Ref. copy

Mathematical Programming Study 28 (1986) 63-93
North-Holland

A LAGRANGIAN FINITE GENERATION TECHNIQUE FOR SOLVING LINEAR-QUADRATIC PROBLEMS IN STOCHASTIC PROGRAMMING

R.T. ROCKAFELLAR
Department of Mathematics, University of Washington, Seattle, WA 98115, USA

R.J.-B. WETS
Department of Mathematics, University of California, Davis, CA 95616, USA

Received 1 March 1984
Revised manuscript received 8 December 1984

A new method is proposed for solving two-stage problems in linear and quadratic stochastic programming. Such problems are dualized, and the dual, althought itself of high dimension, is approximated by a sequence of quadratic programming subproblems whose dimensionality can be kept low. These subproblems correspond to maximizing the dual objective over the convex hull of finitely many dual feasible solutions. An optimizing sequence is produced for the primal problem that converges at a linear rate in the strongly quadratic case. An outer algorithm of augmented Lagrangian type can be used to introduce strongly quadratic terms, if desired.

Key words: Stochastic Programming, Large-Scale Quadratic Programming, Lagrangian Methods.

1. Introduction

In the recourse model in stochastic programming, a vector x must be chosen optimally with respect to present costs and constraints as well as certain expected costs and induced constraints that are associated with corrective actions available in the future. Such actions may be taken in response to the observation of the values of various random variables about which there is only statistical information at the time x is selected. The actions involve costs and constraints that depend on these observed values and on x. The theory of this kind of stochastic programming and the numerical methods that have been proposed for it has been surveyed recently by Wets [12].

We aim here at developing a new solution procedure for the case where the first and second stage problems in the recourse model fit the mold of linear or quadratic (convex) programming. We assume for simplicity that the random variables are discretely distributed with only finitely many values. This restriction is not fully necessary in theory, but it reflects the realities of computation and a natural division among the questions that arise. Every continuous distribution must in practice be replaced by a finite discrete one, whether empirically, or through sampling, mathematical approximation, or in connection with the numerical calculation of integrals expressing expectations. The effects of such discretization raise important questions

This work was supported in part by grants from the National Science Foundation.

of convergence and statistical confidence in the solutions that are obtained, but such matters are best left to separate study.

We assume therefore that the probability space is a finite set Ω: the probability associated with an element $\omega \in \Omega$ is p_ω, and the expectation of a quantity u_ω that depends on ω is

$$Eu_\omega := \sum_{\omega \in \Omega} p_\omega u_\omega.$$

The fundamental problem we want to address is

$$\text{minimize} \quad c \cdot x + \tfrac{1}{2} x \cdot Cx + E\psi_\omega(x) \quad \text{over all} \quad x \in X \subset \mathbb{R}^n, \tag{P}$$

where X is a nonempty convex polyhedron, c is a vector in $\mathbb{R}^n$, C is a symmetric matrix in $\mathbb{R}^{n+n}$ that is positive semidefinite, and $\psi_\omega(x)$ is the minimum cost in a certain recourse subproblem that depends on ω and x. (Here $x \cdot y$ denotes the inner product of x and y.) We view this recourse subproblem as one of linear or quadratic programming, but instead of handling it directly we work with its dual. More will be said about this later (see Proposition 1 in Section 2 and the comments after its proof), but what counts in the end is the following: we suppose a representation

$$\psi_\omega(x) = \max_{z_\omega \in Z_\omega} \{ z_\omega \cdot [h_\omega - T_\omega x] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega \} \tag{1.1}$$

is available, where Z_ω is a nonempty convex polyhedron in $\mathbb{R}^m$, T_ω is a matrix in $\mathbb{R}^{m \times n}$, h_ω is a vector in $\mathbb{R}^m$, and H_ω is a symmetric matrix in $\mathbb{R}^{m \times m}$ that is positive semidefinite. Such a formulation also covers important cases where, as will be explained presently, "recourse" is not the key idea and instead $\psi_\omega(x)$ arises when penalty expressions of a certain general type are introduced to restrain the difference vector $h_\omega - T_\omega x$. Note from the subscript ω that all the elements in the representation (1.1) are in principle allowed to be random, although a particular application might not involve quite so much randomness.

Two basic conditions are imposed on the given data. We assume X and C are such that for every $v \in \mathbb{R}^n$ the set

$$\xi(v) := \operatorname*{argmin}_{x \in X} \{ v \cdot x + \tfrac{1}{2} x \cdot Cx \} \tag{1.2}$$

is nonempty and bounded. We also assume Z_ω, h_ω, T_ω, and H_ω, are such that for every $x \in X$ the set

$$\zeta_\omega(x) := \operatorname*{argmax}_{z\omega \in Z_\omega} \{ z_\omega \cdot [h_\omega - T_\omega x] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega \} \tag{1.3}$$

is nonempty and bounded. Certainly the first condition holds if X is bounded or C is positive definite, and the second holds if Z_ω is bounded or H_ω is positive definite.

The first condition is quite innocuous, since in practice X can always be taken to be bounded. It implies that the function

$$\phi(v) = \inf_{x \in X} \{ v \cdot x + \tfrac{1}{2} x \cdot Cx \}, \tag{1.4}$$

which will have a role in duality, is finite everywhere.

The second condition is more subtle, since it involves dual elements that might not be given directly but derived instead from a primal statement of the recourse subproblem that depends on x and ω. It ensures in particular that for every $x \in X$ and $\omega \in \Omega$, the optimal value $\psi_\omega(x)$ in this subproblem is finite, and an optimal recourse exists. This means that our stochastic programming problem (P) is one of *relatively complete recourse* [10]: there are no induced constraints on x that arise from the need to keep open the possibility of recourse at a later time.

Of course, if our problem were not one of relatively complete recourse, we could make it so by identifying the induced constraints and shrinking the set X until they were all satisfied. The smaller X would still be a convex polyhedron, although its description might be tedious in situations where special approaches such as in [10, Section 1] can't be followed. In this sense our second condition forces no real restriction on the problem either, except in requiring that the induced constraints, if any, be identified thoroughly in advance.

In some of the situations that motivate our model the recourse subproblem is actually trivial and its solution can be given in closed form. Such situations occur when constraints are represented by penalties: the term $E\psi_\omega(x)$ in (P) can then be interpreted as an *expected penalty*. Indeed, using the notation

$$\theta_\omega(u) = \max_{z_\omega \in Z_\omega} \{z_\omega \cdot u - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\} \tag{1.5}$$

we can write

$$\psi_\omega(x) = \theta_\omega(h_\omega - T_\omega x). \tag{1.6}$$

If $0 \in Z_\omega$, then

$$\theta_\omega(u) \geq 0 \quad \text{for all } u, \qquad \theta_\omega(0) = 0, \tag{1.7}$$

so we can view $\theta_\omega(h_\omega - T_\omega x)$ as a penalty attached to certain degrees or directions of deviation of $T_\omega x$ from the vector h_ω. Many useful penalty functions of linear-quadratic type can be expressed as in (1.5). In particular the case where $\theta_\omega(h_\omega - T_\omega x)$ is a sum of separate terms, one for each scalar component of the deviation vector $h_\omega - T_\omega x$, can be identified with the case where each Z_ω is a product of intervals and H_ω is diagonal. This case underlies a special model we have treated in [9].

The solution procedure that we shall present depends on a Lagrangian representation of problem (P) which leads to the dual problem

$$\begin{aligned} &\text{maximize} \quad \phi(c - ET_\omega^* z_\omega) + E\{z_\omega \cdot h_\omega - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\} \\ &\text{subject to} \quad z_\omega \in Z_\omega \ \text{for all } \omega \in \Omega. \end{aligned} \tag{D}$$

Here ϕ is the function in (1.4), for which another representation will later be given (Proposition 2 in Section 2). The asterisk * signals the transpose of a matrix. The maximization in (D) takes place over the convex polyhedron

$$Z = \Pi_{\omega \in \Omega} Z_\omega \subset (\mathbb{R}^m)^\Omega; \tag{1.8}$$

we think of z_ω as the component in Z_ω of a point $z \in Z$. The vector space $(\mathbb{R}^m)^\Omega$ here, which is a product of copies of $\mathbb{R}^m$, one for each $\omega \in \Omega$, is likely to be of very high dimension, since the number of points in Ω may be very large. Despite this formidable dimensionality it is by way of (D), at least in concept, that we propose to solve (P). Properties of expectation, decomposition and quadratic structure, will make this possible. The relationship between (P) and (D) is explored in Section 2 along with other questions of quadratic programming duality that are crucial to in our formulation and our algorithm.

We approach problem (D) by a finite generation technique in which the feasible region Z is approximated from within by polytopes of comparatively low dimension, a *polytope* being a subset generated as the convex hull of finitely many points. This technique is presented in Section 3. It resembles the classical finite-element or Galerkin approach to the unconstrained maximization of a functional defined over an infinite-dimensional space, where one maximizes over finite-dimensional subspaces that grow in size as the approximation is refined. An important difference, however, is that in our case the new element or elements that are introduced at each stage in modifying the polytope over which we maximize are not obtained from some predetermined scheme, as classically, but identified in an 'adaptive' manner. Furthermore, the total number of elements used in generating the polytope does not have to keep increasing; the sequence of polytopes does not have to be nested. We prove in Section 4 that when the matrix C is positive definite these elements can readily be consolidated without threat to ultimate convergence, although the rate of progress may be better if a substantial set of generating elements is maintained. In this way the dimension of the subproblem to be solved in every iteration can be kept as low as seems desirable.

The subproblem of maximizing over a polytope can be represented as a standard type of quadratic programming problem and solved exactly by available codes. It yields as a byproduct an approximate solution vector for (P) along with bounds that provide a test of near optimality. The sequence of such approximate solutions converges to an optimal solution to (P). If not only C but also the matrices H_ω are positive definite, the rate of convergence is linear, in fact with guaranteed progress of a certain sort in every iteration, not just for the tail of the sequence.

In producing a new element to be used in the subrepresentation of Z in terms of a convex polytope, we have a particular x on hand and must carry out the maximization in (1.1) for every $\omega \in \Omega$. In other words, we must solve a large number of closely related linear or quadratic programming problems in $\mathbb{R}^m$. This could be a difficult task in general, but techniques such as have already been developed in connection with other approaches to stochastic programming problems of a more special nature (see Wets [12]) do offer hope. Furthermore, there are cases of definite interest where the maximization in (1.1) is trivial, for instance where Z_ω is a product of intervals and H_ω is diagonal. Such a case has been described in [11].

Not all of the problems we wish to solve have C and H_ω positive definite, but this does not prevent the application of our method and the achievement of a linear rate of convergence. Augmented Lagrangian techniques [7] can be effective in approximating any problem (P) by a sequence of similar problems that do exhibit positive definiteness. We explain this in Section 5 after having established in Section 4 the results that show the advantages of the strongly quadratic case.

Our algorithm has been implemented successfully by Alan King on a VAX 11/780 at IIASA and at the University of Washington for solving quadratic stochastic programs with simple recourse. We have solved some product-mix test problems, and used it in the analysis of investment strategies to control the eutrophication process of a shallow lake. This last class of problems involved 56 decision variables, most of them with upper and lower bounds; the set X was determined by 35 linear constraints. The matrix $T \in \mathbb{R}^{4\times 56}$ and the vector $h \in R^4$ were random, whereas the (nonstochastic) quadratic term involving H was introduced as a result of the augmentation procedure suggested in Section 5. A report on this implementation and the numerical results that have been obtained has been written by A. King [4].

2. Lagrangian representation and duality

As the *Lagrangian* associated with problem (P) under the representation (1.1) of the recourse costs, we shall mean the function

$$L(x, z) = c \cdot x + \tfrac{1}{2} x \cdot Cx + E\{z_\omega \cdot [h_\omega - T_\omega x] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\}$$
$$\text{for } x \in X,\ z \in Z, \tag{2.1}$$

where Z is the convex polyhedron in (1.8). Clearly $L(x, z)$ is convex in x and concave in z, since C and H_ω are positive semidefinite. General duality theory [6] associates with L, X, and Z, the primal problem

$$\text{minimize} \quad F \text{ over } X, \quad \text{where } F(x) := \max_{z \in Z} L(x, z), \tag{2.2}$$

and the dual problem

$$\text{maximize} \quad G \text{ over } Z, \quad \text{where } G(z) := \min_{x \in X} L(x, z). \tag{2.3}$$

The functions F and G are convex and concave, respectively. Our assumptions in Section 1 allow us to write 'max' and 'min' in their definitions rather than 'sup' and 'inf'.

These problems turn out to be the ones already introduced. In terms of the notation in (1.2) and (1.3), we have

$$\operatorname*{argmax}_{z \in Z} L(x, z) = \{z \mid z_\omega \in \zeta_\omega(x), \text{ for all } w \in \Omega\}, \tag{2.4}$$

$$\operatorname*{argmin}_{x \in X} L(x, z) = \xi(c - ET^*_\omega z_v). \tag{2.5}$$

Moreover for $x \in X$ and $z \in Z$ we have

$$F(x) = c \cdot x + \tfrac{1}{2}x \cdot Cx + E\psi_\omega(x), \tag{2.6}$$

$$G(z) = \varphi(c - ET^*_\omega z_\omega) + E\{z_\omega \cdot h_\omega - \tfrac{1}{2}z_\omega \cdot H_\omega z_\omega\}. \tag{2.7}$$

Thus the primal and dual problems (2.2) and (2.3) can be identified with (P) and (D), respectively.

In order to continue with our analysis of these problems, we need to step back briefly for a look at some basic facts about duality in quadratic programming, not only as they might apply to (P) and (D), but also to various subproblems in our schemes. A quadratic programming problem is usually defined as a problem in which a quadratic convex function is minimized (or a quadratic concave function maximized) subject to a system of linear constraints, or in other words, over a convex polyhedron. As is well known, such a problem has an optimal solution whenever its optimal value is finite (see Frank and Wolfe [3, Appendix (i)]); the Kuhn–Tucker conditions are both necessary and sufficient for optimality. For the purpose at hand, it is essential to adopt a more general point of view in which a problem is considered to fall in the category of quadratic programming as long as it can be *represented* in this traditional form, possibly through the introduction of auxiliary variables.

