar{u}_0, \ldots, \bar{u}_T)$ solves the multistage optimization problem $(\mathcal{P})$ if and only if there is a dual control sequence $\bar{v} = (\bar{v}_1, \ldots, \bar{v}_{T+1})$ such that $(\bar{u}, \bar{v})$ is a saddle point of $L(u, v)$ relative to $U \times V$ in (3).*

The reason for calling $\bar{v}$ a "dual control sequence" is that it solves a dual problem of the same general kind. The dual dynamical system goes backward in time with the vectors $v_t \in \mathbb{R}^{l_t}$ as controls and certain vectors $y_t \in \mathbb{R}^{n_t}$ as states:

$$y_t = A_t^* y_{t+1} + C_t^* u_t + c_t \text{ for } t = T, \ldots, 1, \text{ with } y_{T+1} = C_{T+1}^* v_{T+1} + c_{T+1}. \tag{5}$$

The dual states y_t thus represent linear expressions in the vectors $v_1, \ldots, v_{T+1}$:

$$(y_1, \ldots, y_{T+1}) = Y(v_1, \ldots, v_{T+1}) \text{ for an affine mapping } Y : \mathbb{R}^{l_1+\cdots+1_{T+1}} \to \mathbb{R}^{n_1+\cdots+n_{T+1}}.$$

It turns out (see [5]) that the Lagrangian dynamical term $l(u, v)$ defined in (4) can be written equivalently by means of (5) as

$$l(u, v) = \sum\nolimits_{t=0}^{T} y_{t+1} \cdot (B_t u_t + b_t). \tag{6}$$

(Here we see some of the motivation for the indexing conventions.) From this one derives the dual problem as

$$\begin{aligned} &\text{maximize subject to (5) with } v_t \in V_t \text{ for } t = 1 \ldots, T+1 \text{ the expression} \\ &g(v_1, \ldots, v_{T+1}) := q_{T+1} \cdot v_{T+1} - \tfrac{1}{2} q_{T+1} \cdot Q_{T+1} v_{T+1} \\ &\qquad + \sum\nolimits_{t=1}^{T} \big[q_t \cdot v_t - \tfrac{1}{2} v_t \cdot Q_t v_t - \rho_{U_t P_t}(B_t^* y_{t+1} + D_t^* v_t - p_t) - b_t \cdot y_{t+1}\big] \\ &\qquad - \rho_{U_0 P_0}(B_0^* y_1 - p_0) - b_0 \cdot y_1. \end{aligned} \tag{$\mathcal{Q}$}$$

From the dynamics (5) we have of course that

$$(y_1, \ldots, y_{T+1}) = Y(v_1, \ldots, v_{T+1}) \text{ for an affine mapping } Y : \mathbb{R}^{l_1+\cdots+l_{T+1}} \to \mathbb{R}^{n_1+\cdots+n_{T+1}}.$$

Because L in (3) is (at most) quadratic in u and v, while the sets U and V are polyhedral, $(\mathcal{P})$ is in fact a problem of extended linear-quadratic programming in standard form, and so is its dual $(\mathcal{Q})$, except for a change of sign in converting to maximization. The following result then holds in consequence of the general theory of extended linear-quadratic programming in [25] (and [1]).

THEOREM 2 [5]. *Problem* $(\mathcal{P})$ *has a solution if and only if its objective* $f(u)$ *is bounded below on* U *and is finite for at least one* $u \in U$. *Then* $\min(\mathcal{P}) = \max(\mathcal{Q})$.

The saddle point condition in Theorem 1 can be expressed through the gradients of L and the normal cones

$$\begin{aligned} N_U(\bar{u}) &:= \{r \,\big|\, r\cdot(u-\bar{u}) \le 0 \text{ for all } u \in U\} \quad (\text{taken to be } \emptyset \text{ if } \bar{u} \notin U), \\ N_V(\bar{v}) &:= \{s \,\big|\, s\cdot(v-\bar{v}) \le 0 \text{ for all } v \in V\} \quad (\text{taken to be } \emptyset \text{ if } \bar{v} \notin V), \end{aligned} \tag{7}$$

as the condition that

$$-\nabla_u L(\bar{u},\bar{v}) \in N_U(\bar{u}), \qquad \nabla_v L(\bar{u},\bar{v}) \in N_V(\bar{v}). \tag{8}$$

