To appear in Math. Programming, vol. 48 (1990)

# COMPUTATIONAL SCHEMES FOR LARGE-SCALE PROBLEMS IN EXTENDED LINEAR-QUADRATIC PROGRAMMING

R.T. Rockafellar*

Department of Mathematics
University of Washington GN–50
Seattle, WA 98195

September, 1989

**Abstract.** Numerical approaches are developed for solving large-scale problems of extended linear-quadratic programming that exhibit Lagrangian separability in both primal and dual variables simultaneously. Such problems are kin to large-scale linear complementarity models as derived from applications of variational inequalities, and they arise from general models in multistage stochastic programming and discrete-time optimal control. Because their objective functions are merely piecewise linear-quadratic, due to the presence of penalty terms, they do not fit a conventional quadratic programming framework. They have potentially advantageous features, however, which so far have not been exploited in solution procedures. These features are laid out and analyzed for their computational potential. In particular, a new class of algorithms, called finite-envelope methods, is described that does take advantage of the structure. Such methods reduce the solution of a high-dimensional extended linear-quadratic program to that of a sequence of low-dimensional ordinary quadratic programs.

**Keywords.** Extended linear-quadratic programming, large-scale numerical optimization, finite-envelope methods, stochastic programming, nonsmooth optimization, linear complementarity, variational inequalities.

* This work was supported in part by grants AFOSR 87–0821 and AFOSR 89–0081 from the Air Force Office of Scientific Research.

## 1. INTRODUCTION

Large-scale problems in quadratic programming and linear complementarity have received much attention as mathematical models in applications of optimization to such areas as the solution of variational inequalities, cf. [1]. A comprehensive survey of iterative methods for solving such problems has been furnished by Lin and Pang [2]. Generally speaking, these methods try to adapt to special structure by taking the sparsity of certain matrices into account. It has long been recognized that without some such adaptation, there is little hope of solving high-dimensional problems effectively. Less clear, however, is whether quadratic programming and linear complementarity methods for sparse matrices offer a fully satisfactory approach to special structure.

In variational inequalities derived from discretized boundary value problems with partial differential operators, for instance, the large number of constraints may cause difficulties of a kind of degeneracy if handled directly, as called for by standard quadratic programming and linear complementarity formulations. Penalty terms seem desirable instead, as well as improved treatment of the natural dual variables associated with the underlying operators, which should not be suppressed algebraically or imagined just as additional "multipliers." The investigation of computational approaches based on alternative representations of model structure, not relying solely on patterns of matrix sparsity, is therefore worthwhile.

Problems of "extended linear-quadratic programming" have been introduced recently in setting up various large-scale models in stochastic programming and multistage optimization; see Rockafellar [3], Rockafellar and Wets [4], [5], [6]. These problems differ from the usual ones in linear or quadratic programming in allowing for penalty terms of piecewise linear-quadratic type along with more general expressions like those associated with augmented Lagrangians. Their objective functions then are only piecewise linear-quadratic in general. On the other hand, the exact constraints that are present, besides the constraints formulated with penalty terms, may be quite simple. This is frequently the case at least in the applications worked out so far, as just cited.

Such features of problems in extended linear-quadratic programming are of keen interest in a large-scale context, because complicated constraints can greatly aggravate any approach to finding solutions and, unless systematically relaxed, cause unpleasant instabilities. In continuous-time optimal control, for example, the presence of pure state constraints on the trajectories in the primal problem, if not represented by penalties, could lead in the statement of optimality conditions to jumps in the trajectories of the adjoint state variables (the naturally associated dual variables in such a context). This would mean in the corresponding dual problem that impulse controls have to be taken into account. From a practical standpoint, the implication would be that in trying to approximate a continuous-time model by a discrete-time model, serious trouble could arise. The discretization could be quite ill-behaved. Much the same phenomenon can be expected with many applications of variational inequalities involving PDE's, as already indicated.

Whether the use of penalty representations in mathematical modeling is to be viewed as an appealing idea in itself, on a par with other accepted simplifications like linearization of nonlinear expressions, or just as a convenient device on a purely technical level for getting temporary approximations that can later be improved, an important challenge is laid down. How can one compute solutions to problems in which such penalties appear and the dimensionality is likely to be large? The case of high-dimensional extended linear-quadratic programming is fundamental in this respect but has not previously received general attention.

A difficulty with this case lies in the circumstance that the objective function in an extended linear-quadratic programming problem may only be *piecewise* linear-quadratic. Its domain can in principle be decomposed into a finite collection of polyhedral convex cells, on each of which the function is expressible by a linear or quadratic formula. An ordinary direct approach to minimizing such a piecewise linear-quadratic function is unsuitable for the applications we are aiming at because an explicit description of all the cells that are involved would be prohibitive—and the dimension of the space in which the cells lie may itself be very high. (We use the term "linear-quadratic" instead of just "quadratic" to emphasize that extended linear programming models with bounded variables and linear penalization are also in the range of interest.)

The primal context is essentially one of *high-dimensional nonsmooth convex optimization where the second derivatives, and sometimes the first, can be discontinuous.* The high dimensionality precludes the use of primal methods based on simplicial decomposition, say, or the typical numerical techniques for nonsmooth optimization developed until now, such as bundle methods (see Kiwiel [7]). As a matter of fact, no primal method at all exists at present for such problems on the scale appropriate to them, much less a method designed with their special features in mind.

Despite this formidable difficulty, there is a convenient way of expressing the objective function in an extended linear-quadratic programming problem by an "envelope" formula derived from an algebraically simple Lagrangian function—not involving a cell decomposition. In the desired applications this Lagrangian typically can be made to exhibit properties of *decomposability with respect to both the primal and the dual variables*, as demonstrated in [3], [4], [5], [6]. Line searches then become numerically feasible, and new uses of duality open up in which the adjoint variables associated with the dynamics or stochastics in a given application can be assigned a significant role. In any case, a type of structure comes into view which has not previously been studied in its implications for solution procedures. This structure is especially suggestive of primal-dual methods, although it might lead eventually to viable primal methods as well.

There are two goals in this paper. The first is to present in a broad way, not oriented to only one type of algorithm or application, the features of extended linear-quadratic programming that suggest untapped possibilities for the development of numerical methods in large-scale optimization. The second goal is to describe, and lay a theoretical foundation

for, a specific class of methods (by no means the only conceivable ones) which do take advantage of envelope representations and double decomposability. These methods, called *finite-envelope methods*, exploit duality very strongly and operate by generating finitely determined envelope approximations which yield directions of improvement. They reduce the task of solving a problem of high dimensionality to that of solving a sequence of ordinary quadratic programming subproblems of low dimensionality, and they do so in a manner that respects underlying dynamic or stochastic relationships and supports extensive parallelization.

Finite-envelope methods in a much narrower conception have already been introduced in Rockafellar and Wets [4], [5], for the special case of two-stage stochastic programming; see also King [8] and Wagner [9]. The characteristic of this special case is that the primal dimension is low; only the dual dimension is high. In contrast here, we allow, for the first time, both dimensions to be high and we place the ideas in a wider framework which does not invoke probabilistic language or ideas. This forces us to a qualitatively very different level of theory but paves the way to applications of the kind in [3] and [6].

The chief result we obtain here about finite-envelope methods is a proof of convergence for the case of "fully quadratic" problems, where the primal and dual quadratic terms are nonsingular. The fully quadratic case can always be induced for computational purposes, if desired, through an outer procedure derived from the proximal point algorithm [10], [11], [12]. We further obtain a result about the automatic eventual identification of the active constraints.

It must be appreciated, of course, that these results are only a first step toward the understanding of general finite-envelope methods and their practical significance. Much needs to be investigated in the precise choice of the envelope representations, not to mention the alternative ways of solving the low-dimensional subproblems, before a real appraisal can be made of the efficacy of this kind of approach. Such an appraisal is difficult anyway because of the lack of tried-and-true methods, applicable to the large-scale models in question, with which a comparison can be made. The contributions in this paper should mainly be judged therefore in terms of bringing new issues to the fore and attempting to make at least some headway with them.

## 2. EXTENDED LINEAR-QUADRATIC PROGRAMMING

A brief review of the central facts in [3] and [4] about extended linear-quadratic programming problems, whether large-scale or small-scale, will provide a basis for the developments in the rest of the paper. A key ingredient in the formulation of such problems is a class of what we call *monitoring functions*, which serve in the modeling of constraints but can also be used in other ways. A typical function in this class is expressible as

$$\rho_{VQ}(w) = \sup_{v\in V}\{w\cdot v - \tfrac{1}{2}v\cdot Qv\} \text{ for } w \in \mathbb{R}^m, \tag{2.1}$$

where $V$ is a nonempty polyhedral convex set in $\mathbb{R}^m$ and $Q$ is an $m \times m$ symmetric positive semidefinite matrix (possibly the zero matrix). As will be explained presently, such functions also be given other, more direct formulas in important cases.

A general problem of extended linear-quadratic programming in $\mathbb{R}^n$ takes the form

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

The set $U$ is nonempty polyhedral convex in $\mathbb{R}^n$, and the $n \times n$ matrix $P$ is symmetric and positive semidefinite. One has $p \in \mathbb{R}^n$, $q \in \mathbb{R}^m$ and $R \in \mathbb{R}^{m\times n}$.

The polyhedral convexity of $U$ expresses an underlying system of linear constraints which are considered to be relatively easy to handle. All other modeling of constraints is done through the term $\rho_{VQ}(q-Ru)$, which responds to deviations of $Ru$ from $q$. Note that if $0 \in V$ the function $\rho_{VQ}$ is nonnegative everywhere but vanishes at 0 (actually on a certain polyhedral cone $K$ which may or may not reduce to just 0). It thus assigns *penalty* values the differences $q - Ru$ (which are positive when $q - Ru \neq K$). For theoretical reasons that will be clear later, as well as motivations derived from modeling in various applications (see Rockafellar [3,SS3,5] for example), we do not insist always on having $0 \in V$, however, so that $\rho_{VQ}$ may in general have negative as well as positive values. This is why we speak of $\rho_{VQ}$ as a monitoring function rather than a penalty function. (For instance, if $V = \{y\}$ and $Q = 0$, then $\rho_{VQ}$ is the linear function $w \mapsto y\cdot w$ and $\rho_{VQ}(q-Ru)$ is the ordinary Lagrangian expression $y\cdot(q - Ru)$; augmented Lagrangian terms can be represented similarly.)

In the absence of additional assumptions the function $\rho_{VQ}$ could take on the value $\infty$ in some regions of $\mathbb{R}^m$, but the set of points $w$ with $\rho_{VQ}(w) < \infty$ turns out always to be a polyhedral convex cone $C$ [3, Proposition 2.3]. The constraint $q - Ru \in C$ is implicit then in $(\mathcal{P})$ along with $u \in U$, unless $C$ is all of $\mathbb{R}^m$. On $C$, $\rho_{VQ}$ is piecewise linear-quadratic convex [3, Proposition 2.3]. Therefore, the objective function $f$ in $(\mathcal{P})$ is piecewise linear-quadratic convex.

An important case that helps to clarify the formulation of $(\mathcal{P})$ and the motivations behind the monitoring term $\rho_{VQ}(q - Ru)$ is the one we shall refer to in the sequel as the *box-diagonal case*. Then $P$ and $Q$ are diagonal matrices,

$$P = \operatorname{diag}[\alpha_1, \ldots, \alpha_n], \qquad Q = \operatorname{diag}[\beta_1, \ldots, \beta_m], \tag{2.2}$$

and $U$ and $V$ are boxes representing upper and lower bounds (not necessarily finite) on the components of $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_m)$:

$$U = [u_1^-, u_1^+] \times \ldots \times [u_n^-, u_n^+], \qquad V = [v_1^-, v_1^+] \times \ldots \times [v_m^-, v_m^+]. \tag{2.3}$$

Then $(\mathcal{P})$ specializes to the problem of minimizing

$$\sum_{j=1}^{n} [p_j u_j + \tfrac{1}{2}\alpha_j u_j^2] + \sum_{i=1}^{m} \rho_{v_i^-, v_i^+, \beta_i}\left(q_i - \sum_{j=1}^{n} r_{ij} u_j\right) \tag{2.4}$$

subject to $u_j^- \le u_j \le u_j^+$ for $j = 1, \ldots, n$, where

$$\rho_{v_i^-, v_i^+, \beta_i}(w_i) = \begin{cases} v_i^+ w_i - \frac{1}{2}\beta_i (v_i^+)^2 & \text{if } w_i/\beta_i > v_i^+, \\ \frac{1}{2} w_i^2/\beta_i & \text{if } v_i^- \le w_i/\beta_i \le v_i^+, \\ v_i^- w_i - \frac{1}{2}\beta_i (v_i^-)^2 & \text{if } w_i/\beta_i < v_i^-. \end{cases} \tag{2.5}$$

Geometrically, the graph of this $\rho$ function is obtained from the graph of the function $w_i \mapsto \frac{1}{2} w_i^2/\beta_i$ by extrapolating the latter linearly to the right beyond the point where $w_i$ equals $\beta_i v_i^+$, and to the left beyond the point where $w_i$ equals $\beta_i v_i^-$. Thus for instance, if $v_i^- = 0$ and $v_i^+ > 0$, the corresponding term in (2.4) vanishes when the inequality $\sum_{j=1}^n r_{ij} u_j \ge q_i$ is satisfied, but gives a positive penalty when this inequality is violated. The penalty grows smoothly at a quadratic rate initially, as dictated by the value of $\beta_i$, but for violations greater than a certain amount it grows only linearly at a rate specified by $v_i^+$. A modeling decision on what these rates should be determines $V$ and $Q$ and the overall function $\rho_{VQ}$.