Consider an arbitrary Lagrangian of the form

$$l(u, v) = p \cdot u + q \cdot v + \tfrac{1}{2}u \cdot Pu - \tfrac{1}{2}v \cdot Qv - v \cdot Ru \quad \text{for } u \in U,\ v \in V, \tag{2.8}$$

where U and V are nonempty convex polyhedra, and P and Q are symmetric, positive semidefinite matrices. Let

$$f(u) = \sup_{v \in V}\{v \cdot [q - Ru] - \tfrac{1}{2}v \cdot Qv\}, \tag{2.9}$$

$$U_0 = \{u \mid f(u) \text{ finite}\} = \{u \mid \sup \text{ in (2.9) attained}\}, \tag{2.10}$$

$$g(v) = \inf_{u \in U}\{u \cdot [p - R^*v] - \tfrac{1}{2}u \cdot Pu\}, \tag{2.11}$$

$$V_0 = \{v \mid g(v) \text{ finite}\} = \{v \mid \inf \text{ in (2.11) attained}\}. \tag{2.12}$$

The primal and dual problems associated with l, U, and V by general duality theory can then we written as:

$$\text{minimize } p \cdot u + \tfrac{1}{2}u \cdot Pu + f(u) \text{ over } u \in U \cap U_0, \tag{P_0}$$

$$\text{maximize } q \cdot v - \tfrac{1}{2}v \cdot Qv + g(v) \text{ over } v \in V \cap V_0. \tag{D_0}$$

The following duality theorem for (P_0) and (D_0) extends the standard results in quadratic programming that were achieved by Dorn [2] and Cottle [1]. Those authors concentrated in effect on the case where U and V are orthants. The proof that we furnish is directed not only at an extension of theory, however. It explains how the optimal solutions to problems in the general framework of (P_0) and (D_0) can be identified in terms of the input and output of standard algorithms in quadratic programming after a reformulation. This observation is essential in dealing with the variouus subproblems that will play a role in Section 3.

Theorem 1. *Problems* (P_0) *and* (D_0) *are representable as quadratic programming in the traditional sense. If* (P_0) *and* (D_0) *both have feasible solutions, or if either* (P_0) *or* (D_0) *has finite optimal value, then both have optimal solutions, and*

$$\min(\mathrm{P}_0) = \max(\mathrm{D}_0).$$

This occurs if and only if the Lagrangian l *has a saddle point* $(\bar{u}, \bar{v})$ *relative to* $U \times V$, *in which case the saddle value* $l(\bar{u}, \bar{v})$ *coincides with the common optimal value in* (P_0) *and* (D_0), *and the saddle points are the pairs* $(\bar{u}, \bar{v})$ such that $\bar{u}$ *is an optimal solution to* (P_0) *and* $\bar{v}$ *is an optimal solution to* (D_0).

Proof. General duality theory [8] assures us that $\inf(\mathrm{P}_0) \geq \sup(\mathrm{D}_0)$ and in particular that both (P_0) and (D_0) have finite optimal value if both have feasible solutions. It also informs us that $(\bar{u}, \bar{v})$ is a saddle point of l on $U \times V$ if and only if $\bar{u}$ is an optimal solution to (P_0), $\bar{v}$ is an optimal solution to (D_0), and $\min(\mathrm{P}_0) = \max(\mathrm{D}_0)$, this common optimal value then being equal to $l(\bar{u}, \bar{v})$. We know further that a quadratic programming problem in the traditional sense has an optimal solution if it has finite optimal value [3, Appendix (i)]. The Kuhn–Tucker conditions are both necessary and sufficient for optimality in such a problem, because the constraint system is linear. The proof of the theorem can be reduced therefore to demonstrating that (P_0) and (D_0) are representable as quadratic programming in the traditional sense and in such a manner that the Kuhn–Tucker conditions for either problem correspond to the saddle point condition for l on $U \times V$.

The sets U and V are associated with systems of linear constraints that can be expressed in various ways, but to be specific we can suppose that

$$U = \{u \in \mathbb{R}^n \mid Au \geq a\} \neq \emptyset \quad \text{and} \quad V = \{v \in \mathbb{R}^m \mid B^* v \leq b\} \neq \emptyset, \tag{2.13}$$

where A is $m' \times n$ and B is $m \times n'$. Let $u' \in \mathbb{R}^{n'}$ and $v' \in \mathbb{R}^{m'}$ be Lagrange multiplier vectors paired with the conditions $B^* v \leq b$ and $Au \geq a$, respectively.

Formula (2.9) gives $f(u)$ as the optimal value in a classical quadratic programming problem in v. The optimal solutions to this problem are vectors that satisfy the usual Kuhn–Tucker conditions, or in other words, correspond to saddle points of the Lagrangian

$$v \cdot [q - Ru] - \tfrac{1}{2} v \cdot Qv + u' \cdot [b - B^* v] = b \cdot u' + v \cdot [q - Ru - Bu'] - \tfrac{1}{2} v \cdot Qv \tag{2.14}$$

relative to $u' \in \mathbb{R}^{n'}_+$ and $v \in \mathbb{R}^m$. In partcular, then, we have

$$f(u) = \inf_{u' \in \mathbb{R}^{n'}_+} \sup_{v \in \mathbb{R}^m} \{b \cdot u' + v \cdot [q - Ru - Bu'] - \tfrac{1}{2} v \cdot Qv\}. \tag{2.15}$$

The inner supremum here is attained whenever finite, and it is attained at a point $v = u''$. Thus it equals inf unless there exists a vector $u'' \in \mathbb{R}^m$ such that $[q - Ru - Bu'] -$

$Qu''=0$, in which case it equals $b\cdot u'+\frac{1}{2}u''\cdot Qu''$, a value that actually depends only on u and u'. We may conclude that

$$U_0=\{u\in\mathbb{R}^n \mid \exists u'\in\mathbb{R}^{n'}_+, \text{ with } Ru+Bu'+Qu''=q\}, \tag{2.16}$$

$$f(u)=\text{minimum of } b\cdot u'+\tfrac{1}{2}u''\cdot Qu'' \quad \text{subject to } u'\in\mathbb{R}^{n'}_+,\ u''\in\mathbb{R}^m,\ Ru+Bu'+Qu''=q. \tag{2.17}$$

We can therefore represent (P_0) as

$$\begin{aligned} &\text{minimize} && p\cdot u+\tfrac{1}{2}u\cdot Pu+b\cdot u'+\tfrac{1}{2}u''\cdot Qu'' \\ &\text{subject to} && Au\geq a,\ u'\geq 0,\ Ru+Bu'+Qu''=q, \end{aligned} \tag{$\tilde{\mathrm{P}}_0$}$$

where the value of $u''\cdot Qu''$ does not depend on the particular choice of the vector u'' satisfying $Ru+Bu'+Qu''=q$ but only on u and u'. This is a quadratic programming problem in the usual sense, but in which u'' is a sort of vector of dummy variables that can be eliminated, if desired. In any case it follows that (P_0) has an optimal solution if its optimal value is finite, inasmuch as this property holds for $(\tilde{\mathrm{P}}_0)$.

The optimal solutions $(\bar{u}, \bar{u}', \bar{u}'')$ to $(\tilde{\mathrm{P}}_0)$ are characterized by the Kuhn–Tucker conditions that involve multiplier vectors $\bar{v}$ for the constraint $Ru+Bu'+Qu''=q$ and $\bar{v}'$ for the constraint $Au\geq a$. These conditions take the form:

$$\begin{aligned} &A\bar{u}\geq a, \quad \bar{v}'\geq 0, \quad \bar{v}'\cdot[A\bar{u}'-a]=0, \\ &\bar{u}'\geq 0, \quad B^*\bar{v}\leq b, \quad \bar{u}'\cdot[B^*\bar{v}-b]=0, \\ &R^*\bar{v}+A^*\bar{v}'-P\bar{u}=p, \quad R\bar{u}+B\bar{u}'+Q\bar{u}''=q, \quad Q\bar{u}''=Q\bar{v}. \end{aligned}$$

Because of the final condition we can write the next-to-last condition instead as $R\bar{u}+B\bar{u}'+Q\bar{v}=q$. Note that there is no restriction then on $\bar{u}''$, except that $Q\bar{u}''=Q\bar{v}$; we could always take $\bar{u}''=\bar{v}$ in particular. This is in keeping with our observation that $(\tilde{\mathrm{P}}_0)$ is really just a problem in u and u'. We see in fact that the pairs $(\bar{u}, \bar{u}')$ which are optimal for $(\tilde{\mathrm{P}}_0)$ are the ones which, for some pair $(\bar{v}, \bar{v}')$, satisfy the conditions

$$\begin{aligned} &A\bar{u}\geq 0, \quad \bar{v}'\geq 0, \quad \bar{v}'\cdot[A\bar{u}'-a]=0, \\ &\bar{u}'\geq 0, \quad B^*\bar{v}\leq b, \quad \bar{u}'\cdot[B^*\bar{v}-b]=0, \\ &R\bar{u}+B\bar{u}'+Q\bar{v}=q, \quad R^*\bar{v}+A^*\bar{v}'-P\bar{u}=p. \end{aligned} \tag{2.18}$$

Problem (D_0) can be understood in the same way. From the formula (2.11) for $g(v)$ we deduce that

$$V_0=\{v\in\mathbb{R}^m \mid \exists v'\in\mathbb{R}^{m'}_+,\ v''\in\mathbb{R}^N, \text{ with } R^*v+B^*v'-Pv''=p\}, \tag{2.19}$$

$$g(v)=\text{maximum of } a\cdot v'-\tfrac{1}{2}v''\cdot Pv'' \quad \text{subject to } v'\in\mathbb{R}^{m'}_+,\ v''\in\mathbb{R}^n,\ R^*v+A^*v'-Pv''=p. \tag{2.20}$$

These formulas yield for (D_0) the representation

$$\begin{aligned} &\text{maximize} \quad q \cdot v - \tfrac{1}{2} v \cdot Qv + a \cdot v' - \tfrac{1}{2} v'' \cdot Pv'' \\ &\text{subject to} \quad B^* v \leq b, \ v' \geq 0, \ R^* v + A^* v' - Pv'' = p, \end{aligned} \qquad (\tilde{\mathrm{D}}_0)$$

where the value of $v'' \cdot Pv''$ does not depend on the particular v'' satisfying $R^* v + A^* v' - Pv'' = p$ but only on v and v'. This is really a problem in v and v', and the Kuhn–Tucker conditions characterize $\bar{v}$ and $\bar{v}'$ as optimal if and only if there exist $\bar{u}$ and $\bar{u}'$ such that (2.18) holds, the same conditions as before. Since $(\tilde{\mathrm{D}}_0)$ is a quadratic programming problem in the usual sense, it has an optimal solution whenever its optimal value is finite, and (D_0) therefore has this property too.

Our argument demonstrates that if either (P_0) or (D_0) has finite optimal value, then both problems have optimal solutions. The optimal solutions in both cases are characterized by the existence of auxiliary vectors such that (2.18) holds. But (2.18) can also be seen as the Kuhn–Tucker conditions for $(\bar{u}, \bar{v})$ to be a saddle point of the Lagrangian (2.8), when U and V are given by (2.13). Thus for $\bar{u}$ and $\bar{v}$ to be optimal solutions to (P_0) and (D_0) respectively, it is necessary and sufficient that $(\bar{u}, \bar{v})$ be a saddle point in (2.18). Following on the remarks at the beginning of the proof, this establishes the theorem. □

Corollary. *Any standard quadratic programming method can in principle be used to solve problems of the form* (P_0) *or* (D_0), *in fact both simultaneously, thereby determining a saddle point of the corresponding Lagrangian* l *on* $U \times V$, *if such a saddle point exists.*

Proof. The representations in the proof of the theorem show more specifically that if an algorithm is applied to $(\tilde{\mathrm{P}}_0)$, the optimal solution vectors $\bar{u}$, $\bar{u}'$ and multiplier vectors $\bar{v}$ and $\bar{v}'$ which it produces yield optimal solutions $\bar{u}$ to (P_0) and $\bar{v}$ to (D_0), and $(\bar{u}, \bar{v})$ is a saddle point in (2.8). The same holds if an algorithm is applied to $(\tilde{\mathrm{D}}_0)$, except that then $\bar{v}$ and $\bar{v}'$ are the optimal solution vectors, whereas $\bar{u}$ and $\bar{u}'$ are the multiplier vectors. □

Theorem 2. *The stochastic programming problems* (P) *and* (D) *are representable as quadratic programming problems in the traditional sense, although with potentially very high dimensionality. Both problems have optimal solutions, and*

$$\min(\mathrm{P}) = \max(\mathrm{D}).$$

A pair $(\bar{x}, \bar{z})$ *is a saddle point of the Lagrangian* L *relative to* $X \times Z$ *if and only if* $\bar{x}$ *is an optimal solution to* (P) *and* $\bar{z}$ *is an optimal solution to* (D). *The set of such pairs* $(\bar{x}, \bar{z})$ *is bounded.*

Proof. We need only observe that the triple L, X, Z, can be construed as a special case of the triple l, U, V, in Theorem 1. A term like $Ez_\omega \cdot H_\omega z_\omega$ can be expressed as $z \cdot Qz$ for certain matrix Q, and so forth. Our assumption that the extremal sets $\xi(v)$

in (1.2) and $\zeta_\omega(x)$ in (1.3) are nonempty for all $v \in \mathbb{R}^n$, $x \in X$ and $\omega \in \Omega$, guarantees that every $x \in X$ is feasible for (P), and every $z \in Z$ is feasible for (D). Therefore we are in the case of Theorem 1 where both problems have feasible solutions.

As for the boundedness of the set of saddle points $(\bar{x}, \bar{z})$, consider a particular pair of optimal solutions $\bar{x}^*$ and $\bar{z}^*$ to (P) and (D). Observe that for every optimal solution $\bar{x}$ to (P), $(\bar{x}, \bar{z}^*)$ is a saddle point and therefore satisfies

$$\bar{x} \in \underset{x \in X}{\operatorname{argmin}}\, L(x, \bar{z}^*) = \xi(c - E T_\omega^* \bar{z}_\omega^*)$$

(cf. (2.5)). But the set on the right is bounded (one of our basic assumptions in Section 1). Likewise for every optimal solution $\bar{z}$ to (D), $(\bar{x}^*, \bar{z})$ is a saddle point and therefore satisfies

$$\bar{z} \in \underset{z \in Z}{\operatorname{argmax}}\, L(\bar{x}^*, z), \quad \text{so } \bar{z}_\omega \in \zeta_\omega(\bar{x}^*) \text{ for all } w \in \Omega.$$

(cf. (2.4)). The sets $\zeta_\omega(\bar{x}^*)$ are all bounded (again by one of our basic assumptions in Section 1), so $\bar{z}$ belongs to a certain bounded set. The pairs $(\bar{x}, \bar{z})$ thus all belong to a product of bounded sets dependent only on $\bar{x}^*$ and $\bar{z}^*$. □

The following pair of results will help to clarify the quadratic programming nature of problems (P) and (D).