In terms of the affine mapping M defined by

$$M(u,v) = (\nabla_u L(u,v), -\nabla_v L(u,v)) \tag{9}$$

this can be written in turn as

$$-M(\bar{u},\bar{v}) \in N_{U\times V}(\bar{u},\bar{v}), \tag{10}$$

a condition which is the same as the linear variational inequality

$$M(\bar{u},\bar{v})\cdot\big[(u,v)-(\bar{u},\bar{v})\big] \le 0 \text{ for all } (u,v) \in U \times V. \tag{11}$$

We shall not describe here the stochastic version in [5] of the multistage problem $(\mathcal{P})$, but it likewise corresponds to a quadratic convex-concave Lagrangian on a product of polyhedral sets much as in (3). The difference is just that u and v are then random vectors subjected to certain measurability constraints (expressing the state of information at any time t), and the sets U and V must be altered slightly to take this into account. An expectation must also be taken in the formula for the Lagrangian. The various data elements in the problem may be random variables along with the unknowns u and v. See [5].

4. Envelope Methods.

The formulas for the objective functions f and g in $(\mathcal{P})$ and $(\mathcal{Q})$ bring the linear-quadratic structure fully into view. They underline the fact that these functions are not necessarily smooth, because the monitoring terms may fail to be smooth. Even if f and g do have continuous first derivatives, they will generally fail to have continuous second derivatives, since $\rho_{V_tQ_t}$ and $\rho_{P_tU_t}$ fail to have such derivatives except in the very special cases. These circumstances make it important to look for new ways of utilizing the structure of f and g in terms of the primal and dual dynamics.

We shall focus on the *strictly quadratic* case of $(\mathcal{P})$ and $(\mathcal{Q})$, by which we mean the case where the matrices P_t and Q_t are *positive definite* (i.e., nonsingular). Then, as we shall demonstrate, there are special properties which strongly support the design of numerical methods. In particular, it turns out that we do have first-order smoothness of f and g in this case, although second-order smoothness is still lacking. While many problems of interest are not strictly quadratic, a basic device allows them to be approached by solving a sequence of auxiliary problems (of the same multistage form) that are strictly quadratic. This will be explained in the next section.

THEOREM 3. *In the strictly quadratic case, the objective function* f *in* $(\mathcal{P})$ *is strongly convex with continuous first derivatives, and it can be expressed by*

$$f(u) = \max_{v\in V} L(u,v) = L(u,F(u)) \;\; \textit{with} \;\; F(u) := \operatorname*{argmax}_{v\in V} L(u,v). \tag{12}$$

Likewise, the objective function g *in* $(\mathcal{Q})$ *is in this case strongly concave with continuous first derivatives, and it can be expressed by*

$$g(v) = \min_{u\in U} L(u,v) = L(G(v),v) \;\; \textit{with} \;\; G(v) := \operatorname*{argmin}_{u\in U} L(u,v). \tag{13}$$

The mappings F and G are continuous and piecewise linear, and one has

$$\nabla f(u) = \nabla_u L(u, F(u)), \qquad \nabla g(v) = \nabla_v L(G(v), v). \tag{14}$$

Proof. Formulas (12) and (13) follow right from the definitions of the monitoring functions $\rho_{V_t Q_t}$ and $\rho_{U_t P_t}$. The max and min are uniquely attained because of the assumed positive definiteness of the matrices giving the quadratic terms in (3). Indeed, $F(u)$ is the solution to a strictly quadratic programming program in v in which only the linear part of the objective depends on u, from which it follows further (by the theory of parametric quadratic programming, cf. also the proof of our next result) that the mapping F is continuous and piecewise linear; similarly for $G(v)$. The positive definiteness also yields the strong convexity of f and the strong concavity of g, as is evident from the formulas for these functions in the statement of $(\mathcal{P})$ and $(\mathcal{Q})$. The gradient formulas have been proved in a more general context of extended linear-quadratic programming in [6, Prop. 3.2]. □

The expressions for f and g in (12) and (13) are called *envelope representations*: f is the pointwise maximum of the collection of quadratic convex functions $\{L(\,\cdot\,, v)\}_{v \in V}$, whereas g is the pointwise minimum of the collection of quadratic concave functions $\{L(u, \,\cdot\,)\}_{u \in U}$. Numerical schemes that make use of these representations in terms of various iterations on the mappings F and G are *envelope methods*. A basic theory of such methods has been developed by the author in [6].