Formula (2.5) appears to require $\beta_i$ to be positive and $v_i^-$ and $v_i^+$ to be finite, but it is easily extended to limiting cases where possibly $\beta_i = 0$, $v_i^- = -\infty$, or $v_i^+ = \infty$. As an illustration, one has

$$\rho_{0,\infty,0}(w_i) = \begin{cases} 0 & \text{if } w_i \le 0, \\ \infty & \text{if } w_i > 0. \end{cases}$$

Then the term (2.5) gives in (2.4) a strict representation of the constraint $\sum_{j=1}^n r_{ij} u_j \ge q_i$ by infinite penalties. On the other hand, from the case of

$$\rho_{-\infty,\infty,\beta_i}(w_i) = \tfrac{1}{2}\beta_i^{-1} w_i^2 \text{ for all } w_i$$

one would have a purely quadratic penalty representation of the equation $\sum_{j=1}^n r_{ij} u_j = q_j$.

Duality plays a very strong role in extended linear-quadratic programming. According to general theory in [3,S2], the problem dual to $(\mathcal{P})$ is

$$(\mathcal{Q}) \qquad \text{minimize } g(v) \text{ over all } v \in V, \text{ with } g(v) := q \cdot v - \tfrac{1}{2} v \cdot Q v - \rho_{UP}(R^T v - p).$$

The matrix $R^T$ is the transpose of $R$, and the monitoring function $\rho_{UP}$ is defined exactly like $\rho_{VQ}$:

$$\rho_{UP}(z) = \sup_{u \in U} \{z \cdot u - \tfrac{1}{2} u \cdot P u\} \text{ for } z \in \mathbb{R}^n. \tag{2.6}$$

In the box-diagonal case, for example, $(\mathcal{Q})$ is the problem of maximizing

$$\sum_{i=1}^{m}[q_i v_i - \tfrac{1}{2}\beta_i v_i^2] - \sum_{j=1}^{n} \rho_{u_i^-,u_i^+,\alpha_i}\left(\sum_{i=1}^{m} v_i r_{ij} - p_j\right) \tag{2.7}$$

subject to $v_i^- \le v_i \le v_i^+$ for $i = 1, \ldots, m$.

The deep connection between these problem is the following.

**Theorem 2.1** [3],[4]**.** *Unless both* $\inf(\mathcal{P}) = \infty$ *and* $\sup(\mathcal{Q}) = -\infty$, *one has* $\inf(\mathcal{P}) = \sup(\mathcal{Q})$. *Moreover, when this common optimal value is finite the two problems possess optimal solutions.*

The dual problem $(\mathcal{Q})$, like $(\mathcal{P})$, has its objective function only piecewise linear-quadratic in general. The problems enjoy a common expression, though, in terms of their *Lagrangian function*, which is defined by

$$L(u,v) := p{\cdot}u + \tfrac{1}{2}u{\cdot}Pu + q{\cdot}v - \tfrac{1}{2}v{\cdot}Qv - v{\cdot}Ru \quad \text{on } U \times V. \tag{2.8}$$

Indeed, one has from (2.1) and (2.6) that

$$f(u) = \sup_{v \in V} L(u,v), \qquad g(v) = \inf_{u \in U} L(u,v). \tag{2.9}$$

These formulas furnish *envelope representations* of the primal and dual objective functions: $f$ is expressed as the pointwise maximum of the collection of quadratic convex functions $L(\cdot, v)$ indexed by $v \in V$, while $g$ is expressed as the pointwise minimum of the collection of quadratic concave functions $L(u, \cdot)$ indexed by $u \in U$. They have strong potential for numerical developments, particularly in view of the following saddle point characterization of optimality in terms of the Lagrangian.

**Theorem 2.2** [3],[4]**.** *A vector pair* $(\bar{u}, \bar{v})$ *is a saddle point of the Lagrangian* $L(u,v)$ *relative to* $U \times V$ *if and only if* $\bar{u}$ *is an optimal solution to the primal problem* $(\mathcal{P})$ *and* $\bar{v}$ *is an optimal solution to dual problem* $(\mathcal{Q})$.

Especially attractive for numerical purposes is the *fully quadratic case* of $(\mathcal{P})$ and $(\mathcal{Q})$, by which we shall mean the case where *both* of the matrices $P$ and $Q$ are positive *definite*, not just semidefinite. Of course, models with singular matrices $P$ or $Q$ are very much of interest too for many applications. In particular, note should be taken of the situation where $P = 0$ and $Q = 0$, while $U$ and $V$ are cones (for instance the nonnegative orthants of $\mathbb{R}^n$ and $\mathbb{R}^m$). Then $(\mathcal{P})$ and $(\mathcal{Q})$ are classical linear programming problems dual to each other. A thorough analysis of the fully quadratic case is important even for such applications, however, because fully quadratic terms can be introduced iteratively as part of a solution scheme (see S6).

A fundamental connection with variational inequalities and linear complementarity is established when the saddle point optimality condition in Theorem 2.2 is written in terms of the derivatives of $L$. It is equivalent to

$$-\nabla_u L(\bar{u},\bar{v}) \in N_U(\bar{u}), \qquad \nabla_v L(\bar{u},\bar{v}) \in N_V(\bar{v}), \tag{2.10}$$

where $N_U(\bar{u})$ is the normal cone to $U$ at $\bar{u}$ in the sense of convex analysis [13],

$$N_U(\bar{u}) = \{w \in \mathbb{R}^n \,\big|\, w\cdot(u-\bar{u}) \le 0 \text{ for all } u \in U\},$$

and similarly for $N_V(\bar{v})$. The fact that $N_{U\times V}(\bar{u},\bar{v}) = N_U(\bar{u})\times N_V(\bar{v})$ gives us the following.

**Theorem 2.3.** *In terms of the mapping*

$$T(u,v) := (\nabla_u L(u,v), -\nabla_v L(u,v)) = (p + Pu - Rv, -q + R^T u + Qv)$$

*the necessary and sufficient condition for the optimality of $\bar{u}$ and $\bar{v}$ in Theorem 2.2 can be written equivalently as the linear variational inequality $-T(\bar{u},\bar{v}) \in N_{U\times V}(\bar{u},\bar{v})$, or in other words, the condition that*

$$(\bar{u},\bar{v}) \in U\times V, \text{ and } T(\bar{u},\bar{v})\cdot\big((u,v)-(\bar{u},\bar{v})\big) \le 0 \text{ for all } (u,v)\in U\times V.$$

Note that the symmetric part of the mapping T is given by diag $[P,Q]$, which is positive semidefinite. Thus T is a *monotone* mapping. In the fully quadratic case, T is *strongly monotone*.

The problem of solving the variational inequality in Theorem 2.3 turns into a linear complementarity problem in standard form precisely when $U$ and $V$ are the nonnegative orthants of $\mathbb{R}^n$ and $\mathbb{R}^m$. A conversion to this case is always possible in principle by writing the systems of linear inequalities that define $U$ and $V$ explicitly with their associated multipliers, but it could be unwise—in addition to entailing a possibly large increase in dimensionality.

The box-diagonal case illustrates this well. In that case, in order to handle a simple condition like $u_j^- \le u_j \le u_j^+$, with both bounds finite, one would presumably have first to introduce a change of variables so as to shift $u_j^-$ to 0. Then the upper inequality would have to be incorporated into the matrix part of the problem formulation along with a new dual variable $w_j \ge 0$. In parallel fashion, the condition $v_i^- \le v_i \le v_i^+$ would require the introduction of a new primal variable. All variables would end up merely as nonnegative variables, and an important aspect of the problem structure, the presence of both upper and lower bounds primally and dually, would be hidden and difficult to take advantage of. Furthermore, such emphasis on pushing the entire representation of a problem into a single matrix would disrupt natural relationships and forgo the use of positive features such as the ability to calculate $f(u)$ and $g(v)$ readily.

This seems counter to the philosophy one ought to be following with respect to large-scale models. Incentive is thereby added to the quest for iterative numerical methods which, unlike the ones surveyed in Lin and Pang [2], do not start off from a standard linear complementarity or quadratic programming formulation.

## 3. FEATURES AND ASSUMPTIONS FOR USE IN COMPUTATIONS

Although the applications of extended linear-quadratic programming that we are interested in accommodating involve cases of $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ where both $n$ and $m$ could be large, we have at our disposal the capability of solving *low-dimensional subproblems* of extended linear-quadratic programming by established techniques. As explained in detail in [4,S2], such a subproblem can always be converted in a certain way to an ordinary quadratic programming problem, which can then be solved by existing codes. (See Pang [14] for a survey of quadratic programming methods, and form more recent developments see Ye and Tse [15], Monteiro and Adler [16] and Goldfarb and Liu [17].) Transformation to a problem in linear complementarity is also available as an alternative, as just discussed, and in low dimensions, at least, this could work reasonably well. (The corresponding conversion of high-dimensional problems is unattractive because it would greatly increase dimensionality and obscure the given special structure.)

The possibility therefore emerges that a high-dimensional problem in extended linear-quadratic programming might be solved by some algorithm based on solving a sequence of low-dimensional problems. This will be amplified in S4 and made the basis of our "finite-envelope methods" in S5. To begin with, however, we introduce and discuss for the high-dimensional case *three assumptions which will hold for the rest of this paper.*

These assumptions fit a very large class of applications with dynamic or stochastic structure, as formulated in Rockafellar and Wets [6]. They are relevant not only to the particular algorithms developed in S5, but others that aim at exploiting such structure; see for instance Zhu and Rockafellar [18].

**Finite Monitoring Assumption.** *The functions $\rho_{QV}$ and $\rho_{UP}$ are finite everywhere; they do not take on $\infty$. Thus the objective functions $f$ in $(\mathcal{P})$ and $g$ in $(\mathcal{Q})$ are finite everywhere, and there are no (strictly enforced) constraints implicit in these problems beyond $u \in U$ and $v \in V$.*

This is definitely satisfied in the fully quadratic case of problems $(\mathcal{P})$ and $(\mathcal{Q})$ (where $P$ and $Q$ are both nonsingular). It is satisfied also, regardless of the nature of $P$ or $Q$, when the sets $U$ and $V$ are bounded. In finer detail, as proved in [3, Proposition 2.4], the assumption holds if and only if

$$\text{(3.1)} \qquad \operatorname{nl} Q \cap \operatorname{rc} V = \{0\}, \qquad \operatorname{nl} P \cap \operatorname{rc} U = \{0\},$$

where "nl" indicates the null space (kernel) of a matrix and "rc" indicates the recession cone of a convex set (in the sense of [12,S8]).

The Finite Monitoring Assumption merely acts as a weak normalization of the mathematical model being used in a given application. If it were not satisfied, we could, for one thing, just truncate the sets $U$ and $V$. Provided that the truncated sets contained a saddle point $(\bar{u}, \bar{v})$ for $L$ relative to the original $U$ and $V$, any solutions to the truncated problems

would be solutions to the original problem [3, Theorem 2.5]. The truncation would thus be harmless.