Proposition 1. *For the function ψ_ω given by (1.1), if the polytope Z_ω has a representation*

$$Z_\omega = \{z_\omega \in \mathbb{R}^m \mid B_\omega^* z_\omega \leq b_\omega\} \tag{2.21}$$

for some vector $b_\omega \in \mathbb{R}^s$ and matrix $B_\omega \in \mathbb{R}^{m \times s}$ (with s independent of ω), then ψ_ω has an alternative expression of the form

$$\begin{aligned} \psi_\omega(x) = {} & \text{minimum of } d_\omega \cdot y_\omega + \tfrac{1}{2} y_\omega \cdot D_\omega y_\omega \\ & \text{subject to } y_\omega \in Y_\omega,\ \bar{T}_\omega x + W_\omega y_\omega = \bar{h}_\omega, \end{aligned} \tag{2.22}$$

for certain vectors $d_\omega \in \mathbb{R}^s$, $\bar{h}_\omega \in \mathbb{R}^q$, and matrices $\bar{T}_\omega \in \mathbb{R}^{q \times n}$, $W_\omega \in \mathbb{R}^{q \times s}$, and $D_\omega \in \mathbb{R}^{s \times s}$ with D_ω symmetric and positive semidefinite, and where

$$Y_\omega = \{y_\omega \in \mathbb{R}^s \mid A_\omega y_\omega \geq a_\omega\} \tag{2.23}$$

for some $a_\omega \in \mathbb{R}^p$ and $A_\omega \in \mathbb{R}^{p \times s}$.

Conversely, any function ψ_ω having a representation (2.22) as just described (with $\psi_\omega(x)$ finite for all $x \in X$) also has a representation (1.1) with Z_ω of the form (2.21).

Proof. Starting with the representation (1.1) and Z_ω of the form (2.21), view the maximization problem in (1.1) as the dual problem associated with the Lagrangian

$$l_{x,\omega}(u_\omega, z_\omega) = u_\omega \cdot [b_\omega - B_\omega^* z_\omega] + z_\omega \cdot [h_\omega - T_\omega x] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega$$
$$\text{for } u_\omega \in \mathbb{R}_+^s \text{ and } z_\omega \in \mathbb{R}^m.$$

The corresponding primal problem, whose optimal value is also equal to $\psi_\omega(x)$ by Theorem 1 (as long as $x \in X$, so that $\psi_\omega(x)$ is finite by assumption) is

$$\text{minimize} \quad b_\omega \cdot u_\omega + f_\omega(u_\omega) \quad \text{over } u_\omega \in \mathbb{R}^s_+, \quad \text{where}$$

$$f_\omega(u_\omega) = \sup_{z_\omega \in \mathbb{R}^m} \{z_\omega \cdot [h_\omega - T_\omega x - B_\omega u_\omega] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\}.$$

Using the trick in the proof of Theorem 1, we can reformulate the latter as

$$\text{minimize} \quad b_\omega \cdot u_\omega + \tfrac{1}{2} u''_\omega \cdot H_\omega u''_\omega$$

$$\text{subject to} \quad u_\omega \in \mathbb{R}^s_+, \ u''_\omega \in \mathbb{R}^m, \ B_\omega u_\omega + H_\omega u''_\omega = h_\omega - T_\omega x.$$

We can then pass to form (2.22) in terms of $y_\omega = (u_\omega, u''_\omega)$ (or by setting $y_\omega = u_\omega$ after algebraic elimination of u''_ω, if the rank of H_ω is the same for all $\omega \in \Omega$).

Starting with the representation (2.22) and Y_ω of the form (2.23), on the other hand, we can view $\psi_\omega(x)$ as the optimal value for the primal problem associated with the Lagrangian

$$l_{x,\omega}(y_\omega, v_\omega) = d_\omega \cdot y_\omega + \tfrac{1}{2} y_\omega \cdot D_\omega y_\omega + v_\omega \cdot [\bar{h}_\omega - \bar{T}_\omega x - W_\omega y_\omega]$$

$$\text{for } y_\omega \in Y_\omega \text{ and } v_\omega \in \mathbb{R}^q.$$

Then $\psi_\omega(x)$ (when finite) is also the optimal value in the corresponding dual problem

$$\text{maximize} \quad v_\omega \cdot [\bar{h}_\omega - \bar{T}_\omega x] + g_\omega(v_\omega) \quad \text{over } v_\omega \in R^q, \quad \text{where}$$

$$g_\omega(v_\omega) = \inf_{y_\omega \in Y_\omega} \{y_\omega \cdot [d_\omega - W^*_\omega v_w] + \tfrac{1}{2} y_\omega \cdot D_\omega y_\omega\}.$$

As we saw in the proof of Theorem 1, this problem can also be written as

$$\text{maximize} \quad v_\omega \cdot [\bar{h}_\omega - \bar{T}_\omega x] + v'_\omega \cdot a_\omega - \tfrac{1}{2} v''_\omega \cdot D_\omega v''_\omega$$

$$\text{subject to} \quad v_\omega \in \mathbb{R}^q, \ v'_\omega \in \mathbb{R}^p_+, \ W^*_\omega v_\omega + A^*_\omega v'_\omega + D_\omega v''_\omega = d_\omega.$$

With $z_\omega = (v_\omega, v'_\omega, v''_\omega)$, this can be brought into the form (1.1) with Z_ω as in (2.21). (Alternatively one could take $z_\omega = (v_\omega, v'_\omega)$ and eliminate v''_ω algebraically, provided that the rank of D_ω is independent of ω. If also the rank of the matrix W_ω is independent of ω, one could even eliminate v_ω from the problem and just take $z_\omega = v'_\omega$ to get a representation (1.1) in fewer variables.) □

Proposition 2. *The function φ in (1.4) also has a representation*

$$\varphi(v) = \text{maximum of } q \cdot u - \tfrac{1}{2} u \cdot Qu \text{ over all } u \in U \text{ satisfying } Bu = v$$

for some choice of vectors b and q and matrices B and Q with Q symmetric and positive semidefinite, where U is a convex polyhedron.

Proof. Recall that $\varphi(v)$ is finite for all v by assumption. Express X as $\{x \in \mathbb{R}^n \mid Ax \geq a\}$ for some $a \in \mathbb{R}^p$ and $A \in \mathbb{R}^{p \times n}$, and consider the Lagrangian

$$l_v(x, u') = v \cdot x + \tfrac{1}{2} x \cdot Cx + u' \cdot [a - Ax] \quad \text{for } x \in \mathbb{R}^n \text{ and } u' \in \mathbb{R}^p_+.$$

The primal problem associated with this Lagrangian is the minimization problem in (1.4), whereas the dual problem, which also has $\varphi(v)$ as its optimal value, is

$$\text{maximize} \quad a \cdot u' + g(u') \quad \text{over } u' \in \mathbb{R}^p_+,$$

$$\text{where } g(u') = \inf_{x \in R^n} \{x \cdot [v - A^* u'] + \tfrac{1}{2} x \cdot Cx\}.$$

The reformulation trick in Theorem 1 translates this into

$$\text{maximize} \quad a \cdot u' - \tfrac{1}{2} u'' \cdot Cu''$$

$$\text{subject to} \quad u' \in \mathbb{R}^p_+, \; u'' \in \mathbb{R}^n, \; A^* u' - Cu'' = v.$$

We can then get a representation (2.24) in terms of $u = (u', u'')$. □

Propositions 1 and 2 make possible a more complete description of the quadratic programming representation of problems (P) and (D) indicated in Theorem 2. When $\psi_\omega(x)$ is expressed in terms of a recourse subproblem in y_ω as in Proposition 1, we can identify (P) with the problem

$$\begin{aligned} &\text{minimize} \quad c \cdot x + \tfrac{1}{2} x \cdot Cx + E\{d_\omega \cdot y_\omega + \tfrac{1}{2} y_\omega \cdot D_\omega y_\omega\} \\ &\text{subject to} \quad x \in X, \; y_\omega \in Y_\omega, \; \bar{T}_\omega x + W_\omega \cdot y_\omega = \bar{h}_\omega \; \text{ for all } \omega \in \Omega. \end{aligned} \tag{2.25}$$

Similarly, when φ is expressed as in Proposition 2 we can pose (D) as

$$\begin{aligned} &\text{maximize} \quad q \cdot u - \tfrac{1}{2} u \cdot Qu + E\{z_\omega \cdot h_\omega - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\} \\ &\text{subject to} \quad u \in U, \; z_\omega \in Z_\omega, \; \text{and } Bu + E\{T^*_\omega z_\omega\} = c. \end{aligned} \tag{2.26}$$

In the latter, our assumption that $\varphi(v)$ is finite for all $v \in \mathbb{R}^n$ implies that no matter what the choice of vectors $z_\omega \in Z_\omega$, there does exist a $u \in U$ such that the constraint $Bu + E\{T^*_\omega z_w\} = c$ is satisfied.

3. Finite generation method

Our aim is to solve problem (P) by way of (D) according to the following scheme. We replace (D) by a sequence of subproblems

$$\text{maximize} \quad G(z) \quad \text{over all} \quad z \in Z^\nu \subset Z \tag{D$^\nu$}$$

for $\nu = 1, 2, \ldots$, where G is the dual objective function in (2.3) and (2.7), and Z' is a polytope of relatively low dimension generated as the convex hull of finitely many points in Z. Obviously (D^ν) is the dual of the problem

$$\text{minimize} \quad F^\nu(x) \quad \text{over all} \quad x \in X, \tag{P$^\nu$}$$

where F^ν is obtained by substituting Z^ν for Z in the formula (2.2) for the primal objective function F:

$$F^\nu(x) = \max_{z \in Z^\nu} L(x, z)$$
$$= c \cdot x + \tfrac{1}{2}x \cdot Cx + \max_{z \in Z^\nu} E\{z_\omega \cdot [h_\omega - T_\omega x] - \tfrac{1}{2}z_\omega \cdot H_\omega z_\omega\}. \tag{3.1}$$

Indeed, (P^ν) and (D^ν) are the primal and dual problems that correspond to L on $X \times Z^\nu$ rather than $X \times Z$. In calculating a solution $\bar{z}^\nu$ to (D^ν) we obtain also a solution $\bar{x}^\nu$ to (P^ν) that can be viewed as an approximately optimal solution to (P). From $\bar{z}^\nu$ and $\bar{x}^\nu$ we gain information that helps in determining the polytope $Z^{\nu+1}$ to be used in the next iteration. The new polytope $Z^{\nu+1}$ is not necessarily 'larger' than Z^ν.

Problems (P^ν) and (D^ν) belong to the realm of 'generalized' quadratic programming as demarcated in Section 2. Clearly

$$F(x) \geq F^\nu(x) \quad \text{for all } x, \tag{3.2}$$

where F is the primal objective function in (2.2) and (2.6), so (P^ν) can be regarded as a 'lower envelope approximation' to (P). The feasible sets in (P^ν) and (D^ν) are X and Z^ν, respectively, whereas the ones in (P) and (D), are X and Z. From Theorem 1, therefore, we know that optimal solutions $\bar{x}^\nu$ and $\bar{z}^\nu$ to (P^ν) and (D^ν) exist and satisfy

$$F^\nu(\bar{x}^\nu) = G(\bar{z}^\nu), \tag{3.3}$$

$$\bar{x}^\nu \in \operatorname*{argmin}_{x \in X} F^\nu(x) \subset \operatorname*{argmin}_{x \in X} L(x, \bar{z}^\nu), \tag{3.4}$$

$$\bar{z}^\nu \in \operatorname*{argmax}_{z \in Z^\nu} G(z) \subset \operatorname*{argmax}_{z \in Z^\nu} L(\bar{x}^\nu, z). \tag{3.5}$$

Having determined a pair $(\bar{x}^\nu, \bar{z}^\nu)$ of this type, which is a saddle point of L relative to $X \times Z^\nu$, we can test whether it is actually a saddle point of L relative to $X \times Z$. This amounts to checking the maximum of $L(\bar{x}^\nu, z)$ over all $z \in Z$ to see if it occurs at $z = \bar{z}^\nu$. If yes, $\bar{x}^\nu$ and $\bar{z}^\nu$ are optimal solutions to (P) and (D), and we are done. If no, we obtain from the test an element

$$z^\nu \in \operatorname*{argmax}_{z \in Z} L(\bar{x}^\nu, z) \tag{3.6}$$

and have

$$L(\bar{x}^\nu, \bar{z}^\nu) < L(\bar{x}^\nu, z^\nu) = F(\bar{x}^\nu). \tag{3.7}$$

The crucial feature that makes the test possible is the decomposition in (2.4): maximizing $L(\bar{x}^\nu, z)$ in $z \in Z$ reduces a solving a separate quadratic programming problem (perhaps trivial) in $z_\omega \in Z_\omega$ for each $\omega \in \Omega$. Anyway, with such a z^ν we have

$$F(x) \geq L(x, z^\nu) \quad \text{for all } x, \text{ with equality when } x = \bar{x}^\nu. \tag{3.8}$$

We can use this in conjunction with (3.3) in constructing a new lower envelope approximation $F^{\nu+1}$ for F, which in primal terms is what is involved in constructing a new set $Z^{\nu+1}$ to replace Z^ν. More will be said about this later.

Of course the optimality test also furnishes a criterion for termination with suboptimal solutions, if desired. Since $\bar{x}^\nu$ and $\bar{z}^\nu$ are feasible solutions to (P) and (D) and satisfy (by Theorem 2)

$$F(\bar{x}^\nu) \geq \min(\mathrm{P}) = \max(\mathrm{D}) \geq G(\bar{z}^\nu), \tag{3.9}$$

we know that for $\varepsilon_\nu = F(\bar{x}^\nu) - G(\bar{z}^\nu)$, both $\bar{x}^\nu$ and $\bar{z}^\nu$ are ε_ν-optimal:

$$|F(\bar{x}^\nu) - \min(\mathrm{P})| \leq \varepsilon_\nu \quad \text{and} \quad |G(\bar{x}^\nu) - \max(\mathrm{D})| \leq \varepsilon_\nu. \tag{3.10}$$

Our basic procedure can be summarized now as follows.

Algorithm

Step 0 (Initialization). Choose the optimality test parameter $\bar{\varepsilon} \geq 0$ and the initial convex polytope $Z^1 \subset Z$. Set $\nu = 1$.