A starting idea for envelope methods is to use a current point $u^0 \in U$ in $(\mathcal{P})$ and a current point $v^0 \in V$ in $(\mathcal{Q})$ to generate finitely many other points $u^1, \ldots, u^r$ in U and $v^1, \ldots, v^r$ in V by

$$v^k = F(u^{k-1}) \text{ and } u^k = G(v^{k-1}) \text{ for } k = 1, \ldots, r.$$

These points are seen as providing envelope information about the functions f and g through the fact that the finite envelope functions

$$\begin{aligned} \tilde{f}(u) &:= \max_{v \in \tilde{V}} L(u, v) \text{ with } \tilde{V} := \operatorname{conv}\{v^0, v^1, \ldots, v^r\}, \\ \tilde{g}(v) &:= \min_{u \in \tilde{U}} L(u, v) \text{ with } \tilde{U} := \operatorname{conv}\{u^0, u^1, \ldots, u^r\}, \end{aligned}$$

satisfy

$$\begin{aligned} f(u) \geq \tilde{f}(u) \text{ for all } u, \text{ with equality for } u = u^k,\ k = 0, 1, \ldots, r-1, \\ g(v) \leq \tilde{g}(v) \text{ for all } v, \text{ with equality for } v = v^k,\ k = 0, 1, \ldots, r-1. \end{aligned} \tag{15}$$

One type of scheme based on this information has been worked out in some detail in [6] as a generalization of a method that has been successful in two-stage stochastic programming [25], [26], [27]. This calculates a new point for $(\mathcal{P})$ by minimizing $\tilde{f}(u)$ over $u \in \tilde{U}$ (instead of the true primal problem of minimizing $f(u)$ over $u \in U$) along with a new point in $(\mathcal{Q})$ by maximizing $\tilde{g}(v)$ over $v \in \tilde{V}$ (instead of maximizing $g(v)$ over $v \in V$). The calculation of these new u and v points, toward which one can then do line searches from the current primal and dual points, for instance, is equivalent to finding a saddle point of $L(u, v)$ relative to the polyhedral set $\tilde{U} \times \tilde{V}$. This can be converted to an explicit, low-dimensional problem in extended linear-quadratic programming which can be solved by way of standard codes (see [6]).

Another kind of envelope method is developed in the author's joint paper with Zhu [7]. It uses the generation scheme (15) with $r = 2$ to execute a sort of gradient projection method of simultaneous primal descent and dual ascent with restarts triggered by a certain feedback between the primal and dual processes.

In applying any envelope method to the multistage problems $(\mathcal{P})$ and $(\mathcal{Q})$, a critical issue is how to deal effectively with the mappings F and G. The next two theorems provide the answer.

THEOREM 4. *In the strictly quadratic case, the following procedure calculates $F(u)$ and then $\nabla f(u)$ for any given u:*
(a) Determine $x = (x_0, \ldots, x_T)$ from $u = (u_0, \ldots, u_T)$ via the primal dynamics (2).
(b) Calculate $\tilde{v} = (\tilde{v}_1, \ldots, \tilde{v}_{T+1}) \in V$ by

$$\begin{aligned} \tilde{v}_t &= \nabla \rho_{V_t Q_t}(q_t - C_t x_{t-1} - D_t u_t) \text{ for } t = 1, \ldots, T, \\ \tilde{v}_{T+1} &= \nabla \rho_{V_{T+1} Q_{T+1}}(q_{T+1} - C_{T+1} x_T). \end{aligned}$$

(c) Determine $\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_{T+1})$ from $\tilde{v} = (\tilde{v}_1, \ldots, \tilde{v}_{T+1})$ via the dual dynamics (5).
(d) Then $\tilde{v} = F(u)$, while $\nabla f(u)$ is given by

$$\nabla_{u_t} f(u_0, u_1, \ldots, u_t) = \begin{cases} (p_t - B_t^* \tilde{y}_{t+1} - D_t^* \tilde{v}_t) + P_t u_t & \text{for } t = 1, \ldots, T, \\ (p_0 - B_1^* \tilde{y}_1) + P_0 u_0 & \text{for } t = 0. \end{cases}$$

Proof. From the definition of $F(u)$ in Theorem 3, with expression (4) used in the Lagrangian (3), one obtains that $F(u)$ is the vector $\tilde{v}'$ having components

$$\tilde{v}'_t = \operatorname*{argmax}_{v_t \in V_t} \{v_t \cdot (q_t - C_t x_{t-1} - D_t u_t) - \tfrac{1}{2} v_t \cdot Q_t v_t\} \text{ for } t = 1, \ldots, T,$$
$$\tilde{v}'_{T+1} = \operatorname*{argmax}_{v_{T+1} \in V_{T+1}} \{v_{T+1} \cdot (q_{T+1} - C_{T+1} x_T) - \tfrac{1}{2} v_{T+1} \cdot Q_{T+1} v_{T+1}\}.$$