Another way of making sure that the Finite Monitoring Assumption is satisfied is to add strictly quadratic "proximal terms" to $L$. Such a procedure can always be combined iteratively with any other computational scheme as an "outer algorithm" for which there is considerable theoretical backing [10], [11], [12]. This approach will be discussed in S6.

In short, the Finite Monitoring Assumption is natural for the context of large-scale extended linear-quadratic programming and in the long run does not entail any substantial loss of generality. It provides useful focus for a discussion of computational possibilities.

By virtue of the Finite Monitoring Assumption, we can write

$$f(u) = \max_{v \in V} L(u,v), \qquad g(v) = \min_{u \in U} L(u,v), \tag{3.2}$$

where the "max" and "min" indicate attainment in the formulas of (2.9). Correspondingly we have *nonempty* point sets

$$F(u) = \operatorname*{argmax}_{v \in V} L(u,v), \qquad G(v) = \operatorname*{argmin}_{u \in U} L(u,v), \tag{3.3}$$

that depend on $u$ and $v$. The set-valued mappings $F : u \mapsto F(u)$ and $G : v \mapsto G(v)$ will be of considerable interest to us. The attainment of the inf and sup in (2.9) stems, of course, from the fact that one is solving quadratic programming subproblems in these formulas. Such a subproblem has an optimal solution whenever its optimal value is finite.

Because there are no constraints being represented implicitly by infinite values for $f(u)$ or $g(v)$, the feasible solutions to $(\mathcal{P})$ are precisely the vectors $u \in U$, and the feasible solutions to $(\mathcal{Q})$ are precisely the vectors $v \in V$. Further, we can conclude from Theorem 2.1 that

$$\infty > \min(\mathcal{P}) = \max(\mathcal{Q}) > -\infty. \tag{3.4}$$

Contrasting with these advantageous properties of problems $(\mathcal{P})$ and $(\mathcal{Q})$, however, is the noted circumstance that the objective expressions $f(u)$ and $g(v)$ are only *piecewise* linear-quadratic. It is here that the second assumption becomes important and offers compensating opportunities.

**Double Decomposability Assumption.** *For each $u \in U$ it is readily possible to calculate the value $f(u)$ in (3.2) and at least one of the elements of the set $F(u)$ in (3.3). Similarly, for each $v \in V$ it is readily possible to calculate the value $g(v)$ and at least one of the elements of $G(v)$.*

To see how this assumption may well be satisfied despite $U$ and $V$ being high-dimensional, one need only look at the box-diagonal case in the notation of (2.2) and (2.3). There, for

fixed $u \in U$, the problem of maximizing $L(u,v)$ over $v \in V$ decomposes into separate problems in the individual coordinates: for $i = 1, \ldots, m$ one must

$$\text{(3.5)} \qquad \text{maximize} \quad \left[ q_i - \sum_{j=1}^{n} r_{ij} u_j \right] \cdot v_i - \tfrac{1}{2} \beta_i v_i^2 \text{ subject to } v_i^- \leq v_i \leq v_i^+.$$

In parallel fashion, the problem of minimizing $L(u,v)$ over $u \in U$ for given $v \in V$ reduces to the separate problems

$$\text{(3.6)} \qquad \text{minimize} \quad \left[ p_j - \sum_{i=1}^{m} v_i r_{ij} \right] \cdot u_i + \tfrac{1}{2} \alpha_j u_j^2 \text{ subject to } u_j^- \leq u_j \leq u_j^+.$$

These one-dimensional subproblems have very simple *closed-form solutions* in the fully quadratic case, where $\alpha_j > 0$ and $\beta_i > 0$.

Even without complete separability such as in the box-diagonal case, it may be possible nonetheless to break down the maximization of $L(u,v)$ in $v$, or its minimization in $u$, into relatively small subproblems that can be solved by parallel processing. Dynamical structure, for example, always gives such a breakdown into separate problems for each instant of time, cf. [3], [6]. Thus, multistage problems of optimization in which the individual stages are low-dimensional primally and dually can always be seen as satisfying the Double Decomposability Assumption when viewed in the right light. This is a different approach to taking advantage of dynamical structure than the conventional one of looking for a "staircase" pattern in some large matrix, and it opens routes to computation that have so far not been explored.

In specific areas of application there are additional features to the postulated decomposability beyond what might at first be apparent. Roughly, these have to do with the way a fixed $v$ is substituted into the expression $L(u,v)$ to get a function of $u$, and vice versa. In principle, this is seen to involve multiplying the matrix $R$ on the left by the vector $v$, among other things, but in the models set up in terms of finite-difference representations of differential operators, for instance in optimal control, it is precisely at this point that "state variables" can be brought in and traditional schemes of "integration" can be incorporated, with much saving of effort. For more on this, see [6].

It should be noted, finally, that the Double Decomposability Assumption does not actually presuppose that both the $u$-dimension and the $v$-dimension are large. As a special case, if only the $v$-dimension is large and the $u$-dimension is small (which typically holds in two-stage stochastic programming, cf. [4], [5]), one can minimize $L(u,v)$ readily in $u \in U$ for each $v$ even if there is no separability in $u$, as such.

**Line Searchability Assumption.** *It is possible to minimize $f(u)$ over any line segment joining two points in $U$, and likewise, to maximize $g(v)$ over any line segment joining two points in $V$.*

This property is already implied, more or less, by the two preceding assumptions, but we introduce it as an assumption anyway for clarity. Observe that it is not quite the same as

supposing we can readily perform a line search over a *half-line* (unbounded line segment) in which the intersection of the half-line with the boundary of the feasible set might have to be detected as part of the search—a potentially harder task. We shall not go into any details here about ways of carrying out the line searches in the Line Searchability Assumption, but certainly the special nature of $f$ and $g$ could have influence. Besides $f$ being convex and $g$ being concave, there is the fact that both functions are piecewise linear-quadratic.

There may be some hope not only of optimizing $f$ and $g$ over line segments, as part of a broader procedure, but over triangles and other low-dimensional polytopes. We shall return to this issue at the end of S6, when more background will be available for the issues.

The Double Decomposability Assumption speaks of calculating a vector $v \in F(u)$ for any choice of $u \in U$, or a vector $u \in G(v)$ for any choice of $v \in V$. Calculations need not be limited to one such stage, however. Elements of the sets $G(F(u))$, $F(G(v))$, $F(G(F(u)))$, $G(F(G(V)))$, and so forth, can be likewise determined on the basis of the assumption, if one wishes to do so. Such vectors all provide information about the functions $f$ and $g$ that might be useful in solving $(\mathcal{P})$ and $(\mathcal{Q})$. *The challenge is to design methods that make good use of this kind of information.* The following results provide theoretical support.

An elementary property can immediately be stated in terms of subgradients. Recall that for the convex function $f$, the subgradient set $\partial f(u^*)$ at any point $u^*$ consists of the vectors $w$ satisfying

$$f(u) \geq f(u^*) + w \cdot (u - u^*) \text{ for all } u \in \mathbb{R}^n,$$

whereas for the concave function $g$ the set $\partial g(v^*)$ at any point $v^*$ consists of the vectors $z$ satisfying

$$g(v) \leq g(v^*) + z \cdot (v - v^*) \text{ for all } v \in \mathbb{R}^m.$$

These sets reduce to singletons if and only if $f$ is differentiable at $u^*$ and $g$ is differentiable at $v^*$, in which event the singleton elements are in fact the gradients $\nabla f(u^*)$ and $\nabla g(v^*)$; see [13,S25].

**Proposition 3.1.** *Let $v^{**} \in F(u^*)$ and $u^{**} \in G(v^*)$ for arbitrary vectors $u^*$ and $v^*$. Then*

$$L(u, v^{**}) \leq f(u) \text{ for all } u, \text{ with } L(u^*, v^{**}) = f(u^*) \text{ and } \nabla_u L(u^*, v^{**}) \in \partial f(u^*).$$

*Similarly,*

$$L(u^{**}, v) \geq g(v) \text{ for all } v, \text{ with } L(u^{**}, v^*) = g(v^*) \text{ and } \nabla_v L(u^{**}, v^*) \in \partial g(v^*).$$

**Proof.** For any $v \in V$ one would have $L(u, v) \leq f(u)$ for all $u$ by (3.2). In the case of $v = v^{**} \in F(u^*)$ this inequality holds as an equation at $u = u^*$ by the definition (3.3) of $F(u^*)$. Any subgradient of the convex function $L(\cdot, v^{**})$ at $u^*$ is then in particular a subgradient of $f$ at $u^*$. But the unique subgradient of $L(\cdot, v^{**})$ at $u^*$ is $\nabla_u L(u^*, v^{**})$, because $L$ is a differentiable function. The verification of the facts about $g$ proceeds along the same lines. □

The fully quadratic case is particularly attractive in this respect.

**Proposition 3.2.** *In the fully quadratic case, the functions $f$ and $g$ are continuously differentiable everywhere, and the mappings $F$ and $G$ are single-valued. Then*

$$\nabla f(u^*) = \nabla_u L(u^*, F(u^*)) = p + Pu^* - R^T F(u^*) \text{ at any } u^*,$$

*whereas*

$$\nabla g(v^*) = \nabla_v L(G(v^*), v^*) = q - RG(v^*) - Qv^* \text{ at any } v^*.$$

**Proof.** The positive definiteness of $P$ and $Q$ ensures that $L(u, v)$ is strictly convex in $u$ and strictly concave in $v$, so the sets $F(u)$ and $G(v)$ in (3.3) are always singletons. From the original statement of problem $(\mathcal{P})$ it is clear that the continuous differentiability of $f$ hinges on that of the function $\rho_{VQ}$. The latter is seen by definition (2.1) to be the conjugate of the closed, proper, convex function

$$\varphi_{VQ}(v) = \begin{cases} \frac{1}{2} v \cdot Qv & \text{if } v \in V, \\ \infty & \text{if } v \notin V, \end{cases} \tag{3.7}$$

inasmuch as

$$\rho_{VQ}(w) = \sup_{v \in \mathbb{R}^m} \{v \cdot w - \varphi_{VQ}(v)\} =: \varphi^*_{VQ}(w). \tag{3.8}$$

The function $\varphi_{VQ}$ being strictly convex on its effective domain $V$, its conjugate is differentiable [13, Theorem 26.3], hence by convexity continuously differentiable [13, Corollary 25.5.1]. Thus $\rho_{VQ}$ and $f$ are continuously differentiable. The argument for the continuous differentiability of $g$ is parallel. □

In the fully quadratic case one has in addition, obviously, that *$f$ is strictly convex and $g$ is strictly concave*. Then $(\mathcal{P})$ and $(\mathcal{Q})$ have unique optimal solutions. In face, global Lipschitz constants can be derived in this case for the single-valued mappings $F$ and $G$, as will be seen below. In spite of the first-order smoothness in Proposition 3.2, though, there may be jumps in the second derivatives of $f$ and $g$, so one is still essentially in a context of nonsmooth optimization.

Although these agreeable properties may fail outside the fully quadratic case of $(\mathcal{P})$ and $(\mathcal{Q})$, this does not have to be an impediment in the long run, since for computational purposes, strictly quadratic terms can be added to the objectives in $(\mathcal{P})$ and $(\mathcal{Q})$ as part of an interactive process (S6) and the properties are then available.

We conclude this section by recording a fundamental property of problems $(\mathcal{P})$ and $(\mathcal{Q})$ which is useful in formulating stopping criteria for approximate methods of solution. To state it well, we use the notation

$$\|u\|_P = [u \cdot Pu]^{1/2}, \qquad \|v\|_Q = [v \cdot Qv]^{1/2}. \tag{3.9}$$

These expressions are not actually norms, of course, except in the fully quadratic case, where $P$ and $Q$ are positive definite.