Step 1 (Approximate Solution). Determine a saddle point $(\bar{x}^\nu, \bar{z}^\nu)$ of L relative to $X \times Z^\nu$ and the value $\bar{a}_\nu = L(\bar{x}^\nu, \bar{z}^\nu)$.

Step 2 (Decomposition). For each $\omega \in \Omega$, determine an optimal solution z_ω^ν to the problem

$$\text{maximize} \quad z_\omega \cdot [h_\omega - T_\omega \bar{x}^\nu] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega \ \text{ over } z_\omega \in Z_\omega \tag{3.11}$$

and the optimal value α_ω^ν. Let z^ν be the element of Z having component z_ω^ν in Z_ω, and let

$$\alpha_\nu = c \cdot \bar{x}^\nu + \tfrac{1}{2} \bar{x}^\nu \cdot C \bar{x}^\nu + E \alpha_\omega^\nu = L(\bar{x}^\nu, z^\nu). \tag{3.12}$$

Step 3 (Optimality Test). Let $\varepsilon_\nu = \alpha_\nu - \bar{\alpha}_\nu$. Then $\bar{x}^\nu$ is an ε_ν-optimal solution to (P), $\bar{z}^\nu$ is an ε_ν-optimal solution to (D), and

$$\alpha_\nu \geq \min(\mathrm{P}) = \max(\mathrm{D}) \geq \bar{\alpha}_\nu. \tag{3.13}$$

If $\varepsilon_\nu \leq$ terminate.

Step 4 (Polytope Modification). Choose a new convex polytope $Z^{\nu+1}$ that contains both $\bar{z}^\nu$ and z^ν, although not necessarily all of Z^ν. Replace ν by $\nu+1$; return to Step 1.

We proceed to comment on these algorithmic steps individually in more detail, one by one. Properties of the algorithm as a whole will be developed in Section 4 and Section 5.

The most important observation concerns the quadratic programming nature of the subproblem solved in Step 1. Suppose that Z^ν is generated from certain elements $\tilde{z}_k^\nu \in Z$:

$$Z^\nu = \mathrm{co}\{\tilde{z}_k^\nu | k = 1, \ldots, m_\nu\} = \left\{ \sum_{k=1}^{m_\nu} \lambda_k \tilde{z}_k^\nu \,\middle|\, \lambda_k \geq 0, \ \sum_{k=1}^{m_\nu} \lambda_k = 1 \right\}. \tag{3.14}$$

Finding a saddle point $(\bar{x}^\nu, \bar{z}^\nu)$ of $L(x, z)$ relative to $x \in X$ and $z \in Z^\nu$ is equivalent to finding a saddle point $(\bar{x}^\nu, \bar{\lambda}^\nu)$ of

$$L^\nu(x, \lambda) = L\left(x, \sum_{k=1}^{m_\nu} \lambda_k \tilde{z}_k^\nu\right) \tag{3.15}$$

relative to $x \in X$ and $\lambda \in \Lambda^\nu$, where Λ^ν is the unit simplex in $\mathbb{R}^{m_\nu}$,

$$\Lambda^\nu := \left\{\lambda = (\lambda_1, \ldots, \lambda_{m_\nu}) \mid \lambda_k \geq 0, \sum_{k=1}^{m_\nu} \lambda_k = 1\right\}, \tag{3.16}$$

and then setting

$$\bar{z}^\nu = \sum_{k=1}^{m_\nu} \lambda_k \tilde{z}_k^\nu. \tag{3.17}$$

But from the definition (2.1) of $L(x, z)$ we have

$$\begin{aligned} L^\nu(x, \lambda) &= c \cdot x + \tfrac{1}{2} x \cdot Cx + \sum_{k=1}^{m_\nu} \lambda_k E\{\tilde{z}_{k\omega}^\nu \cdot [h_\omega - T_\omega x]\} \\ &\quad - \tfrac{1}{2} \sum_{j=1}^{m_\nu} \sum_{k=1}^{m_\nu} \lambda_j \lambda_k E\{\tilde{z}_{k\omega}^\nu \cdot H_\omega \tilde{z}_{k\omega}^\nu\} \\ &= c \cdot x + \tfrac{1}{2} x \cdot Cx + \lambda \cdot [\tilde{h}^\nu - \tilde{T}^\nu x] - \tfrac{1}{2} \lambda \cdot \tilde{H}^\nu \lambda, \end{aligned} \tag{3.18}$$

where

$$\tilde{h}^\nu \in \mathbb{R}^{m_\nu} \quad \text{with components } \tilde{h}_k^\nu = E\{\tilde{z}_{k\omega}^\nu \cdot h_\omega\}, \tag{3.19}$$

$$\tilde{H}^\nu \in \mathbb{R}^{m_\nu \times m_\nu} \quad \text{with entries } \tilde{H}_{jk}^\nu = E\{\tilde{z}_{j\omega}^\nu \cdot H_\omega \tilde{z}_{k\omega}^\nu\}, \tag{3.20}$$

$$\tilde{T}^\nu \in \mathbb{R}^{m_\nu \times n} \quad \text{with entries } \tilde{T}_{ki}^\nu = E\{\tilde{z}_{k\omega}^\nu \cdot T_\omega^i\}, \tag{3.21}$$

T_ω^i being the ith column of the matrix $T_\omega \in \mathbb{R}^{m \times m}$. Problem (D^ν) thus reduces to a *deterministic quadratic programming problem in which the coefficients are certain expectations*:

$$\text{maximize} \quad \phi(c - \tilde{T}^{\nu *} \lambda) - \tfrac{1}{2} \lambda \cdot \tilde{H}^\nu \lambda \quad \text{over all } \lambda \in \Lambda^\nu. \tag{$\tilde{\mathrm{D}}^\nu$}$$

Here φ is the function in (1.4), which has alternative representations such as in Proposition 2 that can be used to place $(\tilde{\mathrm{D}}^\nu)$ in a more traditional quadratic programming format. Regardless of such reformulation, the dimensionality of this quadratic programming problem will be relatively low as long as m_ν, the number of elements $\tilde{z}_k^\nu$ used in generating Z^ν, is kept modest.

The translation of (D^ν) into $(\tilde{\mathrm{D}}^\nu)$ also sheds light on the lower envelope function F^ν in the approximate primal subproblem (P^ν):

$$F^\nu(x) = \max_{\lambda \in \Lambda^\nu} L(x, \lambda) = c \cdot x + \tfrac{1}{2} x \cdot Cx + \Psi^\nu(x), \tag{3.22}$$

where

$$\Psi^\nu(x) := \max_{\lambda \in \Lambda^\nu} \{\lambda \cdot [\tilde{h}^\nu - \tilde{T}^\nu x] - \tfrac{1}{2} \lambda \cdot \tilde{H}^\nu \lambda\} = \max_{z \in Z^\nu} E\{z_\omega \cdot [h_\omega - T_\omega x] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\}. \tag{3.23}$$

Clearly $\Psi^\nu(x)$ is a lower envelope approximation to the recourse cost function

$$\Psi(x) := \max_{z \in Z} E\{z_\omega \cdot [h_\omega - T_\omega x] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\} = E\psi_\omega(x). \tag{3.24}$$

Especially worth noting in (3.23) is the case where there are no quadratic terms $z_\omega \cdot H_\omega z_\omega$, i.e. where $H_\omega = 0$ for all $\omega \in \Omega$ and consequently $\tilde{H}^\nu = 0$. Then

$$\Psi^\nu(x) = \max_{k=1,\dots,m_\nu} \{\tilde{h}_k^\nu - \tilde{T}_k^\nu \cdot x\}, \tag{3.25}$$

where $\tilde{T}_k^\nu$ is the vector in $\mathbb{R}^n$ given by the kth row of the matrix $\tilde{T}^\nu$ in (3.21):

$$\tilde{T}_k^\nu = E\{T_\omega^* \tilde{z}_{k\omega}^\nu\}. \tag{3.26}$$

In this case Ψ^ν is a *polyhedral* convex envelope representation of Ψ, the pointwise maximum of a collection of affine functions

$$l_k(x) = \tilde{h}_k^\nu - \tilde{T}_k^\nu \cdot x \quad \text{for } k = 1, \dots, m_\nu.$$

Our technique then resembles a cutting-plane method, at least as far as the function Ψ is concerned.

Indeed, if not only $H_\omega = 0$ but $C = 0$, so that there are no quadratic cost terms at all and (P) is a purely linear stochastic programming problem, we can regard F^ν as a polyhedral convex subrepresentation of F. Then the subproblems (P^ν) and ($\tilde{\mathrm{D}}^\nu$) can be solved by linear rather than quadratic programming algorithms. Furthermore the function $L(x, z^\nu)$ determined in (3.8) is then affine in x. If in fact we were to take $Z^{\nu+1} = \mathrm{co}\{Z^\nu, z^\nu\}$, we would get

$$F^{\nu+1}(x) = \max\{F^\nu(x), L(x, z^\nu)\},$$

and this would truly be a cutting-plane method applied to problem (P).

It must be remembered, though, that in such a cutting-plane approach it might generally be necessary to retain more and more affine functions in the polyhedral approximation to F, since the conditions that theoretically validate the dropping of earlier cutting-planes might not be met. The dimension of the linear programming subproblem to be solved in each iteration would become progressively larger. In contrast, by taking advantage of the quadratic structure even to the extent of introducing it when it is not already at hand (as proposed in Section 5), one can avoid the escalation of dimensionality and at the same time get convergence results of a superior character (as presented in Section 4).

Note that with a nonvanishing quadratic term $\lambda \cdot \tilde{H}^\nu \lambda$ in (3.23) (the matrix $\tilde{H}^\nu$ being positive semidefinite, of course) the lower envelope approximation Ψ^ν to Ψ will generally *not* be polyhedral but have 'rounded corners'. As a matter of fact, if $\tilde{H}^\nu$ is nonsingular, then Ψ^ν is a smooth convex function with Lipschitz continuous derivatives.

In Step 2 of the algorithm, we need to solve a potentially large number of quadratic programming problems (3.11) in the vectors z_ω. This could be a trouble spot. If the problems are complicated and require full application of some quadratic programming routine, the secret to success would have to lie in taking advantage of the similarities between neighboring problems. Techniques of parametric programming and 'bunching' might be useful. Not to be overlooked, however, are the situations in which each problem (3.11) decomposes further into something simpler.

Especially important is the case where

$$Z_\omega = Z_{\omega 1} \times Z_{\omega 2} \times \cdots \times Z_{\omega r} \tag{3.27}$$

and H_ω does not involve cross terms between the sets in this product:

$$H_\omega = \text{diag}[H_{\omega 1}, H_{\omega 2}, \ldots, H_{\omega r}]. \tag{3.28}$$

Then (3.11) reduces to a separate problem over each of the sets $Z_{\omega 1}, \ldots, Z_{\omega r}$. If these sets are actually intervals (bounded or unbounded), then the separate problems are one-dimensional, and their solutions can be given *in closed form.* Such is indeed what happens when the costs $\psi_\omega(x)$ in (P) are penalties $\theta_\omega(h_\omega - T_\omega x)$ as in (1.5), (1.6), (1.7), and $\theta_\omega(h_\omega - T_\omega x)$ is a sum of separate terms, one for each real component of the vector $h_\omega - T_\omega x$. The special model we have treated in [11] makes use of this simplification. In such a setting the vector z^ν is readily computed as a simple function of $\bar{x}^\nu$, and indeed one can get away with storing only $\bar{x}^\nu$, which has only a small number of components compared to z^ν; cf. [4].

The product form (3.27) for Z_ω, if it is present, also raises further possibilities for structuring the subproblems introduced in Step 1, by the way. One could write

$$Z = Z_1 \times \cdots \times Z_r \quad \text{with } Z_j = \Pi_{\omega \in \Omega} Z_{\omega j} \tag{3.29}$$

and work with polytopes of the form

$$Z^\nu = Z_1^\nu \times \cdots \times Z_r^\nu \quad \text{with } Z_j^\nu \subset Z_j, \tag{3.30}$$

for instance. This could be advantageous in holding the dimensionality down. If each Z_j^ν is generated as the convex hull of a finite subset of Z_j consisting of n_ν elements, one can get away with describing the points of Z^ν by rn_ν parameters λ_{jk}. On the other hand, if Z^ν is regarded as the convex hull of the product of these finite subsets of $Z_1, \ldots, Z_r$, one would need $(n_\nu)^r$ parameters.

The procedure invoked in Step 4 of the algorithm has been left open to various possibilities, which could be influenced too by such considerations as the foregoing. Two basic possibilities that immediately come to mind are:

$$Z^{\nu+1} = \text{co}\{\bar{z}^\nu, z^\nu\} \quad \text{(generalized Frank-Wolfe rule)} \tag{3.31}$$

and

$$Z^{\nu+1} = \text{co}\{Z^\nu, z^\nu\} \quad \text{(generalized cutting-plane rule).} \tag{3.32}$$

The first of these is adequate for convergence if the matrix C is positive definite, as we shall see in Section 4. It is certaintly the simplest but might suffer from too much information being thrown away between one iteration of Step 1 and the next. It gets its name from the interpretation in terms of problem (D) that will underly the proof of Theorem 5.

The second formula goes to the opposite extreme. It achieves better and better representations of the primal objective F, in the sense that

$$F(x) \geq F^{\nu+1}(x) \geq \max\{F^{\nu}(x), L(x, z^{\nu})\} \quad \text{for all } x, \quad \text{with}$$
$$F(\bar{x}^{\nu}) = F^{\nu+1}(\bar{x}^{\nu}) = L(\bar{x}^{\nu}, z^{\nu}), \tag{3.33}$$

but this is at the expense of keeping all information and continually enlarging the size of the quadratic programming subproblem. A good compromise possibility might be

$$Z^{\nu+1} = \text{co}\{Z^1, \bar{z}^{\nu}, z^{\nu}\}, \tag{3.34}$$

where Z^1 is the fixed initial polytope.

This brings us to the choice of Z^1 in Step 0, which in determining the first approximate solutions $\bar{x}^1$ and $\bar{z}^1$ could have a big effect on the progress of the computations. We can, of course, start with $Z^1 = \{\hat{z}\}$, where $\hat{z}$ is an element of Z that may be regarded as an estimate for an optimal solution to (D). For example, if an initial guess $\hat{x}$ is available for an optimal solution to (P), one might take $\hat{z}$ to be a vector constructed by calculating an element $\hat{z}_{\omega} \in \zeta_{\omega}(\hat{x})$ for each ω. This approach makes sense especially in situations where $\zeta_{\omega}(\hat{x})$ is a singleton for each $\omega \in \Omega$, so that $\hat{z}$ is uniquely determined by the estimate $\hat{x}$.