Consider for each t the function $\varphi_{V_t Q_t}(v_t)$, which has the value $\frac{1}{2} v_t \cdot Q_t v_t$ when $v_t \in V_t$ but equals ∞ when $v_t \notin V_t$. This function is strongly convex (because Q_t is positive definite), and its conjugate is $\rho_{V_t Q_t}$. From this conjugacy, the formula just given can be interpreted in convex analysis (cf. [28, Thm. 23.5]) as saying that $\tilde{v}'_t$ is the unique subgradient of $\rho_{V_t Q_t}$ at the point $q_t - C_t x_{t-1} - D_t u_t$ (or, in the case of $t = T+1$, the point $q_{T+1} - C_{T+1} x_T$). When a convex function has a unique subgradient, that subgradient must actually be the gradient (cf. [28, Thm. 25.1]). Thus, $\tilde{v}'_t = \tilde{v}_t$. This proves the asserted formula for $F(u)$. We have then from Theorem 3 that $\nabla f(u) = \nabla_u L(u, \tilde{v})$, where in the definition (3) of $L(u, \tilde{v})$ one invokes (6) with the dual trajectory $\tilde{y}$. This yields the asserted formula for $\nabla_{u_t} f(u)$. □

THEOREM 5. *In the strictly quadratic case, the following procedure calculates $G(v)$ and then $\nabla g(v)$ for any given v:*
(a) Determine $y = (y_1, \ldots, y_{T+1})$ from $v = (v_1, \ldots, v_{T+1})$ via the dual dynamics (5).
(b) Calculate $\tilde{u} = (\tilde{u}_0, \ldots, \tilde{u}_T) \in U$ by

$$\tilde{u}_t = \nabla \rho_{U_t P_t}(B_t^* y_{t+1} + D_t v_t - q_t) \text{ for } t = 1, \ldots, T,$$
$$\tilde{u}_0 = \nabla \rho_{U_0 P_0}(B_0^* y_1 - p_0).$$

(c) Determine $\tilde{x} = (\tilde{x}_0, \ldots, \tilde{x}_T)$ from $\tilde{u} = (\tilde{u}_0, \ldots, \tilde{u}_T)$ via the primal dynamics (2).
(d) Then $\tilde{u} = G(v)$, while $\nabla g(v)$ is given by

$$\nabla_{v_t} g(v_1, \ldots, v_T, v_{T+1}) = \begin{cases} (q_t - C_t \tilde{x}_{t-1} - D_t^* \tilde{u}_t) - Q_t v_t & \text{for } t = 1, \ldots, T, \\ (q_{T+1} - C_{T+1} \tilde{x}_T) - Q_{T+1} u_{T+1} & \text{for } t = T+1. \end{cases}$$

Proof. The argument is parallel to the one for Theorem 4. □

5. Strictly Quadratic Regularization.

The results stated in Theorems 3, 4 and 5 for the strictly quadratic case can be applied in a fundamental way even when problem $(\mathcal{P})$ is not strictly quadratic. One sets up a regularizing scheme of "outer" iterations which converts the solving of $(\mathcal{P})$ into the solving of a sequence of slightly modified, but strictly quadratic problems. An appropriate mechanism is the general *proximal point algorithm*, developed in Rockafellar [29].

The proximal point algorithm could be applied to our multistage model in a primal, dual, or primal-dual mode, following the pattern already known for its applications to standard problems in convex programming [30]. The primal-dual mode is the one of interest in the general case where neither P_t nor Q_t can be counted on as positive definite, and it is therefore the mode we concentrate on here. We can view the algorithm in this mode as operating on $(\mathcal{P})$ either in terms of the saddle point problem for the Lagrangian (3) or in the context of the variational inequality (10)-(11). Both ways, the implementation requires us to specify auxiliary matrices

$$\bar{P}_t \in \mathbb{R}^{k_t \times k_t}, \qquad \bar{Q}_t \in \mathbb{R}^{l_t \times l_t}, \qquad \text{symmetric and positive definite.}$$

Also required is a *bounded* sequence of values $\gamma_\nu > 0$, where $\nu = 1, 2, \ldots$, is the iteration index.