**Proposition 3.3.** *Let $\bar{u}$ and $\bar{v}$ be optimal solutions (not necessarily unique) to $(\mathcal{P})$ and $(\mathcal{Q})$, and let $u^*$ and $v^*$ be feasible solutions (i.e., elements of $U$ and $V$ respectively) that for a certain $\varepsilon$ satisfy $f(u^*) - g(v^*) \le \varepsilon$. Then $u^*$ and $v^*$ are $\varepsilon$-optimal in the sense that*

$$|f(u^*) - f(\bar{u})| \le \varepsilon, \qquad |g(u^*) - g(\bar{v})| \le \varepsilon,$$

*and furthermore*

$$\|u^* - \bar{u}\|_P \le [2\varepsilon]^{1/2}, \qquad \|v^* - \bar{v}\|_Q \le [2\varepsilon]^{1/2}.$$

**Proof.** The inequalities for $f$ and $g$ are evident from the chain

$$f(u^*) \ge f(\bar{u}) = \min(\mathcal{P}) = \max(\mathcal{Q}) = g(\bar{v}) \ge g(v^*),$$

which holds by Theorem 2.1. The optimality of $\bar{u}$ and $\bar{v}$ is equivalent by Theorem 2.2 to the saddle point condition that $L(u, \bar{v}) \ge L(\bar{u}, \bar{v}) \ge L(\bar{u}, v)$ for all $u \in U$, $v \in V$. Here $f(\bar{u}) = L(\bar{u}, \bar{v}) = g(\bar{v})$. Then

$$\begin{aligned} f(u^*) - L(\bar{u}, \bar{v}) &\ge L(u^*, \bar{v}) - L(\bar{u}, \bar{v}) \\ &= \nabla_u L(\bar{u}, \bar{v}) \cdot (u^* - \bar{u}) + \tfrac{1}{2}(u^* - \bar{u}) \cdot P(u^* - \bar{u}). \end{aligned} \tag{3.10}$$

Inasmuch as $\bar{u}$ minimizes $L(u, \bar{v})$ over $u \in U$, we have $\nabla_u L(\bar{u}, \bar{v}) \cdot (u - \bar{u}) \ge 0$ for all $u \in U$. Applying this in (3.10) we obtain

$$f(u^*) - L(\bar{u}, \bar{v}) \ge \tfrac{1}{2}(u^* - \bar{u}) \cdot P(u^* - \bar{u}) = \tfrac{1}{2}\|u^* - \bar{u}\|_P^2.$$

By a parallel argument,

$$g(v^*) - L(\bar{u}, \bar{v}) \le -\tfrac{1}{2}(v^* - \bar{v}) \cdot Q(v^* - \bar{v}) = -\tfrac{1}{2}\|v^* - \bar{v}\|_Q^2.$$

We subtract this inequality from the preceding one to get

$$\tfrac{1}{2}\|u^* - \bar{u}\|_P^2 + \tfrac{1}{2}\|v^* - \bar{v}\|_Q^2 \le f(u^*) - g(v^*) \le \varepsilon.$$

This yields the final pair of estimates in the proposition. □

Proposition 3.3 underlines a valuable consequence of the fundamental symmetry between our primal and dual problems. It is possible to work on solving both problems simultaneously and in this way not only maintain current candidates for both primal and dual optimal solutions but also have running estimates on how far these are from true optimality.

## 4. ENVELOPE SUBPROBLEMS.

The possibility of approaching $(\mathcal{P})$ and $(\mathcal{Q})$ computationally in terms of some sequence of low-dimensional subproblems has already been suggested. We now look at this more closely to see what it might entail. The idea is that of replacing $U$ and $V$ by special choices of polyhedral convex sets $U^\nu \subset U$ and $V^\nu \subset V$ for $\nu = 1, 2, \ldots$ and solving the corresponding primal and dual problems

$$(\mathcal{P}^\nu) \quad \text{minimize } f^\nu(u) \text{ over } u \in U^\nu, \text{ with } f^\nu(u) = p{\cdot}u + \tfrac{1}{2}u{\cdot}Pu + \rho_{V^\nu Q}(q - Ru),$$
$$(\mathcal{Q}^\nu) \quad \text{maximize } g^\nu(v) \text{ over } v \in V^\nu, \text{ with } g^\nu(v) = q{\cdot}v + \tfrac{1}{2}v{\cdot}Qv + \rho_{U^\nu P}(R^T v - p).$$

Solving such a pair of subproblems, which again belong to the category of extended linear-quadratic programming as outlined in S2, corresponds to finding a saddle point of $L(u, v)$ relative to $U^\nu \times V^\nu$ instead of $U \times V$.

In a moment it will be clearer how this idea might be realized in an essentially low-dimensional manner. Let us note first a basic relationship between the objective functions in these problems.

**Proposition 4.1.** *For any choice of polyhedral convex sets $U^\nu \subset U$ and $V^\nu \subset V$, one has*

$$f(u) \geq f^\nu(u) \text{ for all } u, \text{ with equality when } F(u) \cap V^\nu \neq \emptyset,$$
$$g(v) \leq g^\nu(v) \text{ for all } v, \text{ with equality when } G(v) \cap U^\nu \neq \emptyset.$$

**Proof.** Alongside of the formulas (3.2) and (3.3) for $f$, $g$, $F$, and $G$ in terms of the Lagrangian $L$ we have

$$f^\nu(u) = \max_{v \in V^\nu} L(u, v), \qquad g^\nu(v) = \min_{u \in U^\nu} L(u, v). \tag{4.1}$$

From this and the assumption that $U^\nu$ and $V^\nu$ are subsets of $U$ and $V$, the claims are evident. □

The inequalities in Proposition 4.1 suggest the term *envelope subproblems* for $(\mathcal{P}^\nu)$ and $(\mathcal{Q}^\nu)$. The objective function $f^\nu$ is a lower envelope replacement for $f$ which agrees with it at all points $u$ for which the chosen set $V^\nu$ contains an element of $F(u)$. Likewise, $g^\nu$ is an upper envelope replacement for $g$ that agrees with it at all points $v$ for which $U^\nu$ contains an element of $G(v)$.

There is no simple relationship in general between the common optimal value in $(\mathcal{P}^\nu)$ and $(\mathcal{Q}^\nu)$ and the one in $(\mathcal{P})$ and $(\mathcal{Q})$. The replacement of $f$ by the envelope $f^\nu$ would tend to lower the optimal value from the one in $(\mathcal{P})$, but the simultaneous replacement of $U$ in the minimization by the smaller set $U^\nu$ would tend to raise it. This complicates the analysis of any computational approach making use of a sequence of envelope subproblems, but it is not an ultimate obstacle, as we shall demonstrate in S5.

How actually can the envelope subproblems be solved, considering that they lie in spaces that may intrinsically be high-dimensional? Even if the sets $U^\nu$ and $V^\nu$ are themselves low-dimensional within these spaces, doesn't some difficulty with projections, say, take over? This is a critical juncture where the *extended linear-quadratic* nature of the given problems $(\mathcal{P})$ and $(\mathcal{Q})$ re-enters.

Let us suppose that $U^\nu$ and $V^\nu$ are *polytopes*, i.e., sets represented as the convex hulls of finite collections of points:

$$U^\nu = \mathrm{co}\{u_k^\nu \,\big|\, k = 1, \dots, n^\nu\}, \qquad V^\nu = \mathrm{co}\{v_l^\nu \,\big|\, l = 1, \dots, m^\nu\}. \tag{4.2}$$

A general element of $U^\nu$ can be represented as a convex combination

$$u = \sum_{k=1}^{n^\nu} \xi_k u_k^\nu \text{ with } \xi = (\xi_1, \dots, \xi_{n^\nu}) \in S(n^\nu), \tag{4.3}$$

where $S(n^\nu)$ is the unit simplex in $\mathbb{R}^{n^\nu}$ (consisting of the nonnegative vectors whose coordinates add up to 1). Likewise, a general element of $V^\nu$ has the form

$$v = \sum_{l=1}^{m^\nu} \eta_l v_l^v \text{ with } \eta = (\eta_1, \dots, \eta_{m^\nu}) \in S(m^\nu). \tag{4.4}$$

This seems of little help if we think of minimizing $f^\nu(u)$ directly over $u \in U^\nu$ or maximizing $g^\nu(v)$ over $v \in V^\nu$, but in the Lagrangian framework an important simplification occurs. There we can express

$$L(u, v) = L\left(\sum_{k=1}^{n^\nu} \xi_k u_k^\nu,\ \sum_{l=1}^{m^\nu} \eta_l v_l^\nu\right) = \Lambda^\nu(\xi, \eta) \tag{4.5}$$

for a certain *quadratic form* $\Lambda^\nu$ *with explicit coefficients in the* $n^\nu + m^\nu$ *variables* $\xi_k$ *and* $\eta_l$. Specifically, we have

$$\Lambda^\nu(\xi, \eta) = p^\nu \cdot \xi + \tfrac{1}{2}\xi \cdot P^\nu \xi + q^\nu \cdot \eta - \tfrac{1}{2}\eta \cdot Q^\nu \eta - \eta \cdot R^\nu \xi, \tag{4.6}$$

where $p_k^\nu = p \cdot u_k^\nu$, $P_{ik}^\nu = u_i^\nu \cdot P u_k^\nu$, $q_l^\nu = q \cdot v_l^\nu$, $Q_{jl}^\nu = v_j^\nu \cdot Q v_l^\nu$, and $R_{lk}^\nu = v_l^\nu = v_l^\nu \cdot R u_k^\nu$. We summarize this observation in the next proposition.

**Proposition 4.2.** *Solving the envelope subproblems* $(\mathcal{P}^\nu)$ *and* $(\mathcal{Q}^\nu)$ *in the case where* $U^\nu$ *and* $V^\nu$ *are polytopes expressed as in (4.2) is equivalent to finding a saddle point* $(\bar\xi^\nu, \bar\eta^\nu)$ *of the function* $\Lambda^\nu$ *in (4.6) relative to the simplex product* $S(n^\nu) \times S(m^\nu)$, *and then setting*

$$\bar u^\nu = \sum_{k=1}^{n^\nu} \bar\xi_k^\nu u_k^\nu, \qquad \bar v^\nu = \sum_{l=1}^{m^\nu} \bar\eta_l^\nu v_l^\nu,$$

*to get the desired optimal solutions $\bar{u}^\nu$ and $\bar{v}^\nu$. This reduced saddle point subproblem corresponds to a special primal-dual pair of extended linear-quadratic programming problems in $\mathbb{R}^{n^\nu}$ and $\mathbb{R}^{m^\nu}$. Thus it can be kept low-dimensional by restricting the numbers $n^\nu$ and $m^\nu$ of elements in the expressions (4.6).*

An additional fact about this reduction is that in applications such as to stochastic programming and optimal control, the coefficients incorporated in the quadratic form $\Lambda^\nu(\xi, \eta)$ can be generated by computing certain expectations and variances or integrating dynamical equations [4], [6]. Thus, the reduced form can have meaning appropriate to the context of the given problem.

The reader should note that the reduced subproblems in Proposition 4.2 are examples of extended linear-quadratic programming where the monitoring functions are definitely not penalty functions. This is because the simplexes $S(n^\nu)$ and $S(m^\nu)$ do not contain the origins of their respective spaces, so the functions take on negative as well as positive values. These subproblems can be solved in a practical manner by the transformations mentioned at the beginning of S3. Possibly they could be solved more efficiently by some specialized method yet in the offing.

A brief inspection of the extended linear programming case where $P = 0$ and $Q = 0$ will shed more light on the envelope subproblems and show their connection to cutting-plane methods and other well understood techniques of nonsmooth optimization. In this case $f$ and $g$ are piecewise linear. The formulas in (4.1) reduce under the substitutions in (4.3) and (4.4) to

$$f^\nu(u) = \max_{l=1,\ldots,m^\nu} L(u, v_l^\nu), \qquad g^\nu(v) = \min_{k=1,\ldots,n^\nu} L(u_k^\nu, v). \tag{4.7}$$

Thus, in this linear case, the envelope $f^\nu$ is the pointwise maximum of a finite collection of "lower affine approximants" to $f$, while $g^\nu$ similarly involves "upper affine approximants" to $g$.

In the general instance of problems $(\mathcal{P})$ and $(\mathcal{Q})$, of course, the envelope functions are *not* just piecewise linear or expressible as a finitary max or min but have nontrivial quadratic pieces as well. These can be viewed as affording a smoothing effect to counteract the "corners" present in the linear case in (4.7).

## 5. FINITE-ENVELOPE METHODS

There are many conceivable ways of utilizing numerically the special problem structure we have described in $(\mathcal{P})$ and $(\mathcal{Q})$, but we shall focus now on a particular class of methods. These will be called *finite-envelope* methods, because they proceed in terms of envelope subproblems $(\mathcal{P}^\nu)$ and $(\mathcal{Q}^\nu)$ in which the sets $U^\nu$ and $V^\nu$ are finitely generated as polytopes, thereby fitting the pattern in Proposition 4.2. They can be viewed as generalizations of cutting-plane methods through the connection noted at the end of S4 in the linear case, but the envelope functions they work with are typically not just piecewise linear. Both primal and dual envelopes are present at all times and are able to "communicate with each other" in the solution process.