Another approach to the initial Z^1 requires no guesses or prior information about solutions. A fixed number of elements $a_{k\omega}$ $(k = 1, \ldots, p)$ is chosen from each Z_{ω}, such as the set of extreme points of Z_{ω} augmented by some selected internal points. These yield p elements a_k of Z, where a_k has component $a_{k\omega}$ in Z_{ω}. The convex hull of these a_k's can be taken as Z^1. Such an approach to initialization has turned out to be very effective in the case of our special model in [11] when adapted to a product structure (3.27); see King [4].

In summary, there are many possibilities for choosing the initial polytope Z^1 in Step 0 and modifying it iteratively in Step 4. They can be tailored to the structure of the problem. Various product representations of Z and Z^{ν} could be helpful in particular. Versions of rules (3.31), (3.32), and (3.34), which maintain the product form, can be developed.

See the end of Section 4 for other comments on forming $Z^{\nu+1}$ from Z^{ν}.

4. Convergence results

Properties of the sequences produced by the finite generation algorithm in Section 3 will now be derived. For this purpose we ignore the optimality test in Step 3 of the algorithm, since our interest is centered on what happens when the procedure is iterated indefinitely. Unless otherwise indicated, our assumptions are merely the basic ones in Section 1. The initial polytope Z^1 is arbitrary, and $Z^{\nu+1}$ is not subjected to any requirement stricter than the one in Step 4, namely that $Z^{\nu+1} \supset \{\bar{z}^{\nu}, z^{\nu}\}$. In addition to the symbols already introduced in the statement of the algorithm in

Section 3 we use the supplementary notation

$$\bar{\alpha} = \min(\mathrm{P}) = \max(\mathrm{D}), \tag{4.1}$$

$$\bar{\varepsilon}_\nu = \bar{\alpha} - \bar{\alpha}_\nu, \tag{4.2}$$

$$\bar{w}^\nu = \nabla_x L(\bar{x}^\nu, \bar{z}^\nu) = c + C\bar{x}^\nu - ET^*_\omega \bar{z}^\nu_\omega, \tag{4.3}$$

$$\|x\|_C = [x \cdot Cx]^{1/2}. \tag{4.4}$$

Of course $\|x\|_C$ is a norm on $\mathbb{R}^n$ if C is positive definite. If C is only positive semidefinite, then $\|x\|_C$ vanishes on the subspace $\{x \in \mathbb{R}^n \mid Cx = 0\}$ but is positive elsewhere.

Theorem 3. *The sequences* $\{\bar{x}^\nu\}$, $\{\bar{z}^\nu\}$, *and* $\{z^\nu\}$ *are bounded and satisfy*

$$F(\bar{x}^\nu) = \alpha_\nu \geq \bar{\alpha} \geq \cdots \geq \bar{\alpha}_{\nu+1} \geq \bar{\alpha}_\nu = G(\bar{z}^\nu). \tag{4.5}$$

Furthermore one has the estimate.

$$\tfrac{1}{2}\|\bar{x} - \bar{x}^\nu\|_C^2 \leq \bar{\varepsilon}_\nu - \bar{w}^\nu \cdot (\bar{x} - \bar{x}^\nu) \leq \bar{\varepsilon}_\nu \leq \bar{\varepsilon}_\nu \tag{4.6}$$

for every optimal solution $\bar{x}$ *to* (P), *where*

$$\bar{w}^\nu \cdot (x - \bar{x}^\nu) \geq 0 \quad \text{for every } x \in X. \tag{4.7}$$

If $\varepsilon_\nu \to 0$, *then every cluster point of* $\{\bar{x}^\nu\}$ *is an optmal solution to* (P), *and every cluster point of* $\{\bar{z}^\nu\}$ *is an optimal solution to* (D).

Proof. We have $\bar{\alpha}_\nu = L(\bar{x}^\nu, \bar{z}^\nu)$ and $\alpha_\nu = L(\bar{x}^\nu, z^\nu)$ by definition, so $F(\bar{x}^\nu) = \alpha_\nu$ by (3.7). Then $\alpha_\nu \geq \bar{\alpha} \geq \bar{\alpha}_\nu$ by (3.9). By the same token, $G(\bar{z}^{\nu+1}) = \bar{\alpha}_{\nu+1}$ and $\bar{\alpha} \geq \bar{\alpha}_{\nu+1}$. But also

$$G(\bar{z}^{\nu+1}) = \max_{z \in Z^{\nu+1}} G(z) \geq G(\bar{z}^\nu)$$

because $\bar{z}^\nu \in Z^{\nu+1}$. All the relations in (4.5) are therefore correct.

Next we verify that the sequence $\{\bar{z}^\nu\}$ is bounded. Recall that G is a continuous concave function on Z, since G is given by (2.7), where φ is the concave function defined by (1.4); our basic assumption about the sets $\xi(v)$ being bounded implies φ is finite everywhere. (As is well known, a concave function is continuous at a point if it is finite on a neighborhood of the point [9, Theorem 10.1].) We know from (4.5) that the sequence $\{G(\bar{z}^\nu)\}$ us nondecreasing, so the boundedness of $\{\bar{z}^\nu\}$ can be established by showing that the set $\{z \in Z \mid G(z) \geq G(\bar{z}^1)\}$ is bounded. Consider the closed concave function

$$g(z) = \begin{cases} G(z) & \text{if } z \in Z, \\ \infty & \text{if } z \notin Z. \end{cases}$$

We wish to show that a certain level set $\{z \mid g(z) \geq \bar{\alpha}_1\}$ is bounded. But the level sets $\{z \mid g(z) \geq \alpha\}$, $\alpha \in R$, are all bounded if merely one of them is bounded and nonempty (see [9, Corollary 8.7.1]). In the present case we know that the level set

$$\{z \mid g(z) \geq \bar{\alpha}\} = [\text{set of all optimal solutions to (D)}]$$

is bounded and nonempty (Theorem 2). Therefore the set $\{z \in Z \mid G(z) \geq G(\bar{z}^1)\}$ is indeed bounded, and the sequence $\{\bar{z}^\nu\}$ is bounded as claimed.

We invoke now the fact that

$$\bar{x}^\nu \in \xi(c - ET_\omega^* \bar{z}_\omega^\nu) \quad \text{for all } \nu, \tag{4.8}$$

which is true by (2.5) because $(\bar{x}^\nu, \bar{z}^\nu)$ is a saddle point of L relative to $X \times Z^\nu$. In terms of the finite concave function φ we have

$$\xi(v) = \partial\varphi(v) \quad \text{for all } v \in \mathbb{R}^n. \tag{4.9}$$

Indeed, (1.4) defines φ as the conjugate of the closed proper concave function

$$\gamma(x) = \begin{cases} -\frac{1}{2}x \cdot Cx & \text{if } x \in X, \\ \infty & \text{if } x \notin X, \end{cases}$$

so $\partial\varphi(v)$ consists of the points x which minimize $v \cdot x - \gamma(x)$ over $\mathbb{R}^n$ (see [9, Theorem 23.5]). These are the points that make up the set $\xi(v)$ in (1.2). Thus

$$\bar{x}^\nu \in \partial\varphi(\bar{v}^\nu) \quad \text{for all } \nu, \quad \text{where } \bar{v}^\nu = c - ET_\omega^* \bar{z}_\omega^\nu. \tag{4.10}$$

The sequence $\{\bar{v}^\nu\}$ is bounded, since $\{\bar{z}^\nu\}$ is. Moreover the multifunction $\partial\varphi$ is locally bounded: for every $\bar{v} \in \mathbb{R}^n$ there is a $\delta > 0$ such that the set $\bigcup \{\partial\varphi(v) \mid |v - \bar{v}| \leq \delta\}$ is bounded (see [9, Corollary 24.5.1]). It follows by a simple compactness argument that $\partial\varphi$ carries bounded sets into bounded sets: if $V \subset \mathbb{R}^n$ is bounded, then $\bigcup \{\partial\varphi(v) \mid v \in V\}$ is bounded. Taking $V = \{\bar{v}^\nu\}$, we conclude that the sequence $\{\bar{x}^\nu\}$ is bounded.

The argument establishing that $\{z^\nu\}$ is bounded is similar. We have $z_\omega^\nu \in \zeta_\omega(\bar{x}^\nu)$, where ζ_ω is the multifunction defined in (1.3). Since the sequence $\{\bar{x}^\nu\}$ is now known to be bounded, we need only show that ζ_ω is locally bounded at every $\bar{x}^\nu$ in order to conclude that each of the sequences $\{z_\omega^\nu\}$ is bounded and consequently that $\{z^\nu\}$ is bounded.

In terms of the convex function θ_ω defined in (1.5) we have

$$\zeta_\omega(x) = \partial\theta_\omega(h_\omega - T_\omega x) \quad \text{for all } x \in X. \tag{4.11}$$

This holds because (1.5) expresses θ_ω as the conjugate of the closed proper convex function

$$f_\omega(z_\omega) = \begin{cases} \frac{1}{2}z_\omega \cdot H_\omega z_\omega & \text{if } z_\omega \in Z_\omega, \\ \infty & \text{if } z_\omega \notin Z_\omega. \end{cases}$$

The vectors $z_w \in \partial\theta_\omega(u)$ are therefore the ones that maximize $u \cdot z_\omega - f_\omega(z_\omega)$ (see [9, Theorem 23.5]). Our assumption that $\zeta_\omega(x)$ is nonempty and bounded for every $x \in X$ means that $\partial\theta_\omega(u)$ is nonempty and bounded for every u of the form $h_\omega - T_\omega x$ for some $x \in X$. Every such $u = h_\omega - T_\omega x$ therefore belongs to int(dom θ_ω) (cf.

Section 3 we use the supplementary notation

$$\bar{\alpha} = \min(\mathrm{P}) = \max(\mathrm{D}), \tag{4.1}$$

$$\bar{\varepsilon}_\nu = \bar{\alpha} - \bar{\alpha}_\nu, \tag{4.2}$$

$$\bar{w}^\nu = \nabla_x L(\bar{x}^\nu, \bar{z}^\nu) = c + C\bar{x}^\nu - ET^*_\omega \bar{z}^\nu_\omega, \tag{4.3}$$

$$\|x\|_C = [x \cdot Cx]^{1/2}. \tag{4.4}$$

Of course $\|x\|_C$ is a norm on $\mathbb{R}^n$ if C is positive definite. If C is only positive semidefinite, then $\|x\|_C$ vanishes on the subspace $\{x \in \mathbb{R}^n \mid Cx = 0\}$ but is positive elsewhere.

Theorem 3. *The sequences* $\{\bar{x}^\nu\}$, $\{\bar{z}^\nu\}$, *and* $\{z^\nu\}$ *are bounded and satisfy*

$$F(\bar{x}^\nu) = \alpha_\nu \geq \bar{\alpha} \geq \cdots \geq \bar{\alpha}_{\nu+1} \geq \bar{\alpha}_\nu = G(\bar{z}^\nu). \tag{4.5}$$

Furthermore one has the estimate.

$$\tfrac{1}{2}\|\bar{x} - \bar{x}^\nu\|_C^2 \leq \bar{\varepsilon}_\nu - \bar{w}^\nu \cdot (\bar{x} - \bar{x}^\nu) \leq \bar{\varepsilon}_\nu \leq \bar{\varepsilon}_\nu \tag{4.6}$$

for every optimal solution $\bar{x}$ *to* (P), *where*

$$\bar{w}^\nu \cdot (x - \bar{x}^\nu) \geq 0 \quad \text{for every } x \in X. \tag{4.7}$$

If $\varepsilon_\nu \to 0$, *then every cluster point of* $\{\bar{x}^\nu\}$ *is an optmal solution to* (P), *and every cluster point of* $\{\bar{z}^\nu\}$ *is an optimal solution to* (D).

Proof. We have $\bar{\alpha}_\nu = L(\bar{x}^\nu, \bar{z}^\nu)$ and $\alpha_\nu = L(\bar{x}^\nu, z^\nu)$ by definition, so $F(\bar{x}^\nu) = \alpha_\nu$ by (3.7). Then $\alpha_\nu \geq \bar{\alpha} \geq \bar{\alpha}_\nu$ by (3.9). By the same token, $G(\bar{z}^{\nu+1}) = \bar{\alpha}_{\nu+1}$ and $\bar{\alpha} \geq \bar{\alpha}_{\nu+1}$. But also

$$G(\bar{z}^{\nu+1}) = \max_{z \in Z^{\nu+1}} G(z) \geq G(\bar{z}^\nu)$$

because $\bar{z}^\nu \in Z^{\nu+1}$. All the relations in (4.5) are therefore correct.

Next we verify that the sequence $\{\bar{z}^\nu\}$ is bounded. Recall that G is a continuous concave function on Z, since G is given by (2.7), where φ is the concave function defined by (1.4); our basic assumption about the sets $\xi(v)$ being bounded implies φ is finite everywhere. (As is well known, a concave function is continuous at a point if it is finite on a neighborhood of the point [9, Theorem 10.1].) We know from (4.5) that the sequence $\{G(\bar{z}^\nu)\}$ us nondecreasing, so the boundedness of $\{\bar{z}^\nu\}$ can be established by showing that the set $\{z \in Z \mid G(z) \geq G(\bar{z}^1)\}$ is bounded. Consider the closed concave function

$$g(z) = \begin{cases} G(z) & \text{if } z \in Z, \\ \infty & \text{if } z \notin Z. \end{cases}$$

We wish to show that a certain level set $\{z \mid g(z) \geq \bar{\alpha}_1\}$ is bounded. But the level sets $\{z \mid g(z) \geq \alpha\}$, $\alpha \in R$, are all bounded if merely one of them is bounded and nonempty (see [9, Corollary 8.7.1]). In the present case we know that the level set

$$\{z \mid g(z) \geq \bar{\alpha}\} = [\text{set of all optimal solutions to (D)}]$$

is bounded and nonempty (Theorem 2). Therefore the set $\{z \in Z \mid G(z) \geq G(\bar{z}^1)\}$ is indeed bounded, and the sequence $\{\bar{z}^\nu\}$ is bounded as claimed.