THEOREM 6. *Starting from any choice of elements $\bar{u}^0 \in U$ and $\bar{v}^0 \in V$, generate $\bar{u}^\nu$ and $\bar{v}^\nu$ iteratively for $\nu = 1, 2, \ldots$, by taking $\bar{u}^\nu$ to be the unique solution to the strictly quadratic problem $(\mathcal{P}^\nu)$ obtained from $(\mathcal{P})$ in replacing the elements P_t, Q_t, p_t and q_t by*

$$P_t^\nu := P_t + \gamma_\nu \bar{P}_t, \quad Q_t^\nu := Q_t + \gamma_\nu \bar{Q}_t, \quad p_t^\nu := p_t - \gamma_\nu \bar{P}_t \bar{u}^{\nu-1}, \quad q_t^\nu := q_t + \gamma_\nu \bar{Q}_t \bar{v}^{\nu-1}, \tag{16}$$

and by taking $\bar{v}^\nu$ to be the unique solution to the corresponding dual problem $(\mathcal{Q}^\nu)$. If the original problem $(\mathcal{P})$ has a solution at all (cf. Theorem 2), then $\bar{u}^\nu$ converges to a solution $\bar{u}$ to $(\mathcal{P})$ (even though $(\mathcal{P})$ may have more than one solution), while $\bar{v}^\nu$ converges to a solution $\bar{v}$ to $(\mathcal{Q})$.

Proof. The uniqueness of $\bar{u}^\nu$ and $\bar{v}^\nu$ follows from the objective function f^ν in $(\mathcal{P}^\nu)$ being strictly convex because P^ν is positive definite, and the objective function g^ν in $(\mathcal{Q}^\nu)$ being strictly concave because Q^ν is positive definite. From Theorem 1 as applied to $(\mathcal{P}^\nu)$ we know that the pair $(\bar{u}^\nu, \bar{v}^\nu)$ is the unique saddle point of $L^\nu(u, v)$ relative to $U \times V$, where L^ν is the Lagrangian for $(\mathcal{P}^\nu)$. In the norm notation

$$\|u\|_{\bar{P}} := \big(\sum\nolimits_{t=0}^{T} u_t \cdot \bar{P}_t u_t\big)^{\frac{1}{2}}, \qquad \|v\|_{\bar{Q}} := \big(\sum\nolimits_{t=1}^{T+1} v_t \cdot \bar{Q}_t v_t\big)^{\frac{1}{2}}, \tag{17}$$

this Lagrangian has the form

$$L^\nu(u, v) = L(u, v) + \tfrac{1}{2}\gamma_\nu \|u - \bar{u}^{\nu-1}\|_{\bar{P}}^2 - \tfrac{1}{2}\gamma_\nu \|v - \bar{v}^{\nu-1}\|_{\bar{Q}}^2 + \text{const.} \tag{18}$$

The norms (17) can be taken as inducing a Euclidean structure on the u and v spaces, and the procedure then fits the pattern of the minimax version of the proximal point algorithm in [29, Theorem 5]. In particular, one obtains the claimed convergence. □

The saddle point interpretation of the algorithm in Theorem 3, in terms of the pair $(\bar{u}^\nu, \bar{v}^\nu)$ being the unique saddle point of the Lagrangian $L^\nu(u, v)$ in (18) relative to $U \times V$, provides the connection with variational inequalities and leads to a result on the *rate* of convergence.

THEOREM 7. *The algorithm in Theorem 6 generates the pair $(\bar{u}, \bar{v})$ from $(\bar{u}^{\nu-1}, \bar{v}^{\nu-1})$ as the solution to the variational inequality*

$$M^\nu(\bar{u}^\nu, \bar{v}^\nu) \cdot \big[(u, v) - (\bar{u}^\nu, \bar{v}^\nu)\big] \leq 0 \text{ for all } (u, v) \in U \times V, \tag{19}$$

where $M^\nu(u, v) = (\nabla_u L^\nu(u, v), -\nabla L^\nu(u, v))$, or equivalently in terms of the positive definite matrix $H = \operatorname{diag}\{\bar{P}_0, \ldots, \bar{P}_T; \bar{Q}_1, \ldots, \bar{Q}_{T+1}\}$ and the affine mapping M in (12),

$$M^\nu(u, v) = M(u, v) + \gamma_\nu H\big[(u, v) - (\bar{u}^\nu, \bar{v}^\nu)\big]. \tag{20}$$

Relative to the multivalued mapping $\bar{M}(u,v) := M(u,v) + N_{U\times V}(u,v)$, which is maximal monotone, this rule of generation can be written as

$$(\bar{u}^\nu, \bar{v}^\nu) = (I + \gamma_\nu^{-1} H^{-1}\bar{M})^{-1}(\bar{u}^{\nu-1}, \bar{v}^{\nu-1}). \tag{21}$$

Proof. The reduction of the saddle point condition for L^ν in the proof of Theorem 6 to the variational inequality (19) follows through equivalence with the intermediate condition

$$-\nabla_u L^\nu(\bar{u}^\nu, \bar{v}^\nu) \in N_U(\bar{u}^\nu), \qquad \nabla_v L^\nu(\bar{u}^\nu, \bar{v}^\nu) \in N_V(\bar{v}^\nu).$$