A basic conceptual algorithm will serve to define the class of finite-envelope methods in general. Different implementations of the algorithm will give specific methods within the class. The main source of flexibility in implementation is in the choice and updating of the finite point sets whose convex hulls are taken to be the polytopes $U^\nu$ and $V^\nu$ in iteration $\nu$. Our goal at this stage is to establish minimal conditions on these sets that ensure the algorithm's convergence. The three assumptions introduced in S3 remain in force.

**Basic Finite-Envelope Algorithm.** *Starting from any choice of elements $\bar{u}_0^0 \in U$ and $\bar{v}_0^0 \in V$, sequences $\{\bar{u}_0^\nu\} \subset U$ and $\{\bar{v}_0^\nu\} \subset V$ are constructed by some implementation of the following steps.*

Step 1 (Optimality Test). *Set $\varepsilon^\nu = f(\bar{u}_0^\nu) - g(\bar{v}_0^\nu)$; this is the current duality gap. If $\varepsilon^\nu$ is sufficiently small, terminate. Otherwise proceed with Step 2.*

Step 2 (Envelope Generation). *For an integer $l \geq 2$ (which for simplicity is taken here to be fixed in advance), calculate elements*

$$\bar{u}_k^\nu \in G(\bar{v}_{k-1}^\nu) \text{ and } \bar{v}_k^\nu \in F(\bar{u}_{k-1}^\nu) \text{ for } k = 1, \ldots, l. \tag{5.1}$$

*Select polytopes $U^\nu \subset U$ and $V^\nu \subset V$ such that*

$$\{\bar{u}_0^\nu, \bar{u}_1^\nu, \ldots, \bar{u}_l^\nu\} \subset U^\nu, \qquad \{\bar{v}_0^\nu, \bar{v}_1^\nu, \ldots, \bar{v}_l^\nu\} \subset V^\nu. \tag{5.2}$$

Step 3 (Envelope Subproblems). *Determine a saddle point $(\bar{u}^\nu, \bar{v}^\nu)$ of $L(u, v)$ relative to $U^\nu \times V^\nu$.*

Step 4 (Line Search). *Minimize $f(u)$ over the line segment from $\bar{u}_0^\nu$ to $\bar{u}^\nu$ to get $\bar{u}_0^{\nu+1}$ and $f(\bar{u}_0^{\nu+1})$. Maximize $g(v)$ over the line segment from $\bar{v}_0^\nu$ to $\bar{v}^\nu$ to get $\bar{v}_0^{\nu+1}$ and $g(\bar{v}_0^{\nu+1})$. Return to Step 1.*

The optimality test in Step 1 of the algorithm relies on the estimates furnished by Proposition 3.3. One can terminate as soon as $\varepsilon^\nu \leq \varepsilon$ for a predetermined value of $\varepsilon$, for instance, and know that the current vectors $\bar{u}_0^\nu$ and $\bar{v}_0^\nu$ will be $\varepsilon$-optimal. More broadly, though, Step 1 is a reminder that the methods in question have the property of providing

us at all times, by way of Proposition 3.3 and the value $\varepsilon^\nu$, with a measure of how far the current values $f(\bar{u}_0^\nu)$ and $g(\bar{v}_0^\nu)$ and solution candidates $\bar{u}_0^\nu$ and $\bar{v}_0^\nu$ are from optimality.

In fact the duality gap $\varepsilon^\nu$ can never increase, only decrease (or perhaps sometimes stay the same) from one iteration to the next, because of the line searches in Step 4—at least under the theoretical assumption, which we make here for simplicity, that the line searches can be executed with full precision. Finite-envelope methods therefore always have the "fail-safe" property of never making things worse. They allow their progress readily to be measured, and in any case they never generate sequences that might diverge from optimality.

The main task, as made apparent by these observations, is to identify schemes for generating the envelopes in Step 2 that ensure the convergence of $\varepsilon^\nu$ to 0 and, even better, provide a good rate for the convergence. There is wide territory for exploration, and only some initial results can be presented in this paper. A further issue, always to be kept in mind, is that of finding the most efficient approach to the saddle point subproblems in Step 3. Although a couple of practical techniques are already available (via Proposition 4.2 and the discussion at the beginning of S3), there is little reason to suppose there aren't better ones still waiting to be discovered. The best design of a line search routine to be used in the context of Step 4 remains open, too.

Turning our attention now fully to the matter of envelope generation in Step 2, we note that formula (5.1) ensures by Proposition 4.1 that

$$f(\bar{u}_k^\nu) = f^\nu(\bar{u}_k^\nu), \qquad g(\bar{v}_k^\nu) = g^\nu(\bar{v}_k^\nu) \quad \text{for } k = 0, 1, \ldots, l-1. \tag{5.3}$$

In particular then, since $l \geq 2$, the primal and dual envelope representations are always exact at the current points $\bar{u}_0^\nu$ and $\bar{v}_0^\nu$ as well as at $\bar{u}_1^\nu$ and $\bar{v}_1^\nu$.

The points $\bar{u}_k^\nu$ and $\bar{v}_k^\nu$ for $k = 0, 1, \ldots, l$ do not necessarily have to be included *directly* among the ones chosen for the convex hull representations (4.2) of $U^\nu$ and $V^\nu$, though. One can anticipate special cases where $U$, but not $V$, say, happens already to be a low-dimensional polytope. Then one could apply the algorithm simply with $U^\nu = U$ for all $\nu$, and the current values of $\bar{u}_0^\nu$, $\bar{u}_1^\nu$,..., $\bar{u}_l^\nu$, would play no role in envelope generation except through their influence on the calculation of $\bar{v}_k^\nu \in F(\bar{u}_{k-1}^\nu)$. A further possibility is that of augmenting the points generated by (5.1) by additional ones kept or derived from previous iterations and thereby serving as a memory bank for what has already been learned about $f$ and $g$.

In order to be sure that $\varepsilon^\nu$ will be replaced by a strictly lower value $\varepsilon^{\nu+1}$, we need to know that at least one of the line searches in Step 4 will yield an improvement. In other words, we need to know either that $\bar{u}^\nu - \bar{u}_0^\nu$ gives a direction of descent for $f$ at $\bar{u}_0^\nu$ relative to $U$, or that $\bar{v}^\nu - \bar{v}_0^\nu$ gives a direction of ascent for $g$ at $\bar{v}_0^\nu$ relative to $V$. Thus $(\mathcal{P}^\nu)$ and $(\mathcal{Q}^\nu)$ must operate effectively as *direction-finding* subproblems for improvements in the current vectors $\bar{u}_0^\nu$ and $\bar{v}_0^\nu$. (A vector $w$ gives a direction of descent for $f$ at $\bar{u}_0^\nu$ relative to $U$ if and only if $\bar{u}_0^\nu + tw$ belongs to $U$ for all $t > 0$ sufficiently small, and $f'(\bar{u}_0^\nu; w) < 0$.

Here $f'$ denotes one-sided directional derivatives of the convex function $f$ as in [13,S23].) A first result is the following.

**Proposition 5.1.** *Suppose in Step 2 of the basic algorithm that actually $F(\bar{u}_0^\nu) \subset V^\nu$ and $G(\bar{v}_0^\nu) \subset U^\nu$ (instead of just that the sets $F(\bar{u}_0^\nu) \cap V^\nu$ and $G(\bar{v}_0^\nu) \cap U^\nu$ contain vectors $\bar{u}_1^\nu$ and $\bar{v}_1^\nu$). Then, unless $\bar{u}_0^\nu$ happens already to be optimal for the subproblem $(\mathcal{P}^\nu)$ and can thus be taken as $\bar{u}^\nu$, the vector $\bar{u}^\nu - \bar{u}_0^\nu$ gives in $(\mathcal{P})$ a direction of descent for $f$ at $\bar{u}_0^\nu$ relative to $U$. Likewise, unless $\bar{v}_0^\nu$ happens already to be optimal for the subproblem $(\mathcal{Q}^\nu)$, the vector $\bar{v}^\nu - \bar{v}_0^\nu$ gives in $(\mathcal{Q})$ a direction of ascent for $g$ at $\bar{v}_0^\nu$ relative to $V$.*

**Proof.** We need only deal with the primal; the argument for the dual is analogous. When $\bar{u}_0^\nu$ is not already optimal for $(\mathcal{P}^\nu)$, the vector $\bar{u}^\nu - \bar{u}_0^\nu$ surely gives a direction of descent for $f^\nu$ at $\bar{u}_0^\nu$ relative to $U^\nu$. This is then also a direction of descent for $f^\nu$ at $\bar{u}_0^\nu$ relative to $U$, of course, because $U^\nu \subset U$. To establish that it is at the same time a direction of descent for $f$ at $\bar{u}_0^\nu$ relative to $U$, it will suffice to show that $f$ and $f^\nu$ have the same directional derivatives at $\bar{u}_0^\nu$. This amounts to showing that the convex functions $\rho_{VQ}$ and $\rho_{V^\nu Q}$ have the same directional derivatives at the point $z^\nu = q - R\bar{u}_0^\nu$, since it is only in these monitoring expressions that $f$ and $f^\nu$ differ.

We have seen in (3.7)-(3.8) that $\rho_{VQ}$ is conjugate to a certain function $\varphi_{VQ}$; in like manner, $\rho_{V^\nu Q}$ is conjugate to $\varphi_{V^\nu Q}$. We know from these formulas and general rules of convex analysis [13, Theorem 23.5] that the subgradients of these functions are given by

$$\begin{aligned}\partial\rho_{VQ}(z^\nu) &= \operatorname*{argmax}_{v\in\mathbb{R}^m}\{v\cdot z^\nu - \varphi_{VQ}(v)\} = \operatorname*{argmax}_{v\in\mathbb{R}^m}\{v\cdot[q - R\bar{u}_0^\nu] - \varphi_{VQ}(v)\}\\ &= \operatorname*{argmax}_{v\in V}\{v\cdot[q - R\bar{u}_0^\nu] - \tfrac{1}{2}v\cdot Qv\} = \operatorname*{argmax}_{v\in V} L(\bar{u}_0^\nu, v) =: F(\bar{u}_0^\nu).\end{aligned}$$

By the same token,

$$\partial\rho_{V^\nu Q}(z^\nu) = \operatorname*{argmax}_{v\in V^\nu} L(\bar{u}_0^\nu, v).$$

Inasmuch as the set $F(\bar{u}_0^\nu)$ is by hypothesis entirely contained in the subset $V^\nu$ of $V$, the "argmax" set in the second case must be the same as the one in first case, i.e., again equal to $F(\bar{u}_0^\nu)$. We conclude that $\partial\rho_{VQ}(z^\nu) = \partial\rho_{V^\nu Q}(z^\nu)$. The fact that the two monitoring functions (which are *finite* convex functions under the Finite Monitoring Assumption in S3) have the same subgradients at $z^\nu$ implies that they have the same directional derivatives at $z^\nu$, as needed [13, Theorem 23.4]. □

A simple case where the hypothesis of Proposition 5.1 is fulfilled is the one where $F(\bar{u}_0^\nu)$ and $G(\bar{v}_0^\nu)$ consist of single elements, as is true in particular in the fully quadratic case in Proposition 3.2. Then additional conclusions can be drawn.

**Theorem 5.2.** *Suppose that both $F(\bar{u}_0^\nu)$ and $G(\bar{v}_0^\nu)$ happen to be singletons, i.e., that the max and min in their respective definitions are uniquely attained. Then $f$ and $f^\nu$ are differentiable at $\bar{u}_0^\nu$ with $\nabla f^\nu(\bar{u}_0^\nu) = \nabla f(\bar{u}_0^\nu)$, and $g$ and $g^\nu$ are differentiable at $\bar{v}_0^\nu$ with*

$\nabla g^\nu(\bar v_0^\nu) = \nabla g(\bar v_0^\nu)$. *Furthermore, unless* $\bar u_0^\nu$ *and* $\bar v_0^\nu$ *are already optimal solutions to* $(\mathcal{P})$ *and* $(\mathcal{Q})$, *one has then that either* $\bar u^\nu - \bar u_0^\nu$ *gives a direction of descent for* $f$ *at* $\bar u_0^\nu$ *relative to* $U$ *in* $(\mathcal{P})$, *or* $\bar v^\nu - \bar v_0^\nu$ *gives a direction of ascent for* $g$ *at* $\bar v_0^\nu$ *relative to* $V$ *in* $(\mathcal{Q})$, *or both. Thus the combination of the two line searches in Step 3 of the basic algorithm will in this case necessarily achieve a reduction in the duality gap: one will have* $\varepsilon^{\nu+1} < \varepsilon^\nu$.