We invoke now the fact that

$$\bar{x}^\nu \in \xi(c - ET_\omega^* \bar{z}_\omega^\nu) \quad \text{for all } \nu, \tag{4.8}$$

which is true by (2.5) because $(\bar{x}^\nu, \bar{z}^\nu)$ is a saddle point of L relative to $X \times Z^\nu$. In terms of the finite concave function φ we have

$$\xi(v) = \partial\varphi(v) \quad \text{for all } v \in \mathbb{R}^n. \tag{4.9}$$

Indeed, (1.4) defines φ as the conjugate of the closed proper concave function

$$\gamma(x) = \begin{cases} -\frac{1}{2}x \cdot Cx & \text{if } x \in X, \\ \infty & \text{if } x \notin X, \end{cases}$$

so $\partial\varphi(v)$ consists of the points x which minimize $v \cdot x - \gamma(x)$ over $\mathbb{R}^n$ (see [9, Theorem 23.5]). These are the points that make up the set $\xi(v)$ in (1.2). Thus

$$\bar{x}^\nu \in \partial\varphi(\bar{v}^\nu) \quad \text{for all } \nu, \quad \text{where } \bar{v}^\nu = c - ET_\omega^* \bar{z}_\omega^\nu. \tag{4.10}$$

The sequence $\{\bar{v}^\nu\}$ is bounded, since $\{\bar{z}^\nu\}$ is. Moreover the multifunction $\partial\varphi$ is locally bounded: for every $\bar{v} \in \mathbb{R}^n$ there is a $\delta > 0$ such that the set $\bigcup \{\partial\varphi(v) \mid |v - \bar{v}| \leq \delta\}$ is bounded (see [9, Corollary 24.5.1]). It follows by a simple compactness argument that $\partial\varphi$ carries bounded sets into bounded sets: if $V \subset \mathbb{R}^n$ is bounded, then $\bigcup \{\partial\varphi(v) \mid v \in V\}$ is bounded. Taking $V = \{\bar{v}^\nu\}$, we conclude that the sequence $\{\bar{x}^\nu\}$ is bounded.

The argument establishing that $\{z^\nu\}$ is bounded is similar. We have $z_\omega^\nu \in \zeta_\omega(\bar{x}^\nu)$, where ζ_ω is the multifunction defined in (1.3). Since the sequence $\{\bar{x}^\nu\}$ is now known to be bounded, we need only show that ζ_ω is locally bounded at every $\bar{x}^\nu$ in order to conclude that each of the sequences $\{z_\omega^\nu\}$ is bounded and consequently that $\{z^\nu\}$ is bounded.

In terms of the convex function θ_ω defined in (1.5) we have

$$\zeta_\omega(x) = \partial\theta_\omega(h_\omega - T_\omega x) \quad \text{for all } x \in X. \tag{4.11}$$

This holds because (1.5) expresses θ_ω as the conjugate of the closed proper convex function

$$f_\omega(z_\omega) = \begin{cases} \frac{1}{2}z_\omega \cdot H_\omega z_\omega & \text{if } z_\omega \in Z_\omega, \\ \infty & \text{if } z_\omega \notin Z_\omega. \end{cases}$$

The vectors $z_w \in \partial\theta_\omega(u)$ are therefore the ones that maximize $u \cdot z_\omega - f_\omega(z_\omega)$ (see [9, Theorem 23.5]). Our assumption that $\zeta_\omega(x)$ is nonempty and bounded for every $x \in X$ means that $\partial\theta_\omega(u)$ is nonempty and bounded for every u of the form $h_\omega - T_\omega x$ for some $x \in X$. Every such $u = h_\omega - T_\omega x$ therefore belongs to int(dom θ_ω) (cf.

[9, Theorem 23.4]). It follows then that $\partial\theta_\omega$ is locally bounded at u (cf. [9, Corollary 24.5.1]). The mapping $x \to h_\omega - T_\omega x$ is continuous, so this implies ζ_ω is locally bounded at x for every $x \in X$, as we needed to prove.

The argument just given shows also that the convex function θ_ω is continuous at $h_\omega - T_\omega x$ for every $x \in X$ (since θ_ω is continuous on int(dom θ_ω) [9, Theorem 10.1]). Therefore F is continuous on X by (1.6) and (2.6). We observed earlier in the proof that G is also continuous on Z. Of course X and Z, being convex polyhedra, are closed sets. Hence if $\varepsilon_\nu \to 0$, so that $F(\bar{x}^\nu) \to \bar{\alpha}$ and $G(\bar{z}^\nu) \to \bar{\alpha}$, any cluster points $\bar{x}^\infty$ of $\{\bar{x}^\nu\}$ and $\bar{z}^\infty$ of $\{\bar{z}^\nu\}$ must satisfy $F(\bar{x}^\infty) = \bar{\alpha} = G(\bar{z}^\infty)$ and be optimal solutions to (P) and (D).

We turn finally to the estimate (4.6). The saddle point condition on $(\bar{x}^\nu, \bar{z}^\nu)$ entails

$$\bar{x}^\nu \in \underset{x \in X}{\operatorname{argmin}}\, L(x, \bar{z}^\nu).$$

Since X is a closed convex set and $L(x, \bar{z}^\nu)$ is a differentiable convex function of x, this condition implies that the vector $-\bar{w}^\nu = -\nabla_x L(\bar{x}^\nu, \bar{z}^\nu)$ belongs to the normal cone to X at $\bar{x}^\nu$ (cf. [9, Theorem 27.4]), which is exactly the assertion of (4.7). We have

$$\begin{aligned} L(x, \bar{z}^\nu) &= L(\bar{x}^\nu, \bar{z}^\nu) + \nabla_x L(\bar{x}^\nu, \bar{z}^\nu) \cdot (x - \bar{x}^\nu) + \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu) \\ &= \bar{\alpha}_\nu + \bar{w}^\nu \cdot (x - \bar{x}^\nu) + \tfrac{1}{2}\|x - \bar{x}^\nu\|_C^2 \text{ for all } x \end{aligned} \tag{4.12}$$

from the quadratic nature of L, and also

$$L(x, \bar{z}^\nu) \leq F(x) \quad \text{for all } x \in X$$

by (2.2). For any optimal solution $\bar{x}$ to (P), then, we have

$$\bar{\alpha}_\nu + \bar{w}^\nu \cdot (\bar{x} - \bar{x}^\nu) + \tfrac{1}{2}\|\bar{x} - \bar{x}^\nu\|_C^2 \leq F(\bar{x}) = \bar{\alpha}.$$

In terms of $\bar{\varepsilon}_\nu = \bar{\alpha} - \bar{\alpha}_\nu$, this can be written as the first inequality in (4.6). The rest of (4.6) then follows from (4.7), inasmuch as $\varepsilon_\nu = \alpha_\nu - \bar{\alpha}_\nu = \bar{\varepsilon}_\nu + \alpha_\nu - \bar{\alpha} \geq \bar{\varepsilon}_\nu$. □

Theorem 3 focuses our attention on finding conditions that guarantee $\varepsilon_\nu \to 0$. Our first result in this direction makes no additional assumptions on the data in the problem and therefore serves as a baseline. It relies on an increasing sequence of polytopes in Step 4, however. The generalized cutting-plane rule in (3.32) is covered as a special case.

Theorem 4. *If $Z^{\nu+1} \supset Z^\nu \cup \{z^\nu\}$ in Step 4 of the algorithm, then $\varepsilon_\nu \to 0$.*

Proof. Let $\bar{\alpha}_\infty = \lim_\nu \bar{\alpha}_\nu$ and $\alpha_\infty = \limsup_\nu \alpha_\nu$. (The first limit exists because $\{\bar{\alpha}_\nu\}$ is nondecreasing in (4.5).) Since $\varepsilon_\nu = \alpha_\nu - \bar{\alpha}_\nu \geq 0$ for all ν, we need only demonstrate that $\alpha_\infty \leq \bar{\alpha}_\infty$. The sequences $\{\bar{x}^\nu\}$, $\{\bar{z}^\nu\}$, and $\{z^\nu\}$, are bounded by Theorem 3, so we can extract convergent subsequences with a common index set $N \subset \{1, 2, \ldots\}$ such that

$$\bar{x}^\nu \xrightarrow[\nu \in N]{} \bar{x}^\infty, \qquad \bar{z}^\nu \xrightarrow[\nu \in N]{} \bar{z}^\infty, \qquad z^\nu \xrightarrow[\nu \in N]{} z^\infty, \qquad \alpha_\nu \xrightarrow[\nu \in N]{} \alpha_\infty.$$

Then since

$$\bar{\alpha}_\nu = L(\bar{x}^\nu, \bar{z}^\nu) \xrightarrow[\nu \in N]{} L(\bar{x}^\infty, \bar{z}^\infty), \qquad \alpha_\nu = L(\bar{x}^\nu, z^\nu) \xrightarrow[\nu \in N]{} L(\bar{x}^\infty, z^\infty),$$

we have $\bar{\alpha}_\infty = L(\bar{x}^\infty, \bar{z}^\infty)$ and $\alpha_\infty = L(\bar{x}^\infty, z^\infty)$. Our task now is to prove that $L(\bar{x}^\infty, z^\infty) \leq L(\bar{x}^\infty, \bar{z}^\infty)$.

From the saddle point condition on $(\bar{x}^\nu, \bar{z}^\nu)$ we have

$$L(\bar{x}^\nu, z) \leq L(\bar{x}^\nu, \bar{z}^\nu) \quad \text{for all } z \in Z^\nu.$$

Let $Z^\infty = \bigcup_{\nu=1}^\infty Z^\nu$. Since $Z^\nu \subset Z^{\nu+1} \subset \cdots$ we know that for any fixed $z \in Z^\infty$ the inequality $L(\bar{x}^\nu, z) \leq L(\bar{x}^\nu, \bar{z}^\nu)$ holds for all ν sufficiently high. Taking the limit as $\nu \to \infty$, $\nu \in N$, we obtain $L(\bar{x}^\infty, z) \leq L(\bar{x}^\infty, \bar{z}^\infty)$. This holds for arbitrary $z \in Z^\infty$, so

$$L(\bar{x}^\infty, z) \leq L(\bar{x}^\infty, \bar{z}^\infty) \quad \text{for all } z \in \operatorname{cl} Z^\infty.$$

But z^∞ is one of the elements of cl Z^∞, since $z_\nu \in Z^{\nu+1}$ for all ν. Therefore $L(\bar{x}^\infty, z^\infty) \leq L(\bar{x}^\infty, \bar{z}^\infty)$ in particular, and the proof is complete. □

Our main result comes next. It assures us that when C is positive definite, we do *not* have to keep increasing the size of the polytope Z^ν in order to have convergence. The number of elements used to generate Z^ν can be kept at whatever level seems adequate in maintaining a robust representation of F and G.

Theorem 5. *Suppose the matrix C in* (P) *is positive definite. Then under the minimal requirement $Z^{\nu+1} \supset \{\bar{z}^\nu, z^\nu\}$ in Step 4 of the algorithm, one has $\varepsilon_\nu \to 0$ and also $\bar{x}^\nu \to \bar{x}$, where $\bar{x}$ is the unique optimal solution to* (P).

If in addition there exists $\rho \geq 0$ such that

$$z_\omega \cdot T_\omega C^{-1} T_\omega^* z_\omega \leq \rho z_\omega \cdot H_\omega z_\omega \quad \textit{for all } z_\omega \in \mathbb{R}^m,\ \omega \in \Omega, \tag{4.13}$$

(*as is true in particular if every H_ω is positive definite*), *then in the estimate* (4.6) *one has*

$$\bar{\varepsilon}_{\nu+1} \leq \tau \bar{\varepsilon}_\nu \quad \textit{for } \nu = 1, 2, \ldots \tag{4.14}$$

where the factor $\tau \in [0, 1)$ is given by

$$\tau = \begin{cases} \rho & \text{if } 0 \leq \rho \leq \frac{1}{2}, \\ 1 - \frac{1}{4}\rho^{-1} & \text{if } \rho \geq \frac{1}{2}. \end{cases} \tag{4.15}$$

Thus

$$\bar{\varepsilon}_{\nu+\mu} \leq \tau^\mu \bar{\varepsilon}_\nu \leq \tau^\mu \varepsilon_\nu \quad \textit{for } \nu = 1, 2, \ldots, \textit{ and } \mu = 1, 2, \ldots. \tag{4.16}$$

Note that Theorem 5 asserts in (4.14) a linear rate of convergence of $\bar{\alpha}_\nu$ to $\bar{\alpha}$ with modulus τ, and the estimate (4.6) effectively translates this into a linear rate of convergence of $\bar{x}^\nu$ to $\bar{x}$ with modulus $\tau^{1/2}$. Indeed, from (4.6) and (4.16) we have

$$\|\bar{x} - \bar{x}^{\nu+\mu}\|_C \leq [2\tau^\mu \varepsilon_\nu]^{1/2} \quad \text{for } \nu = 1, 2, \ldots \text{ and } \mu = 1, 2, \ldots.$$

This is an unusual sort of result, because it applies not just to the tail of the sequence $\{\bar{x}^\nu\}$ but right from the beginning. Moreover the value of ε_ν is known in each iteration, and the value of $\tau \in [0, 1)$ can be estimated in advance.

Theorem 5 makes no assertion about the convergence of $\{\bar{z}^\nu\}$ beyond the one in Theorem 3. Of course if there is a *unique* optimal solution $\bar{z}$ to (D), then by Theorem 3 we have $\bar{z}^\nu \to \bar{z}$ whenever $\varepsilon_\nu \to 0$, as is the case here. In particular (D) has a unique optimal solution if the matrices H_ω are all positive definite.

The proof of Theorem 5 depends on further analysis of the dual objective function G. Essentially what we must provide is a lower estimate of G that ensures that the direction $z^\nu - \bar{z}^\nu$ determined in Step 2 of the algorithm is always a direction of ascent for G.