The affine mapping M is itself maximal monotone because $L(u,v)$ is convex in u and concave in v (see [31]). The maximal monotonicity of the multivalued mapping $\bar{M}$ is then follows because $U \times V$ is a closed convex set (see [32, Theorem 3].) The single-valuedness of the mapping on the right side of (21) is due to $(\bar{u}^\nu, \bar{v}^\nu)$ being uniquely determined as a saddle point of $L^\nu(u,v)$ on $U \times V$, which in turn is a consequence of the strictly quadratic terms in the definition of $L^\nu(u,v)$ in (18). □

THEOREM 8. *Suppose that $(\mathcal{P})$ and $(\mathcal{Q})$ have unique solutions $\bar{u}$ and $\bar{v}$, and let γ_ν be chosen such that $\gamma_\nu \to \gamma_\infty \geq 0$. Then, unless the algorithm in Theorem 6 actually terminates in finitely many steps (with $(\bar{u}^\nu, \bar{v}^\nu) = (\bar{u}, \bar{v})$ for all ν sufficiently large), there is a constant $a \geq 0$ such that*

$$\lim_{\nu\to\infty} \frac{\|(\bar{u}^\nu, \bar{v}^\nu) - (\bar{u}, \bar{v})\|_H}{\|(\bar{u}^{\nu-1}, \bar{v}^{\nu-1}) - (\bar{u}, \bar{v})\|_H} \leq \frac{a\gamma_\infty}{(1 + (a\gamma_\infty)^2)^{\frac{1}{2}}} < 1, \tag{22}$$

where the norm is

$$\|(u,v)\|_H := \sqrt{(u,v)\cdot H(u,v)} = \big(\sum\nolimits_{t=0}^{T} u_t \cdot \bar{P}_t u_t + \sum\nolimits_{t=1}^{T+1} v_t \cdot \bar{Q}_t v_t\big)^{\frac{1}{2}}. \tag{23}$$

Proof. Letting $K = H^{-\frac{1}{2}}$, we make the change of variables $\bar{w}^\nu = K^{-1}(\bar{u}^\nu, \bar{v}^\nu)$ to convert (21) into

$$\bar{w}^\nu = (I + \gamma_\nu^{-1}\tilde{M})^{-1}(\bar{w}^{\nu-1}) \text{ for } \tilde{M}(w) := K\bar{M}(Kw).$$

The mapping $\tilde{M}$ inherits maximal monotonicity from $\bar{M}$, and the procedure is thus converted to the fundamental form of the proximal point algorithm in Rockafellar [29]. Our hypothesis on the uniqueness of $\bar{u}$ and $\bar{v}$ as primal and dual solutions means that these elements uniquely satisfy $(0,0) \in \bar{M}(\bar{u}, \bar{v})$, and this translates to the pair $\bar{w} = (\bar{u}, \bar{v})$ being the unique solution to $0 \in \tilde{M}(\bar{w})$. According to Theorem 2 of the cited paper [29], we will have

$$\lim_{\nu\to\infty} \frac{\|(\bar{w}^\nu - \bar{w}\|}{\|\bar{w}^{\nu-1} - \bar{w}\|} \leq \frac{a\gamma_\infty}{(1 + (a\gamma_\infty)^2)^{\frac{1}{2}}}$$

if we can establish the existence of $a \geq 0$ and $\varepsilon > 0$ such that

$$\|w - \bar{w}\| \leq a\|w'\| \text{ when } w' \in \tilde{M}(w) \text{ and } \|w'\| \leq \varepsilon. \tag{24}$$

This will suffice for our result, because $\|\bar{w}^\nu - \bar{w}\| = \|(\bar{u}^\nu, \bar{v}^\nu) - (\bar{u}, \bar{v})\|_H$ from the definitions.

The mapping $\tilde{M}$ is polyhedral in the sense of Robinson [33]: its graph is the union of a finite collection of polyhedral sets. This follows from the definition of $\tilde{M}$ in terms of $\bar{M}$, which by (20) is the sum of the affine mapping M and the normal cone mapping $(u,v) \mapsto N_{U\times V}(u,v)$. The latter is polyhedral because the sets U and V are polyhedral, see [33]. We know on the other hand that $\tilde{M}$ is maximal monotone with $\bar{w}$ the unique solution to $0 \in \tilde{M}(w)$. Therefore, the inverse $\tilde{M}^{-1}$ is maximal monotone and polyhedral as well as single-valued at 0 (with the value $\bar{w}$ there). The maximal monotonicity and single-valuedness at 0 imply that $\tilde{M}^{-1}(w') \neq \emptyset$ for all w' in some neighborhood of 0; see [34, Theorem 1]. The polyhedral property then assures that $\tilde{M}^{-1}$ is actually single-valued and piecewise affine in some neighborhood of 0. Such a mapping in particular has, for some $a \geq 0$ and $\varepsilon > 0$, the property that