**Proof.** If $F(\bar u_0^\nu)$ and $G(\bar u_0^\nu)$ are singletons, their sole elements must be the vectors $\bar u_1^\nu$ and $\bar v_1^\nu$ that according to the criterion in Step 2 belong to $U^\nu$ and $V^\nu$. The hypothesis of Proposition 5.1 is clearly fulfilled in this case. The argument in the proof of Proposition 5.1 can then be continued: Both of the subgradient sets $\partial\rho_{VQ}(z^\nu)$ and $\partial\rho_{V^\nu Q}(z^\nu)$, where $z^\nu = q - R\bar u_0^\nu$, must be singletons because both coincide with $F(\bar u_0^\nu)$. A convex function is differentiable at any point where its subgradient set is a singleton [13, Theorem 25.1]. Therefore $\rho_{VQ}$ and $\rho_{V^\nu Q}$ are both differentiable at $z^\nu$ with the same gradient, and this is equivalent to $f$ and $f^\nu$ both being differentiable at $\bar u_0^\nu$ with the same gradient. The same reasoning works for $g$ and $g^\nu$.

The main assertion of the theorem, about directions of descent and ascent, is a further consequence of Proposition 5.1. Let $\mu = \min(\mathcal{P}^\nu) = \max(\mathcal{Q}^\nu)$. Then

$$\mu = f^\nu(\bar u^\nu) \le f^\nu(\bar u_0^\nu) = f(\bar u_0^\nu), \tag{5.4}$$

where the middle inequality holds because of the condition $\bar u_0^\nu \in U^\nu$ in Step 2. Similarly

$$\mu = g^\nu(\bar v^\nu) \ge g^\nu(\bar v_0^\nu) = g(\bar v_0^\nu). \tag{5.5}$$

If $\bar u_0^\nu$ and $\bar v_0^\nu$ are not already optimal we have $f(\bar u_0^\nu) > g(\bar v_0^\nu)$, and at least one of the inequalities in (5.4) or (5.5) must be strict. This means through Proposition 5.1 that the line searches will improve either the primal objective or the dual objective or both. □

The next theorem gives our cornerstone result on convergence. To state it succinctly we need a notation for "relative matrix norms." For two symmetric matrices $N$ and $M$ of the same dimensions, with $M$ positive definite and $N$ positive semidefinite, we write

$$\|N/M\| = \max_{w\ne 0} \frac{\|w\|_N}{\|w\|_M} = \max_{w\ne 0} \frac{[w\cdot Nw]^{1/2}}{[w\cdot Mw]^{1/2}}. \tag{5.6}$$

**Theorem 5.3.** *In the fully quadratic case, the optimality gap* $\varepsilon^\nu$ *in the basic algorithm decreases to* 0 *at a geometric rate. Specifically, one at least has*

$$\varepsilon^{\nu+1} \le \theta_{PQR}\,\varepsilon^\nu \ \text{ for } \nu = 0, 1, 2, \ldots \tag{5.7}$$

*where*

$$\theta_{PQR} = 1 - \frac{1/4}{1+\gamma_{PQR}^2} < 1 \ \text{ with } \gamma_{PQR} = \|R^TQ^{-1}R/P\| = \|RP^{-1}R^T/Q\|. \tag{5.8}$$

*In this case the sequences* $\{\bar{u}_0^\nu\}$ *and* $\{\bar{v}_0^\nu\}$ *converge to the unique optimal solutions* $\bar{u}$ *and* $\bar{v}$ *to* $(\mathcal{P})$ *and* $(\mathcal{Q})$ *with*

$$\|\bar{u}_0^\nu - \bar{u}\|_P^2 + \|\bar{v}_0^\nu - \bar{v}\|_Q^2 \le 2(\theta_{PQR})^\nu \varepsilon^0 \text{ for } \nu = 1, 2, \ldots. \tag{5.9}$$

*In fact the sequence* $\{\bar{u}^\nu\}$ *and all the sequences* $\{\bar{u}_k^\nu\}$ *for* $k \le l$ *converge to* $\bar{u}$*, while the sequence* $\{\bar{v}^\nu\}$ *and all the sequences* $\{\bar{v}_k^\nu\}$ *for* $k \le l$ *converge to* $\bar{v}$*.*

**Remark.** Although the convergence factor $\theta_{PQR}$ provided by Theorem 5.3 may turn out to be close to 1, a noteworthy fact is that this convergence rate is guaranteed from the very start of the calculations. The result is therefore not of the usual sort, which would focus only on the tail of the generated sequence. The question of what sharper rate of convergence might ultimately be achieved in the tail is not addressed theoretically here. It would require a different type of analysis, not to mention further specification of the way the sets $U^\nu$ and $V^\nu$ are to be generated by the algorithm. □

**Proof.** As a preliminary, we verify that the two expressions given for $\gamma_{PQR}$ in (5.8) truly are equal. By definition we have

$$\begin{aligned}\|R^TQ^{-1}R/P\|^2 &= \max_{w\neq 0} \frac{w\cdot R^TQ^{-1}Rw}{w\cdot Pw} \\ &= \max_{z\neq 0} \frac{z\cdot P^{-\frac{1}{2}}R^TQ^{-1}RP^{-\frac{1}{2}}z}{z\cdot z} = \|(Q^{-\frac{1}{2}}RP^{-\frac{1}{2}})^T(Q^{-\frac{1}{2}}RP^{-\frac{1}{2}})\|^2,\end{aligned}$$

and by symmetry

$$\|RP^{-1}R^T/Q\|^2 = \|(P^{-\frac{1}{2}}R^TQ^{-\frac{1}{2}})^T(P^{-\frac{1}{2}}R^TQ^{-\frac{1}{2}})\|^2.$$

In terms of $S := Q^{-\frac{1}{2}}RP^{-\frac{1}{2}}$ the first expression is $\|S^TS\|^2$ while the second is $\|SS^T\|^2$. The fact that $\|S^TS\| = \|SS^T\|$ for any matrix $S$ tells us then that $\|R^TQ^{-1}R/P\| = \|RP^{-1}R^T/Q\|$.

Proceeding now with the main part of the proof, we note that because the Lagrangian $L$ is quadratic, it can be expanded as

$$\begin{aligned}L(u,v) =& L(\bar{u}_0^\nu, \bar{v}_1^\nu) + \nabla_u L(\bar{u}_0^\nu, \bar{v}_1^\nu)\cdot(u - \bar{u}_0^\nu) + \nabla_v L(\bar{u}_0^\nu, \bar{v}_1^\nu)\cdot(v - \bar{v}_1^\nu) \\ &+ \tfrac{1}{2}(u - \bar{u}_0^\nu)\cdot P(u - \bar{u}_0^\nu) - \tfrac{1}{2}(v - \bar{v}_1^\nu)\cdot Q(v - \bar{v}_1^\nu) - (v - \bar{v}_1^\nu)\cdot R(u - \bar{u}_0^\nu).\end{aligned} \tag{5.10}$$

Here, by the definition of $\bar{v}_1^\nu$ belonging to $F(\bar{u}_0^\nu)$, we have

$$L(\bar{u}_0^\nu, \bar{v}_1^\nu) = f(\bar{u}_0^\nu), \qquad \nabla_v L(\bar{u}_0^\nu, \bar{v}_1^\nu)\cdot(v - \bar{v}_1^\nu) \le 0 \text{ for all } v \in V. \tag{5.11}$$

Moreover the positive definiteness of $P$ yields from Theorem 5.2 (via Proposition 3.2) that

$$\nabla_u L(\bar{u}_0^\nu, \bar{v}_1^\nu)\cdot(u - \bar{u}_0^\nu) = f'(\bar{u}_0^\nu; u - \bar{u}_0^\nu) = \nabla f(\bar{u}_0^\nu)\cdot(u - \bar{u}_0^\nu). \tag{5.12}$$

Thus

$$L(u,v) \le f(\bar{u}_0^\nu) + \nabla f(\bar{u}_0^\nu)\cdot(u-\bar{u}_0^\nu) \\ + \tfrac{1}{2}(u-\bar{u}_0^\nu)\cdot P(u-\bar{u}_0^\nu) - \tfrac{1}{2}(v-\bar{v}_1^\nu)\cdot Q(v-\bar{v}_1^\nu) - (v-\bar{v}_1^\nu)\cdot R(u-\bar{u}_0^\nu) \text{ for all } v \in V,$$

and we obtain from the envelope formula (3.2) for $f$ in terms of $L$ that

$$\begin{aligned} f(u) - f(\bar{u}_0^\nu) - \nabla f(\bar{u}_0^\nu)\cdot(u-\bar{u}_0^\nu) - \tfrac{1}{2}(u-\bar{u}_0^\nu)\cdot P(u-\bar{u}_0^\nu) \\ &\le \max_{v\in V}\{-(v-\bar{v}_1^\nu)\cdot R(u-\bar{u}_0^\nu) - \tfrac{1}{2}(v-\bar{v}_1^\nu)\cdot Q(v-\bar{v}_1^\nu)\} \\ &\le \max_{w\in\mathbb{R}^m}\{-w\cdot R(u-\bar{u}_0^\nu) - \tfrac{1}{2}w\cdot Qw\} \\ &= \tfrac{1}{2}[R(u-\bar{u}_0^\nu)]\cdot Q^{-1}[R(u-\bar{u}_0^\nu)] = \tfrac{1}{2}(u-\bar{u}_0^\nu)\cdot R^T Q^{-1} R(u-\bar{u}_0^\nu). \end{aligned}$$

Applying this to $u = \bar{u}_0^\nu + \lambda(\bar{u}^\nu - \bar{u}_0^\nu)$ for $0 \le \lambda \le 1$, we reach the estimate that

$$f(\bar{u}_0^\nu + \lambda(\bar{u}^\nu - \bar{u}_0^\nu)) - f(\bar{u}_0^\nu) \le \lambda\nabla f(\bar{u}_0^\nu)\cdot(\bar{u}^\nu - \bar{u}_0^\nu) + \tfrac{1}{2}\lambda^2\|\bar{u}^\nu - \bar{u}_0^\nu\|^2_{P_0}, \tag{5.13}$$

where $P_0 = P + R^T Q^{-1} R$. The determination of $\bar{u}_0^{\nu+1}$ by line search over the segment joining $\bar{u}_0^\nu$ with $\bar{u}^\nu$ means that

$$f(\bar{u}_0^{\nu+1}) = \min_{0\le\lambda\le 1} f(\bar{u}_0^\nu + \lambda(\bar{u}^\nu - \bar{u}_0^\nu)),$$

and therefore

$$f(\bar{u}_0^{\nu+1}) - f(\bar{u}_0^\nu) \le \min_{0\le\lambda\le 1}\{\lambda\nabla f(\bar{u}_0^\nu)\cdot(\bar{u}^\nu - \bar{u}_0^\nu) + \tfrac{1}{2}\lambda^2\|\bar{u}^\nu - \bar{u}_0^\nu\|^2_{P_0}\}. \tag{5.14}$$

Invoking Theorem 5.2 again, we see that

$$\nabla f(\bar{u}_0^\nu)\cdot(\bar{u}^\nu - \bar{u}_0^\nu) = f^\nu(\bar{u}_0^\nu; \bar{u}^\nu - \bar{u}_0^\nu) \le f^\nu(\bar{u}^\nu) - f^\nu(\bar{u}_0^\nu), \tag{5.15}$$

while at the same time by (5.12)

$$\begin{aligned} \|\bar{u}^\nu - \bar{u}_0^\nu\|^2_{P_0} &= \|\bar{u}^\nu - \bar{u}_0^\nu\|^2_P + \|\bar{u}^\nu - \bar{u}_0^\nu\|_{R^T Q^{-1} R} \\ &\le \|\bar{u}^\nu - \bar{u}_0^\nu\|^2_P (1 + \|R^T Q^{-1} R/P\|^2). \end{aligned} \tag{5.16}$$