Proposition 3. *Let*

$$f^\nu(w) = \max_{x \in X} \{(w - \bar{w}^\nu) \cdot (x - \bar{x}^\nu) - \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu)\} \quad \textit{for } w \in \mathbb{R}^n. \tag{4.17}$$

Then f^ν is a finite convex function on $\mathbb{R}^n$ with $0 = f^\nu(0) \leq f^\nu(w)$ for all w, and

$$\begin{aligned} 0 &\leq L(\bar{x}^\nu, z) - G(z) - f^\nu(ET_\omega^*(z_\omega - \bar{z}_\omega^\nu)) \\ &\leq Ef^\nu(T_\omega^*(z_\omega - \bar{z}_\omega^\nu)) \quad \textit{for all } z \in Z. \end{aligned} \tag{4.18}$$

If C is positive definite, then

$$f^\nu(w) \leq \tfrac{1}{2}[(w - \bar{w}^\nu) + s\bar{w}^\nu] \cdot C^{-1}[(w - \bar{w}^\nu) + s\bar{w}^\nu] \quad \textit{for all } s \geq 0, \tag{4.19}$$

so that in particular (for $s = 1$)

$$G(z) \geq L(\bar{x}^\nu, z) - \tfrac{1}{2}E\{(z_\omega - \bar{z}_\omega^\nu) \cdot T_\omega C^{-1} T_\omega^*(z_\omega - \bar{z}_\omega^\nu)\} \quad \textit{for all } z \in Z. \tag{4.20}$$

Proof. First re-express f^ν in terms of the finite concave function φ in (1.4), so as to verify that f^ν is a finite convex function and that 'max' rather than 'sup' is appropriate in (4.17):

$$\begin{aligned} -f^\nu(w) &= \min_{x \in X} \{(\bar{w}^\nu - w) \cdot (x - \bar{x}^\nu) + \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu)\} \\ &= (w - \bar{w}^\nu) \cdot \bar{x}^\nu + \tfrac{1}{2}\bar{x}^\nu \cdot C\bar{x}^\nu + \max_{x \in X} \{[\bar{w}^\nu - C\bar{x}^\nu - w] \cdot x + \tfrac{1}{2}x \cdot Cx\} \\ &= (w - \bar{w}^\nu) \cdot \bar{x}^\nu + \tfrac{1}{2}\bar{x}^\nu \cdot C\bar{x}^\nu + \varphi(\bar{w}^\nu - C\bar{x}^\nu - w). \end{aligned}$$

Clearly $f^\nu(w) \geq 0$ for all w, because $x = \bar{x}^\nu$ is one of the points considered in taking the maximum in (4.17). Furthermore

$$-f^\nu(0) = \min_{x \in X} \{\bar{w}^\nu \cdot (x - \bar{x}^\nu) + \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu)\}.$$

Recalling the expansion (4.12) of $L(x, \bar{z}^\nu)$ around $\bar{x}^\nu$ and the fact that $\bar{x}^\nu$ minimizes $L(x, \bar{z}^\nu)$ over X (since $(\bar{x}^\nu, \bar{z}^\nu)$ is a saddle point of L on $X \times Z^\nu$), we see that $f^\nu(0) = 0$.

To get the equation in (4.18), from which the two inequalities in (4.18) immediately follows (the first because $f^\nu(w) \geq 0$ and the second by Jensen's inequality, because f^ν is convex), we look at the expansion

$$L(x, z) = L(\bar{x}^\nu, z) + \nabla_x L(\bar{x}^\nu, z) \cdot (x - \bar{x}^\nu) + \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu),$$

where

$$\nabla_x L(\bar{x}^\nu, z) = c + C\bar{x}^\nu - ET^*_\omega z_\omega = \bar{w}^\nu - ET^*_\omega(z_\omega - \bar{z}^\nu_\omega).$$

From this we calculate

$$L(\bar{x}^\nu, z) - G(z) = L(\bar{x}^\nu, z) - \min_{x \in X} L(x, z) = \max_{x \in X} \{L(\bar{x}^\nu, z) - L(x, z)\}$$

$$= \max_{x \in X} \{[ET^*_\omega(z_\omega - \bar{z}^\nu_\omega) - \bar{\omega}^\nu] \cdot (x - \bar{x}^\nu)$$

$$- \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu)\}$$

$$= f^\nu(ET^*_\omega(z_\omega - \bar{z}^\nu_\omega)).$$

This establishes (4.18).

Finally we use property (4.7) in Theorem 3 to estimate for arbitrary $s \geq 0$:

$$f^\nu(w) \leq \sup_{x \in X} \{[(w - \bar{w}^\nu) + s\bar{w}^\nu] \cdot (x - \bar{x}^\nu) - \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu)\}$$

$$\leq \sup_{x \in \mathbb{R}^n} \{[(w - \bar{w}^\nu) + s\bar{w}^\nu] \cdot (x - \bar{x}^\nu) - \tfrac{1}{2}(x - \bar{x}^\nu) \cdot C(x - \bar{x}^\nu)\}.$$

When C is positive definite, this last supremum equals the quadratic expression on the right side of (4.19).

Proof of Theorem 5. Since $(\bar{x}^{\nu+1}, \bar{z}^{\nu+1})$ is a saddle point of L relative to $X \times Z^\nu$, we have

$$\bar{\alpha}_{\nu+1} = G(\bar{z}^{\nu+1}) = \max_{z \in Z^{\nu+1}} G(z).$$

But $Z^{\nu+1}$ includes the line segment joining $\bar{z}^\nu$ and z^ν. Therefore

$$\bar{\alpha}_{\nu+1} \geq \max_{0 \leq t \leq 1} G(\bar{z}^\nu + t(z^\nu - \bar{z}^\nu)). \tag{4.21}$$

To see what this implies, we substitute $z = \bar{z}^\nu + t(z^\nu - \bar{z}^\nu)$ into the estimate (4.20) of Proposition 3 and make use of the fact that, for $0 \leq t \leq 1$,

$$L(\bar{x}^\nu, \bar{z}^\nu + t(z^\mu - \bar{z}^\nu)) = L(\bar{x}^\nu, (1-t)\bar{z}^\nu + tz^\nu)$$

$$\geq (1-t)L(\bar{x}^\nu, \bar{z}^\nu) + tL(\bar{x}^\nu, z^\nu) = (1-t)\bar{\alpha}_\nu + t\alpha_\nu - \bar{\alpha}_\nu + t\varepsilon_\nu. \tag{4.22}$$

This yields

$$G(\bar{z}^\nu + t(z^\nu)) \geq \bar{\alpha}_\nu + t\varepsilon_\nu - \tfrac{1}{2}t^2\delta_\nu \quad \text{for } 0 \leq t \leq 1 \tag{4.23}$$

where

$$\delta_\nu := E\{(z_\omega^\nu - \bar{z}_\omega^\nu) \cdot T_\omega C^{-1} T_\omega^*(z_\omega^\nu - \bar{z}_\omega^\nu)\}. \tag{4.24}$$

Combining (4.23) with (4.21), we get

$$\bar{\alpha}_{\nu+1} \geq \bar{\alpha}_\nu + \sigma(\varepsilon_\nu, \delta_\nu), \tag{4.25}$$

where

$$\sigma(\varepsilon, \delta) := \max_{0 \leq t \leq 1} \{t\varepsilon - \tfrac{1}{2}t^2\delta\} = \begin{cases} \varepsilon - \frac{1}{2}\delta & \text{if } 0 \leq \delta \leq \varepsilon, \\ \frac{1}{2}\varepsilon^2\delta^{-1} & \text{if } \delta > \varepsilon. \end{cases} \tag{4.26}$$

Note that σ is a continuous function of $(\varepsilon, \delta) \in \mathbb{R}_+^2$ with $\sigma(\varepsilon, \delta) = 0$ if $\varepsilon = 0$, but $\sigma(\varepsilon, \delta) > 0$ if $\varepsilon > 0$. The sequence $\{\bar{\alpha}_\nu\}$ is nondecreasing and bounded above by $\bar{\alpha}$ (cf. (4.5)), so $\sigma(\varepsilon_\nu, \delta_\nu) \to 0$. The sequence $\{\delta_\nu\}$ is bounded, because the sequences $\{\bar{z}^\nu\}$ and $\{z^\nu\}$ are bounded (Theorem 3). From the cited properties of σ, it follows then that $\varepsilon_\nu \to 0$. This implies $x^\nu \to \bar{x}$ by property (4.6) in Theorem 1.

We can also write (4.25) as

$$\bar{\varepsilon}_{\nu+1} \leq \bar{\varepsilon}_\nu - \sigma(\varepsilon_\nu, \delta_\nu). \tag{4.27}$$

Under the additional assumption in Theorem 5 that (4.13) holds, we have

$$\delta_\nu \leq \rho\beta_\nu, \quad \text{where } \beta_\nu := E\{(z_\omega^\nu - \bar{z}_\omega^\nu) \cdot H_\omega(z_\omega^\nu - \bar{z}_\omega^\nu)\}. \tag{4.28}$$

Consider now the quadratic function

$$q(t) = L(\bar{x}^\nu, \bar{z}^\nu + t(z^\nu - \bar{z}^\nu)) \quad \text{for } 0 \leq t \leq 1.$$

This has $q(0) = L(\bar{x}^\nu, \bar{z}^\nu) = \bar{\alpha}_\nu$, $q(1) = L(\bar{x}^\nu, z^\nu) = \alpha_\nu$, $q''(t) \equiv -\beta_\nu$, so q must be of the form

$$q(t) = (1-t)\bar{\alpha}_\nu + t\alpha_\nu + \tfrac{1}{2}t(1-t)\beta_\nu.$$

Moreover the maximum of $q(t)$ over $0 \leq t \leq 1$ is attained at $t = 1$, since the maximum of $L(\bar{x}^\nu, z)$ over $z \in Z$ is attained at $z = z^\nu$. Therefore

$$(1-t)\bar{\alpha}_\nu + t\alpha_\nu + \tfrac{1}{2}t(1-t)\beta_\nu \leq \alpha_\nu \quad \text{for } 0 \leq t \leq 1,$$

or in other words,

$$t(1-t)\beta_\nu \leq 2(1-t)(\alpha_\nu - \bar{\alpha}_\nu) = 2(1-t)\varepsilon_\nu \quad \text{for } 0 \leq t \leq 1.$$

This implies $\beta_\nu \leq 2\varepsilon_\nu$, and then (4.28) yields

$$\delta_\nu \leq 2\rho\varepsilon_\nu. \tag{4.29}$$

Formula (4.26) now gives us

$$\sigma(\varepsilon_\nu, \delta_\nu) \geq \sigma(\varepsilon_\nu, 2\rho\varepsilon_\nu) = \varepsilon_\nu\sigma(1, 2\rho) \geq \bar{\varepsilon}_\nu\sigma(1, 2\rho).$$

Substituting in (4.27) we get

$$\bar{\varepsilon}_{\nu+1} \leq \bar{\varepsilon}_\nu - \bar{\varepsilon}_\nu\sigma(1, 2\rho) = [1 - \sigma(1, 2\rho)]\bar{\varepsilon}_\nu, \tag{4.30}$$

where

$$\sigma(1, 2\rho) = \begin{cases} 1-\rho & \text{if } 0 \leq \rho \leq \frac{1}{2}. \\ \frac{1}{4}\rho^{-1} & \text{if } \rho \leq \frac{1}{2}. \end{cases}$$

The factor $1-\sigma(1, 2\rho)$ is the number τ defined in (4.15), and (4.30) is thus the desired condition (4.14). □

Remark. Proposition 3 provides additional information that could be used in the direction search and polytope modification steps in the algorithm. Inequality (4.18) asserts that

$$L(\bar{x}^\nu, z) \geq G(z) \geq L(\bar{x}^\nu, z) - Ef^\nu(T_\omega^*(z_\omega - \bar{z}_\omega^\nu))$$
$$\text{for all } z \in Z, \text{ with equality when } z = \bar{z}^\nu. \tag{4.31}$$

The vector z^ν maximizes $L(\bar{x}^\nu, z)$ over all $z \in Z$ and thus provides not only the needed value $L(\bar{x}^\nu, \bar{z}^\nu) = F(\bar{x}^\nu)$ but also a clue as to where we might look to move next in trying to improve on the current value $G(\bar{z}^\nu)$ of G. A further clue can be found by maximizing the right side of (4.31) over Z to get a vector $\bar{z}^\nu$. This is possible because the right side decomposes into separate terms for each ω. Indeed, the components $\bar{z}_\omega^\nu$ of $\hat{z}^\nu$ can be determined by

$$\hat{z}_\omega^\nu \in \underset{z_\omega \in Z_\omega}{\operatorname{argmax}} \{f^\nu(T_\omega^*(z_\omega - \bar{z}_\omega^\nu)) + z_\omega \cdot [h_\omega - T_\omega \bar{x}^\nu] - \tfrac{1}{2} z_\omega \cdot H_\omega z_\omega\}. \tag{4.32}$$

In view of the form of f^ν in (4.17), this amounts to solving a special quadratic programming problem for each $\omega \in \Omega$.

If $\hat{z}^\nu$ is calculated in this way along with z^ν in Step 2, it can also be incorporated in the new polytope $Z^{\nu+1}$ in Step 4 in order to enrich the representation of G.

5. Adding strongly quadratic terms

The theoretical convergence properties of the finite generation algorithm are markedly superior when the quadratic forms that are involved are positive definite. But many problems lack this positive definiteness. Stochastic linear programming problems, for instance, have no quadratic terms at all. Such problems can be handled by a procedure which combines the finite generation algorithm with an augmented Lagrangian technique that introduces the desired property.

The technique in question was developed by Rockafellar [7] in a general context of minimax problems and variational inequalities. As applied to the present situation, it concerns the replacement of the saddle point problem for L on $X \times Z$ by a sequence of saddle point problems for augmented Lagrangians of the form

$$L_\mu(x, z) = L(x, z) + \frac{\eta}{2}(x - \bar{x}_*^\mu) \cdot \bar{C}(x - \bar{x}_*^\mu) - \frac{\eta}{2} E\{(z_\omega - \bar{z}_{*\omega}^\mu) \cdot \bar{H}_\omega(z_\omega - \bar{z}_{*\omega}^\mu)\}$$
$$\text{on } X \times Z \text{ for } \mu = 1, 2, \ldots. \tag{5.1}$$

Here $\bar{C}$ and $\bar{H}_\omega$ are fixed positive definite matrices, η is a penalty parameter value that helps to control the rate of convergence, and $(\bar{x}_*^\mu, \bar{z}_*^\mu)$ is a current 'estimate' for a saddle point of L itself on $X \times Z$, i.e. for an optimal solution pair for problems (P) and (D).

When the augmenting terms in L_μ are expanded and combined with those in L, the expression (5.1) turns into

$$L_\nu(x, z) = c_*^\mu \cdot x + \tfrac{1}{2}x \cdot C_* x + E\{z_\omega \cdot [h_{*\omega} - T_\omega x] - \tfrac{1}{2}z_\omega \cdot H_{*\omega} z_\omega\} + \text{const.} \tag{5.2}$$

where

$$C_* = C + \eta\bar{C}, \qquad H_{*\omega} = H_\omega + \eta\bar{H}_\omega, \tag{5.3}$$

$$c_*^\mu = c - \eta\bar{C}\bar{x}_*^\mu, \qquad h_{*\omega}^\mu = h_\omega - \eta\bar{H}_\omega\bar{z}_{*\omega}^\mu. \tag{5.4}$$

Note that the vectors c_*^μ and $h_{*\omega}^\mu$ giving the linear terms in L_μ depend on the μth solution estimates, but the matrices C_* and $H_{*\omega}$ giving the quadratic terms remain fixed as long as the value of η is not varied. Since $\eta > 0$, these matrices are positive definite. Therefore the saddle point problem for L_μ on $X \times Z$ can be solved by the finite generation algorithm with an essentially linear rate of convergence (cf. Theorem 5).