$$\|\tilde{M}^{-1}(w') - \tilde{M}^{-1}(0)\| \leq a\|w'\| \text{ when } \|w'\| \leq \varepsilon,$$

and this is equivalent to the desired statement (22). □

The version of the proximal point algorithm in Theorems 6, 7 and 8 calls for $\bar{u}^\nu$ and $\bar{v}^\nu$ to be the exact solutions to the strictly quadratic primal and dual subproblems in each iteration. Actually, the supporting theory in [29] do not require this. It is possible without great difficulty to develop a version in which approximate solutions suffice in each iteration, under a certain type of stopping condition.

Furthermore, the assumption in Theorem 8 that $(\mathcal{P})$ and $(\mathcal{Q})$ have unique solutions is not really needed. Making use of the refinements obtained by Luque [35] in the convergence properties of the general proximal point algorithm in [29], it can be demonstrated that (22) always holds in the broader sense of the distance (with respect to the norm induced by H) of $(\bar{u}^\nu, \bar{v}^\nu)$ from the set of all saddle points, rather than the distance from the unique saddle point.

6. Splitting Methods and Lagrangian Decomposition.

The special formulation of the multistage problem $(\mathcal{P})$ also invites the application of so-called splitting methods. Such methods aim at solving a variational inequality like (11) by decomposing it into simpler conditions in the form of auxiliary variational inequalities. Iteratively these simpler conditions are solved relative to certain parameter elements, and their solutions are combined to get new parameter elements. Examples include the alternating direction method for convex programming [8], [9], [10], and the more general algorithm of Lions and Mercier [11]. Spingarn [12], [13], developed a class of splitting methods based on applying the proximal point algorithm [29] to the partial inverse of a maximal monotone mapping and showed that the alternating direction method and many other decomposition techniques were covered as a special case. Recently Eckstein [14], [15], has carried this further and provided an overview showing that even the Lions-Mercier algorithm falls essentially in this class.

Thus, a rich family of numerical methods for large-scale problems can be seen in terms of instances of the general proximal point algorithm. The properties of that algorithm have been invoked in Theorems 6, 7 and 8 with respect to strict quadratic regularization, but the larger question arises of whether splitting methods likewise have a special role to play in this framework of multistage optimization. Tseng [16], [17], has specifically applied a type of splitting method to $(\mathcal{P})$ and shown that this leads to a solution procedure allowing for massive parallelization. Many other possibilities can be explored, however, beyond the one discovered by Tseng. The structure of the Lagrangian in (3), with highly separable terms in t supplemented by a bi-affine form expressing the dynamics, can be made the basis of a kind of decomposition in which, on the one hand, low-dimensional problems of extended linear-quadratic programming are repeatedly solved for each t, while on the other hand, the dynamics are treated through subproblems of ordinary linear-quadratic optimal control without constraints. This will be discussed elsewhere.

REFERENCES

1. R. T. ROCKAFELLAR, *Linear-quadratic programming and optimal control*, SIAM J. Control Opt., 25 (1987), pp. 781–814.

2. R. T. ROCKAFELLAR, *On the essential boundedness of solutions to problems in piecewise linear-quadratic optimal control*, in: *Analyse Mathématique et Applications*, F. Murat and O. Pironneau (eds.), Gauthier-Villars, Paris, 1988, pp. 437–444.

3. R. T. ROCKAFELLAR, *Multistage convex programming and discrete-time optimal control*, Control and Cybernetics, 17 (1988), pp. 225–246.

4. R. T. ROCKAFELLAR, *Hamiltonian trajectories and duality in the optimal control of linear systems with convex costs*, SIAM J. Control Opt., 27 (1989), pp.

5. R. T. ROCKAFELLAR and R. J-B WETS, *Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time*, SIAM J. Control Opt., to appear.

6. R. T. ROCKAFELLAR, *Computational schemes for solving large-scale problems in extended linear-quadratic programming*, Math. Programming, Ser. B, to appear.

7. C. ZHU and R. T. ROCKAFELLAR, *Finite-envelope gradient projection methods for extended linear-quadratic programming*, preprint.