The fact that $(\bar{u}^\nu, \bar{v}^\nu)$ is a saddle point of $L$ relative to $U^\nu \times V^\nu$ gives also that $f^\nu(\bar{u}^\nu) = L(\bar{u}^\nu, \bar{v}^\nu)$ and

$$\begin{aligned} f^\nu(\bar{u}_0^\nu) - f^\nu(\bar{u}^\nu) &\ge L(\bar{u}_0^\nu, \bar{v}^\nu) - L(\bar{u}^\nu, \bar{v}^\nu) \\ &= \nabla_u L(\bar{u}^\nu, \bar{v}^\nu)\cdot(\bar{u}_0^\nu - \bar{u}^\nu) + \tfrac{1}{2}(\bar{u}^\nu - \bar{u}_0^\nu)\cdot P(\bar{u}^\nu - \bar{u}_0^\nu) \\ &= (f^\nu)'(\bar{u}^\nu; \bar{u}_0^\nu - \bar{u}^\nu) + \tfrac{1}{2}\|\bar{u}_0^\nu - \bar{u}^\nu\|^2_P, \end{aligned}$$

where the derivative term is nonnegative. In combination with (5.16) this yields

$$\tfrac{1}{2}\|\bar{u}^\nu - \bar{u}_0^\nu\|^2_{P_0} \le (1 + \gamma^2_{PQR})[f^\nu(\bar{u}_0^\nu) - f^\nu(\bar{u}^\nu)].$$

We use this inequality and the identity

$$f^\nu(\bar{u}^\nu) - f^\nu(\bar{u}_0^\nu) = L(\bar{u}^\nu, \bar{v}^\nu) - f(\bar{u}_0^\nu)$$

together with (5.15) to estimate the right side of (5.14) and obtain

$$\begin{aligned} f(\bar{u}_0^{\nu+1}) - f(\bar{u}_0^\nu) &\le [f(\bar{u}_0^\nu) - L(\bar{u}^\nu, \bar{v}^\nu)] \min_{0\le\lambda\le 1}\{-\lambda + (1+\gamma_{PQR}^2)\lambda^2\} \\ &= [f(\bar{u}_0^\nu) - L(\bar{u}^\nu, \bar{v}^\nu)][-1/4(1+\gamma_{PQR}^2)]. \end{aligned} \tag{5.17}$$

By a similar argument we get

$$g(\bar{v}_0^{\nu+1}) - g(\bar{v}_0^\nu) \ge [g(\bar{v}_0^\nu) - L(\bar{u}^\nu, \bar{v}^\nu)][-1/4(1+\gamma_{PQR}^2)]. \tag{5.18}$$

We can subtract (5.17) from (5.18) and have

$$g(\bar{v}_0^{\nu+1}) - g(\bar{v}_0^\nu) - f(\bar{u}_0^{\nu+1}) + f(\bar{u}_0^\nu) \ge [f(\bar{u}_0^\nu) - g(\bar{v}_0^\nu)]/4(1+\gamma_{PQR}^2),$$

which can be written as $\varepsilon^\nu - \varepsilon^{\nu+1} \ge \varepsilon^\nu/4(1+\gamma_{PQR}^2)$. This is identical to (5.7)–(5.8).

Condition (5.9) now follows from Proposition 3.3 and gives us the convergence of $\bar{u}_0^\nu$ to $\bar{u}$ and $\bar{v}_0^\nu$ to $\bar{v}$.

Because we are in the fully convex case of our problems, $F$ is a single-valued mapping, as we know from Proposition 3.2. As a matter of fact, $F$ is continuous, even Lipschitz continuous. To see this, observe that in terms of $v^* = F(u^*)$ we have the first-order optimality condition $\nabla_v L(u^*, v^*)\cdot(v - v^*) \le 0$ for all $v \in V$. If also $v^{**} = F(u^{**})$ we have $\nabla_v L(u^{**}, v^{**})\cdot(v - v^{**}) \le 0$ for all $v \in V$ as well, and the two inequalities yield

$$\begin{aligned} 0 &\ge \nabla_v L(u^*, v^*)\cdot(v^{**}, v^*) + \nabla_v L(u^{**}, v^{**})\cdot(v^* - v^{**}) \\ &= [\nabla_v L(u^*, v^*) - \nabla_v L(u^{**}, v^{**})]\cdot(v^{**} - v^*) \\ &= [q - Ru^* - Qv^* - q + Ru^{**} + Qv^{**}]\cdot(v^{**} - v^*) \\ &= (v^{**} - v^*)\cdot R(u^{**} - u^*) + (v^{**} - v^*)\cdot Q(v^{**} - v^*). \end{aligned}$$

Then

$$\begin{aligned} \|v^{**} - v^*\|_Q^2 &\le -(v^{**} - v^*)\cdot Q[Q^{-1}R(u^{**} - u^*)] \\ &\le \|v^{**} - v^*\|_Q \, \|Q^{-1}R(u^{**} - u^*)\|_Q, \end{aligned}$$

where

$$\begin{aligned} \|Q^{-1}R(u^{**} - u^*)\|_Q &= [(u^{**} - u^*)\cdot R^T Q^{-1} R(u^{**} - u^*)]^{1/2} \\ &\le \|R^T Q^{-1} R/P\| \, \|u^{**} - u^*\|_P. \end{aligned}$$

Thus

$$\|F(u^{**}) - F(u^*)\|_Q \le \|R^T Q^{-1} R/P\| \, \|u^{**} - u^*\|_P \text{ for all } u^* \text{ and } u^{**}, \tag{5.19}$$

and $F$ is Lipschitz continuous as claimed. Similarly, $G$ is a Lipschitz continuous mapping satisfying

$$\|G(v^{**}) - G(v^*)\|_P \le \|RP^{-1}R^T/Q\|\,\|v^{**} - v^*\|_Q \text{ for all } v^* \text{ and } v^{**}. \tag{5.20}$$

We have $\bar v = F(\bar u)$ and $\bar v_1^\nu = f(\bar u_0^\nu)$ with $\bar u_0^\nu \to \bar u$, so the continuity of $F$ gives us $\bar v_1^\nu \to \bar v$. Then because $\bar u = G(\bar v)$ and $\bar u_2^\nu = G(\bar v_1^\nu)$, the continuity of $G$ gives us $\bar u_2^\nu \to \bar u$. A parallel argument shows that $\bar u_1^\nu \to \bar u$ and $\bar v_2^\nu \to \bar v$. This pattern can be continued to show eventually that $\bar u_k^\nu \to \bar u$ and $\bar v_k^\nu \to \bar v$ for all $k \le l$.

It still must be demonstrated that $\bar u^\nu \to \bar u$ and $\bar v^\nu \to \bar v$. From the expansion (5.10) with (5.11) and (5.12) we have

$$\begin{aligned} f(\bar u_0^\nu) &= f^\nu(\bar u_0^\nu) \ge f^\nu(\bar u^\nu) \ge L(\bar u^\nu, \bar v_1^\nu) \\ &= L(\bar u_0^\nu, \bar v_1^\nu) + \nabla_u L(\bar u_0^\nu, \bar v_1^\nu)\cdot(\bar u^\nu - \bar u_0^\nu) + \tfrac12(\bar u^\nu - \bar u_0^\nu)\cdot P(\bar u^\nu - \bar u_0^\nu) \\ &= f(\bar u_0^\nu) + \nabla f(\bar u_0^\nu)\cdot(\bar u^\nu - \bar u_0^\nu) + \tfrac12\|\bar u^\nu - \bar u_0^\nu\|_P^2 \end{aligned}$$

and therefore

$$\tfrac12\|\bar u^\nu - \bar u_0^\nu\|_P^2 \le -\nabla f(\bar u_0^\nu)\cdot(\bar u^\nu - \bar u_0^\nu) \text{ for all } \nu. \tag{5.21}$$

In particular this tells us that

$$\|\bar u^\nu - \bar u_0^\nu\|_P^2 \le 2[P^{-1}\nabla f(\bar u_0^\nu)]\cdot P(\bar u^\nu - \bar u_0^\nu) \le 2\|P^{-1}\nabla f(\bar u_0^\nu)\|_P\,\|\bar u^\nu - \bar u_0^\nu\|_P$$

and gives us the bound

$$\|\bar u^\nu - \bar u_0^\nu\|_P \le 2\|P^{-1}\nabla f(\bar u_0^\nu)\|_P. \tag{5.22}$$

The gradient mapping $\nabla f$ is continuous by Proposition 3.2. Therefore the right side of (5.21) remains bounded as $\bar u_0^\nu \to \bar u$, and the sequence $\{\bar u^\nu\}$ must be bounded as well. Consider any cluster point $\tilde u$ of the sequence $\{\bar u^\nu\}$. Taking limits in (5.21) for the corresponding subsequences, we obtain

$$\tfrac12\|\tilde u - \bar u\|_P^2 \le -\nabla f(\bar u)\cdot(\tilde u - \bar u), \tag{5.23}$$

where, as we know from the optimality of $\bar u$ in $(\mathcal{P})$, we have $\nabla f(\bar u)\cdot(u - \bar u) \ge 0$ for all $u \in U$. We conclude from (5.23) that $\tilde u = \bar u$. Since $\tilde u$ was any cluster point of the bounded sequence $\{\bar u^\nu\}$, this means that $\bar u^\nu \to \bar u$. The proof that $\bar v^\nu \to \bar v$ is analogous. □

**Remark.** The quadratic expansion (5.10) of L used in the proof of Theorem 5.3 yields an interesting interpretation for the elements $\bar u_2^\nu$ and $\bar v_2^\nu$. Using the fact that $L(\bar u_0^\nu, \bar v_1^\nu) = f(\bar u_0^\nu)$ (by the definition of the condition $\bar v_1^\nu \in F(\bar u_0^\nu)$) and $\nabla_u L(\bar u_0^\nu, \bar v_1^\nu) = \nabla f(\bar u_0^\nu)$ (by Proposition 3.2), we see that

$$L(u, \bar v_1^\nu) = f(\bar u_0^\nu) + \nabla f(\bar u_0^\nu)\cdot(u - \bar u_0^\nu) + \tfrac12(u - \bar u_0^\nu)\cdot P(u - \bar u_0^\nu).$$

If we were to replace the standard euclidean structure on $\mathbb{R}^n$ by that corresponding to $P$ through the inner product $\langle u, w\rangle_P = u{\cdot}Pw$, we could write this as

$$L(u, \bar{v}_1^\nu) = f(\bar{u}_0^\nu) + \langle \nabla_P f(\bar{u}_0^\nu), u - \bar{u}_0^\nu\rangle_P + \tfrac{1}{2}\|u - \bar{u}_0^\nu\|_P^2,$$

where $\nabla_P f(u)$ symbolizes the gradient of $f$ relative to this other structure, namely the vector $P^{-1}\nabla f(u)$. In minimizing this over $u \in U$, we get by definition the elements of $G(\bar{v}_1^\nu)$, which in this case reduce to one, namely $\bar{u}_2^\nu$. It follows that

$$\bar{u}_2^\nu - \bar{u}_0^\nu = [\,P\text{-projection of } \nabla_P f(\bar{u}_0^\nu) \text{ on } U - \bar{u}_0^\nu\,].$$

Thus $\bar{u}_2^\nu - \bar{u}_0^\nu$ gives a sort of modified steepest descent direction for $f$ at $\bar{u}_0^\nu$. □

The virtue of Theorem 5.3 is that it guarantees convergence under quite minimal conditions. It can be used as a stepping stone to other convergence results such as the one to be presented next. For this result, which concerns the identification of the "active constraints" in the course of the basic algorithm, we need another concept, that of the "critical faces" of $U$ and $V$ relative to problems $(\mathcal{P})$ and $(\mathcal{Q})$.

Recall that in minimizing $f(u)$ over $u \in U$ in $(\mathcal{P})$ one has as a first-order necessary condition for the optimality of $\bar{u}$ that $0 \in \partial f(\bar{u}) + N_U(\bar{u})$, where $N_U(\bar{u})$ is the normal cone to $U$ at $\bar{u}$ in the sense of convex analysis, cf. [13, Theorem 27.4]. This means the existence of a vector $w \in \mathbb{R}^n$ such that $w \in \partial f(\bar{u})$ and

$$\bar{u} \in \operatorname*{argmin}_{u\in U} w{\cdot}u. \tag{5.24}$$

The set of all such vectors $w$ is actually independent of the particular choice of $\bar{u}$ if $(\mathcal{P})$ happens to have more than one optimal solution. (It is the set of optimal solutions to the Fenchel dual of $(\mathcal{P})$ when $(\mathcal{P})$ is interpreted in the format of minimizing $f - \varphi$ over $\mathbb{R}^n$, where $\varphi = -\delta_U$; see [13,S31].) For each of these vectors $w$, the set on right of (5.24) is a certain *face* of the polyhedron $U$, and hence so is the intersection of all such faces. The latter is what we define to be the *critical face* $U_0$ *of* $U$ *in* $(\mathcal{P})$. From (5.24), one obviously has $\bar{u} \in U_0$ when $\bar{u}$ is optimal.

The critical face $V_0$ of $V$ in problem $(\mathcal{Q})$ is defined analogously. One has $\bar{v} \in V_0$ when $\bar{v}$ is dual optimal. The feasible solutions to $(\mathcal{P})$ and $(\mathcal{Q})$ that belong to $U_0$ and $V_0$ rather than just to $U$ and $V$ can be interpreted as the ones that "satisfy as equations the linear constraints in the expression of $U$ and $V$ that must be active in an essential way at optimality." This will be explained more fully at the end of this section.

Our actual use here of the critical face concept will be limited to the relatively simple case where the objective functions $f$ and $g$ are differentiable and the optimal solutions $\bar{u}$ and $\bar{v}$ are unique. Then the two critical faces are expressed by

$$U_0 = \operatorname*{argmin}_{u\in U} \nabla f(\bar{u}){\cdot}u = \operatorname*{argmin}_{u\in U} \nabla_u L(\bar{u}, \bar{v}){\cdot}u \tag{5.25}$$

$$V_0 = \operatorname*{argmax}_{v\in V} \nabla g(\bar{v}){\cdot}v = \operatorname*{argmax}_{v\in V} \nabla_v L(\bar{u}, \bar{v}){\cdot}v. \tag{5.26}$$

**Theorem 5.4.** *Consider the fully quadratic case. Let $U_0$ be the critical face of $U$ in $(\mathcal{P})$ and let $V_0$ be the critical face of $V$ in $(\mathcal{Q})$, as just defined. Then, in any realization of the basic algorithm, the sequences $\{\bar{u}_k^\nu\}$ for $0 < k \leq l$ will eventually lie entirely in $U_0$, while the sequences $\{\bar{v}_k^\nu\}$ for $0 < k \leq l$ will eventually lie entirely in $V_0$.*

**Proof.** Our proof will hinge on a general property of polyhedral convex sets like $U$: in terms of

$$M(w) := \operatorname*{argmax}_{u \in U} w \cdot u = \{u \in U \,\big|\, w \in N_U(u)\}$$

one has that $M(w) \subset M(\bar{w})$ for all $w$ in some neighborhood of $\bar{w}$. This is true because $M(w)$ must be one of the finitely many faces of $U$, and the graph of the set-valued mapping $M : u \mapsto M(u)$ is closed.

The application we wish to make of this general fact is to $\bar{w} = -\nabla f(\bar{u})$. (Recall that $f$ is differentiable by Proposition 3.2, because we are in the fully quadratic case.) Inasmuch as $\bar{u}$ minimizes $f$ over $U$, we do have $-\nabla f(\bar{u}) \in N_U(\bar{u})$, i.e., $\bar{u} \in M(-\nabla f(\bar{u}))$. The set $M(-\nabla f(\bar{u}))$ is in this case exactly the critical face $U_0$ defined in (5.25). It will be helpful to us that the mapping $\nabla f$ is continuous, as noted in the proof of Theorem 5.3.

First we shall demonstrate that $\bar{u}_1^\nu \in U_0$ for all $\nu$ sufficiently large. By definition, $\bar{u}_1^\nu$ minimizes $L(u, \bar{v}_0^\nu)$ over $u \in U$, so that $-\nabla_u L(\bar{u}_1^\nu, \bar{v}_0^\nu) \in N_U(\bar{u}_1^\nu)$, i.e.

$$\bar{u}_1^\nu \in M(-\nabla_u L(\bar{u}_1^\nu, \bar{v}_0^\nu)). \tag{5.27}$$

According to Theorem 5.3 we have $\bar{u}_1^\nu \to \bar{u}$ and $\bar{v}_0^\nu \to \bar{v}$, hence also $\nabla_u L(\bar{u}_1^\nu, \bar{v}_0^\nu) \to \nabla_u L(\bar{u}, \bar{v})$. The latter vector is just $\nabla f(\bar{u})$, because $\bar{v} \in F(\bar{u})$. Thus the vector $w^\nu = -\nabla_u L(\bar{u}_1^\nu, \bar{v}_0^\nu)$ in (5.27) converges to $\bar{w} = -\nabla f(\bar{u})$ while $\bar{u}_1^\nu \to \bar{u}$, and therefore, by our basic property, we have for all $\nu$ sufficiently large that $\bar{u}_1^\nu \in M(-\nabla f(\bar{u})) = U_0$.

The argument verifying that $\bar{u}_2^\nu \in U_0$ for all $\nu$ sufficiently large is virtually the same, being based instead on the fact that $\bar{u}_2^\nu$ minimizes $L(u, \bar{v}_1^\nu)$ over $u \in U$. Likewise we get $\bar{v}_1^\nu \in V_0$ and $\bar{v}_2^\nu \in V_0$ for large $\nu$. The pattern continues up to $\{\bar{u}_l^\nu\}$ and $\{\bar{v}_l^\nu\}$. □

## 6. VARIANTS AND EXTENSIONS

The main convergence results in Theorems 5.3 and 5.4 for finite-envelope methods require the fully quadratic case, but of course not all problems of interest come with $P$ and $Q$ positive definite. There is a device that can be used in such problems to create positive definiteness. It is the general *proximal point algorithm*, developed by the author in [10] (also [11], [12]).

With this device a shell of "outer" iterations is added to the "inner" iterations of the basic finite-envelope algorithm. In outer iteration $\mu$ we have points $u_*^\mu \in U$ and $v_*^\mu \in V$ and form the modified Lagrangian

$$L_*^\mu(u, v) = L(u, v) + \tfrac{1}{2}(u - u_*^\mu) \cdot P_*^\mu (u - u_*^\mu) - \tfrac{1}{2}(v - v_*^\mu) Q_*^\mu (v - v_*^\mu) \tag{6.1}$$

for some choice of positive definite matrices $P_*^\mu$ and $Q_*^\mu$. This modified Lagrangian does, of course, exhibit the desired positive definiteness. The finite-envelope algorithm as already laid out in S5 can be applied to determine a saddle point $(u_*^{u+1}, v_*^{\mu+1})$ of $L_*^\mu$ on $U \times V$, and the sequences it generates will enjoy all the properties in Theorems 5.3 and 5.4. With the new pair $(u_*^{\mu+1}, v_*^{\mu+1})$ we are able to form $L_*^{\mu+1}$ and reapply the basic algorithm to that.

It is not really necessary that $(u_*^{\mu+1}, v_*^{\mu+1})$ be a true saddle point in each outer iteration to be sure that this sequence will converge to an optimal solution pair $(\bar{u}, \bar{v})$ for $(\mathcal{P})$ and $(\mathcal{Q})$. The duality gap in the sequence of inner iterations just has to be brought down sufficiently far. The details can be gleaned from the cited papers on the proximal point algorithm, but we shall not work them out here. Convergence can be guaranteed under quite mild conditions, and rates of convergence derived as well.

There are many issues here to be investigated further, particularly the effects of the parameters used in the outer algorithm. It deserves notice that in the specialized context of two-stage stochastic programming in [4], it was possible to show that a linear rate of convergence could be guaranteed for the combined inner-outer algorithm whenever the optimal solutions $\bar{u}$ and $\bar{v}$ to the given problems (not necessarily possessed of positive definiteness) are unique.

Another mode of extension concerns the role of the envelope subproblems. Although we have not mentioned it until now in order to have our notation be simple, there is no reason why the points $\bar{u}^\nu$ and $\bar{v}^\nu$ used to get the directions of primal and dual line search must be calculated from the same, joint saddle point subproblem. In principle we could calculate a saddle point $(\bar{u}^\nu, \tilde{v}^\nu)$ of $L$ relative to $U^\nu \times \tilde{V}^\nu$ to get $\bar{u}^\nu$ and then a saddle point $(\tilde{u}^\nu, \bar{v}^\nu)$ of $L$ relative to $\tilde{U}^\nu \times V^\nu$ to get $\bar{v}^\nu$. We would then need to generate and maintain finite collections of points representing $\tilde{U}^\nu$ and $\tilde{V}^\nu$ in addition to $U^\nu$ and $V^\nu$, but the decoupling might possibly lead to better directions being obtained. The basic convergence theory generalizes easily to this broader picture.

The line searches in the algorithm of S5 might be replaced by searches over triangles and other small polytopes, if in computational developments yet to come it turns out that problems in our category where only the primal or only the dual dimension happens to be high can be solved relatively quickly. The minimization of $f$ over a line segment $[\bar{u}_0^\nu, \bar{u}^\nu]$ can in fact be interpreted as a very special case of our model in which $U$ is replaced by this segment, which is in particular a one-dimensional polytope, the convex hull of two points. It corresponds to finding a saddle point of $L(u, v)$ relative to $[\bar{u}_0^\nu, \bar{u}^\nu] \times V$. The line segment could be replaced quite easily by the convex hull of three or more points to get a subproblem that would be only a little more difficult perhaps, in principle. Such a subproblem could be reduced by the reformulation device in Proposition 4.2, applied only in the primal argument, to one of finding a saddle point of $L(u, v)$ over $S(r) \times V$ for some simplex $S(r)$, $r > 1$.

Finally, we mention that some of the ideas involved in finite-generation methods could well be extended beyond linear-quadratic problems. A general theory of finding saddle

points by primal-dual decomposition has been worked out recently by Oettli [19].

**Concluding remark.** It is worth emphasizing once more that finite-envelope methods are only one class of methods that in principle could take advantage of the special structure we have outlined. The main point of this paper is that such previously unrecognized structure is present, and numerical methods should try to utilize it.

## REFERENCES


1. R.W. Cottle, F. Giannessi, and J.-L. Lions (eds.), *Variational Inequalities and Complementarity Problems*, Wiley, 1980.
2. Y.Y. Lin and J.-S. Pang, "Iterative methods for large convex quadratic programs: a survey," SIAM Journal on Control and Optimization **25** (1987), 383–411.
3. R.T. Rockafellar, "Linear-quadratic programming and optimal control," SIAM Journal on Control and Optimization **25** (1987), 781–814.
4. R.T. Rockafellar and R.J-B Wets,"A Lagrangian finite-generation technique for solving linear-quadratic problems in stochastic programming," Mathematical Programming Studies **28** (1986), 63–93.
5. R.T. Rockafellar and R.J-B Wets, "Linear-quadratic problems with stochastic penalties: the finite generation algorithm," in: *Stochastic Optimization*, V.I. Arkin, A. Shiraev, and R.J-B Wets (eds.), Springer-Verlag Lecture Notes in Control and Information Sciences No. 81, 1987, 545–560.
6. R.T. Rockafellar and R.J-B Wets, "Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time," SIAM Journal on Control and Optimization, to appear.
7. K.C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics No. 1133, Springer-Verlag, Berlin, 1985.
8. 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.
9. J.M. Wagner, *Stochastic Programming with Recourse Applied to Groundwater Quality Management*, doctoral dissertation, M.I.T, 1988.
10. R.T. Rockafellar, "Monotone operators and the proximal point algorithm," SIAM Journal on Control and Optimization **14** (1976), 877–898.
11. R.T. Rockafellar, "Augmented Lagrangians and applications of the proximal point algorithm in convex programming," Mathematics of Operations Research **1** (1976), 97–116.
12. R.T. Rockafellar, "Lagrange multipliers and variational inequalities," in: [1], 303-322.
13. R.T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, NJ, 1970.
14. J.-S. Pang, "Methods for quadratic programming: a survey," Computers and Chemical Engineering **7** (1983), 583–594.
15. Y. Ye and E. Tse, "A polynomial-time algorithm for convex quadratic programming," preprint.

16. R.C. Monteiro and I. Adler, "An $0(n^3L)$ interior point algorithm for convex quadratic programming," preprint.
17. D. Goldfarb and S. Liu, "An $0(n^3L)$ primal interior point algorithm for convex quadratic programming," preprint.
18. C. Zhu and R.T. Rockafellar, "Primal-dual gradient projection methods for extended linear-quadratic programming," preprint.
19. W. Oettli, "Decomposition schemes for finding saddle points of quasi-convex-concave functions," preprint.