8. D. GABAY and B. MERCIER, *A dual algorithm for the solution of nonlinear variational problems via finite element approximations*, Computers and Math. with Appl., 2 (1976), pp. 17–40.

9. M. FORTIN and R. GLOWINSKI, *On decomposition-coordination methods using an augmented Lagrangian*, in: *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, M. Fortin and R. Glowinski (eds.), North-Holland, Amsterdam, 1983.

10. D. GABAY, *Applications of the method of multipliers to variational inequalities*, in: *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, M. Fortin and R. Glowinski (eds.), North-Holland, Amsterdam, 1983.

11. P.-L. LIONS and B. MERCIER, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Analysis, 16 (1979), 964–979.

12. J. E. SPINGARN, *Partial inverse of a monotone operator*, Appl. Math. Optimization, 10 (1983), pp. 247–265.

13. J. E. SPINGARN, *Applications of the method of partial inverses to convex programming*, Math. Programming, 32 (1985), pp. 199–223.

14. J. ECKSTEIN, *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, doctoral dissertation, M.I.T., 1989.

15. J. ECKSTEIN, *The Lions-Mercier algorithm and the alternating direction method are instances of the proximal point algorithm*, report LIDS-P-1769, Laboratory for Information and Decision Sciences, M.I.T., 1989.

16. P. TSENG, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, LIDS Report P-1836, M.I.T., Cambridge, MA, 1989.

17. P. TSENG, *Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming*, LIDS Report P-1866, M.I.T., Cambridge, MA, 1989.

18. J.-S. PANG, *Methods for quadratic programming: a survey*, Computers and Chem. Engineering, 7 (1983), pp. 583–594.

19. Y.-Y. LIN and J.-S. PANG, *Iterative methods for large convex quadratic programs: a survey*, SIAM J. Control Opt., 25 (1987), pp. 383–411.

20. Y. YE and E. TSE, *A polynomial-time algorithm for convex quadratic programming*, preprint.

21. R. C. MONTEIRO and I. ADLER, *An $0(n^3L)$ interior point algorithm for convex quadratic programming*, preprint.

22. D. GOLDFARB and S. LIU, *An $0(n^3L)$ primal interior point algorithm for convex quadratic programming*, preprint.

23. R. T. ROCKAFELLAR, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, SIAM J. Control, 12 (1974), pp. 268–285.

24. R. T. ROCKAFELLAR and R. J-B WETS, *Linear-quadratic problems with stochastic penalties: the finite generation algorithm*, in: *Numerical Techniques for Stochastic Optimization Problems*, Y. Ermoliev and R. J-B Wets (eds.), Springer-Verlag Lecture Notes in Control and Information Sciences No. 81, 1987, pp. 545-560.

25. R. T. ROCKAFELLAR and R. J-B WETS, *A Lagrangian finite generation technique for solving linear-quadratic problems in stochastic programming*, Math. Programming Studies, 28 (1986), pp. 63–93.

26. A. KING, *An implementation of the Lagrangian finite generation method*, in: *Numerical Techniques for Stochastic Programming Problems*, Y. Ermoliev and R. J-B Wets (eds.), Springer-Verlag, 1988.

27. J. M. WAGNER, *Stochastic Programming with Recourse Applied to Groundwater Quality Management*, doctoral dissertation, M.I.T, 1988.

28. R. T. ROCKAFELLAR, *Convex Analysis*, Princeton Univ. Press, Princeton, NJ, 1970.

29. R. T. ROCKAFELLAR, *Monotone operators and the proximal point algorithm*, SIAM J. Control Opt., 14 (1976), pp. 877–898.

30. R. T. ROCKAFELLAR, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. of Op. Research, 1 (1976), pp. 97–116.

31. R. T. ROCKAFELLAR, *Monotone operators associated with saddle functions and minimax problems*, in: *Nonlinear Functional Analysis, Part 1*, F. E. Browder (ed.), Symposia in Pure Math., vol. 18, Amer. Math. Soc., Providence, RI, 1970, pp. 397–407.

32. R. T. ROCKAFELLAR, *On the maximality of sums of nonlinear monotone operators*, Trans. Amer. Math. Soc., 149 (1970), pp. 75–88.

33. S. M. ROBINSON, *Some continuity properties of polyhedral multifunctions*, Math. Programming Study *14*, 1981, pp. 206–214.

34. R. T. ROCKAFELLAR, *Local boundedness of nonlinear monotone operators*, Michigan Math. J., 16 (1969), pp. 397–407.

35. F. J. LUQUE, *Asymptotic convergence analysis of the proximal point algorithm*, SIAM J. Control Opt., 22 (1984), pp. 277–293.