We make use of this as follows.

Master Algorithm

Step 0 (Initialization). Fix the matrices $\bar{C}$, $\bar{H}_\omega$, and the parameter value $\eta > 0$. Choose initial points $\bar{x}_*^1 \in X$ and $\bar{z}_*^1 \in Z$. Set $\mu = 1$.

Step 1 (Finite Generation Method). Use the finite generation algorithm to determine an approximate saddle point $(\bar{x}_*, \bar{z}_*)$ of L_μ on $X \times Z$ (according to a stopping criterion given below).

Step 2 (Update). Set $(\bar{x}_*^{\mu+1}, \bar{z}_*^{\mu+1}) = (\bar{x}_*, \bar{z}_*)$. Replace μ by $\mu + 1$ and return to Step 1 (with the same value of η).

The finite generation method in Step 1 generates for the function L_μ a sequence of pairs $(\bar{x}^\nu, \bar{z}^\nu)$ and test values ε_ν. To get an approximate saddle point we take

$$(\bar{x}_*, \bar{z}_*) = (\bar{x}^\nu, \bar{z}^\nu) \quad \text{when } \varepsilon_\nu \leq \bar{\varepsilon}_*^\mu(\bar{x}^\nu, \bar{z}^\nu), \tag{5.5}$$

where the function $\bar{\varepsilon}_*^\mu$ in the stopping criterion is defined as follows. In terms of the norms

$$\|x\|_* = [x \cdot \bar{C}x]^{1/2} \text{ for } x \in \mathbb{R}^n, \qquad \|z\|_* = [E\{z_\omega \cdot \bar{H}_\omega z_\omega\}]^{1/2} \text{ for } z \in (\mathbb{R}^m)^\Omega,$$

$$\|(x, z)\|_* = [\|x\|_*^2 + \|z\|_*^2]^{1/2} \tag{5.6}$$

we set

$$\bar{\varepsilon}_*^\mu(x, z) = \theta_\mu^2 \min\{1, (\eta/2)\|(x, z) - (\bar{x}^\mu, \bar{z}^\mu)\|_*^2\} \quad \text{with } \theta_\mu > 0, \sum_{\mu=1}^{\infty} \theta_\mu < \infty. \tag{5.7}$$

Obviously $\bar{\varepsilon}_*^\mu(x, z) > 0$ unless $(x, z) = (\bar{x}^\mu, \bar{z}^\mu)$. The sequence $\{(\bar{x}^\nu, \bar{z}^\nu)\}$ converges to the unique saddle point of L_μ on $X \times Z$, so except in the lucky, degenerate case where $(\bar{x}_*^\mu, \bar{z}_*^\mu)$ is already that saddle point, the values $\varepsilon_*^\mu(\bar{x}^\nu, \bar{z}^\nu)$ will be bounded away from zero, and the stopping criterion in (5.5) will eventually be satisfied. (In the degenerate case, $(\bar{x}_*^\mu, \bar{z}_*^\mu)$ must in fact be a saddle point of L itself and there is no need to leave Step 1: the sequence $\{(\bar{x}^\nu, \bar{z}^\nu)\}$ converges to this saddle point at a linear rate.)

Theorem 6. *The sequences $\{\bar{x}_*^\mu\}$ and $\{\bar{z}_*^\mu\}$ generated by the master algorithm converge to particular optimal solutions $\bar{x}$ and $\bar{z}$ to problems* (P) *and* (D), *respectively. If $\bar{x}$ and $\bar{z}$ are the unique optimal solution to* (P) *and* (D), *then there is a number $\beta(\eta) \in [0, 1)$ such that $(\bar{x}_*^\mu, \bar{z}_*^\mu)$ converges to $(\bar{x}, \bar{z})$ at a linear rate with modulus $\beta(\eta)$. Moreover $\beta(\eta) \to 0$ as $\eta \to 0$.*

Proof. We shall deduce this from [7, Theorems 1 and 2], which are general results applicable to the calculation of a saddle point of a convex-concave function on a product of Hilbert spaces. The Hilbert spaces in this case are $\mathbb{R}^n$ and $(\mathbb{R}^m)^\Omega$ under the norms in (5.5). The convex-concave function in question is

$$\bar{L}(x, z) = \begin{cases} L(x, z) & \text{if } x \in X \text{ and } z \in Z, \\ -\infty & \text{if } x \in X \text{ but } z \notin Z, \\ \infty & \text{if } x \notin X. \end{cases}$$

The saddle points of $\bar{L}$ on $\mathbb{R}^n \times (\mathbb{R}^m)^\Omega$ are the same as those of L on $X \times Z$. The problem of finding a saddle point of $\bar{L}_\mu$ on $X \times Z$ reduces to the one for

$$\bar{L}_\mu(x, z) = \bar{L}(x, z) + \frac{\eta}{2}\|x - \bar{x}_*^\mu\|_*^2 - \frac{\eta}{2}\|z - \bar{z}_*^\mu\|_*^2$$

on $\mathbb{R}^n \times (\mathbb{R}^m)^\Omega$.

Denote by $P(\bar{x}^\mu, \bar{z}^\mu)$ the unique saddle point of $\bar{L}_\mu$ on $R^n \times (R^m)^\Omega$, which is also the unique saddle point of L_μ on $X \times Z$. The mapping P is the 'proximal mapping' associated with the maximal monotone multifunction T that corresponds to $\eta^{-1}\bar{L}$ in the sense of [7, Section 1 and Section 5]. In consequence of [7, Theorem 1], the sequence $\{(\bar{x}_*^\mu, \bar{z}_*^\mu)\}$ generated by the master algorithm will converge to a particular saddle point $(\bar{x}, \bar{z})$ of $\bar{L}$ on $\mathbb{R}^n \times (\mathbb{R}^m)^\omega$ (the same as a saddle point of L on $X \times Z$) if

$$\|(\bar{x}^{\mu+1}, \bar{z}^{\mu+1}) - P(\bar{x}^\mu, \bar{z}^\mu)\|_* \le \gamma_\mu \quad \text{with } \gamma_\mu > 0, \sum_{\mu=1}^{\infty} \gamma_\mu < \infty. \tag{5.8}$$

Under the more stringent condition

$$\|(\bar{x}^{\mu+1}, \bar{z}^{\mu+1}) - P(\bar{x}^{\mu}, \bar{z}^{\mu})\|_* \leq \theta_\mu \|(\bar{x}^{\mu+1}, \bar{z}^{\mu+1}) - (\bar{x}^{\mu}, \bar{z}^{\mu})\|_*$$

$$\text{with } \theta_\mu > 0, \quad \sum_{\mu=1}^{\infty} \theta_\mu < \infty, \tag{5.9}$$

we know from [7, Theorem 2] that if $(\bar{x}, \bar{z})$ is the unique saddle point of $\bar{L}$ and a certain Lipschitz property holds in terms of a constant $\sigma \geq 0$, we will have

$$\limsup_{\mu \to \infty} \|(\bar{x}_*^{\mu+1}, \bar{z}_*^{\mu+1}) - (\bar{x}, \bar{z})\|_* / \|(\bar{x}_*^{\mu}) - (\bar{x}, \bar{z})\|_* = \beta(\eta),$$

where

$$\beta(\eta) = \sigma\eta/(1+\sigma^2\eta^2)^{1/2} < 1. \tag{5.10}$$

The Lipschitz property in question is the following: for $\sigma \geq 0$ and some $\varepsilon > 0$, all the saddle points $(\tilde{x}, \tilde{z})$ of any perturbed Lagrangian of the form

$$\tilde{L}(x, y) = L(x, y) + \tilde{c} \cdot x + E\{\tilde{h}_\omega \cdot z_\omega\} \quad \text{on } X \times Z,$$

with $\tilde{c} \in \mathbb{R}^n$ and $\tilde{h} = (\ldots, \tilde{h}_\omega, \ldots) \in (\mathbb{R}^m)^\Omega$, will satisfy

$$\|(\tilde{x}, \tilde{z}) - (\bar{x}, \bar{z})\|_* \leq \sigma \|(\tilde{c}, \tilde{h})\|_{**} \quad \text{when } \|(\tilde{c}, \tilde{h})\|_{**} \leq \varepsilon.$$

(Here $\|\cdot\|_{**}$ is the norm dual to $\|\cdot\|_*$.) This needed property does hold, because of the quadratic nature of our problem. The optimality conditions that characterize $(\tilde{x}, \tilde{z})$ as a saddle point of $\tilde{L}$ on $X \times Z$ are all linear; the multifunction that associates with each $(\tilde{c}, \tilde{h})$ this corresponding set of saddle points is in fact a *polyhedral* multifunction in the sense of Robinson, i.e. its graph is the union of finitely many convex polyhedra. Any such multifunction has the Lipschitz property in question; see Robinson [6].

We shall show now that our stopping criterion (5.5), (5.7), does imply (5.8) and (5.9) with $\gamma_\nu = \theta_\mu[2/\eta]^{1/2}$. Consider the primal and dual objective functions associated with L_μ namely

$$F_\mu(x) = \max_{z \in Z} L_\mu(x, z), \qquad G_\mu(z) = \min_{x \in X} L_\mu(x, z). \tag{5.11}$$

The approximate saddle point $(\bar{x}_*^{\mu+1}, \bar{z}_*^{\mu+1}) = (\bar{x}^\nu, \bar{z}^\nu)$ satisfies

$$F_\mu(\bar{x}_*^{\mu+1}) - G_\mu(\bar{z}_*^{\mu+1}) \leq \varepsilon_\nu \leq \varepsilon_*^\mu(\bar{x}_*^{\mu+1}, \bar{z}_*^{\mu+1}) \tag{5.12}$$

by Theorem 3 (as applied to L_μ) and (5.5). The true saddle point $(\hat{x}_*^\mu, \hat{z}_*^\mu) = P(\bar{x}_*^\mu, \bar{z}_*^\mu)$ satisfies

$$\min_{x \in X} L_\mu(x, \bar{z}_*^\mu) = L_\mu(\hat{x}_*^\mu, \hat{z}_*^\mu) = \max_{z \in Z} L_\mu(\hat{x}_*^\mu, z),$$

and because L_μ is strongly quadratic by virtue of the terms added to form it from L, this must actually hold in the strong sense that

$$L_\mu(x, \hat{z}_*^\mu) \geq L_\mu(\hat{x}_*^\mu, \hat{z}_*^\mu) + (\eta/2)\|x - \hat{x}_*^\mu\|_*^2 \quad \text{for all } z \in Z,$$

$$L_\mu(\hat{x}_*^\mu, z) \leq L_\mu(\hat{x}_*^\mu, \hat{z}_*^\mu) - (\eta/2)\|z - \hat{z}_*^\mu\|_*^2 \quad \text{for all } z \in Z.$$

Taking $x=\bar{x}_*^{\mu+1}$ and $z=\bar{x}_*^{\mu+1}$ in these inequalities and observing from definition (5.11) that

$$F_\mu(\bar{x}_*^\mu) \geq L_\mu(\hat{x}_*^\mu, \hat{z}_*^\mu) \geq G_\mu(\bar{z}_*^\mu),$$

we obtain

$$\begin{aligned} F_\mu(\bar{x}_*^{\mu+1}) - G_\mu(\bar{z}_*^{\mu+1}) &\geq (\eta/2)\|\bar{x}_*^{\mu+1} - \hat{x}_*^\mu\|_*^2 + (\eta/2)\|\bar{z}_*^{\mu+1} - \hat{z}_*^\mu\|_*^2 \\ &= (\eta/2)\|(\bar{x}_*^{\mu+1}, \bar{z}_*^{\mu+1}) - P(\bar{x}_*^\mu, \bar{z}_*^\mu)\|_*^2. \end{aligned}$$

This, combined with (5.12) and (5.7), yields

$$\begin{aligned} &(\eta/2)\|(\bar{x}_*^{\mu+1}, \bar{z}_*^{\mu+1}) - P(\bar{x}_*^\mu, \bar{z}_*^\mu)\|_*^2 \\ &\quad \leq \theta_\mu^2 \min\{1, (\eta/2)\|(\bar{x}_*^{\mu+1}, \bar{z}_*^{\mu+1}) - (\bar{x}_*^\mu, \bar{z}_*^{\mu+1})\|_*^2\}. \end{aligned}$$

Then (5.8) and (5.9) hold as claimed, with $\gamma_\mu = \theta_\mu[2/\gamma]^{1/2}$. □

We conclude by connecting the choice of the matrices $\bar{C}$ and $\bar{H}_\omega$ in (5.1) with the convergence rate of the finite generation algorithm in Step 1 of the master algorithm.

Proposition 4. *Suppose $\bar{C}$ and $\bar{H}_\omega$ are selected so that for a certain $\bar{\rho}>0$,*

$$z_\omega \cdot [T_\omega \bar{C}^1 T_\omega^*] z_\omega \leq \bar{\rho}[z_\omega \cdot \bar{H}_\omega z_\omega] \quad \text{for all } z_\omega \in \mathbb{R}^m. \tag{5.13}$$

Then the matrices C_* and $H_{*\omega}$ in (5.3) have

$$z_\omega \cdot [T_\omega C_*^{-1} T_\omega^*] z_\omega \leq (\bar{\rho}/\eta^2)[z_\omega \cdot H_{*\omega} z_\omega] \quad \text{for all } z_\omega \in \mathbb{R}^m, \tag{5.14}$$

so that when the finite generation algorithm is applied to finding a saddle point of L_μ, the convergence results in Theorem 3 will be valid for $\rho = \bar{\rho}/\eta^2$.

Proof. Let us simplify notation by writing $A \leq B$ for positive definite symmetric matrices A and B to mean that $B-A$ is positive semidefinite. Since A and B can be diagonalized simultaneously, this relation can be interpreted also as a coordinate-wise inequality on the corresponding vectors of eigenvalues. In this notation, our assumption (5.13) is that $T_\omega \bar{C}^{-1} T_\omega^* \leq \bar{\rho}\bar{H}_\omega$. Since $C_* = C + \eta\bar{C}$ we know $C_* \geq \eta\bar{C}$ and therefore $C_*^{-1} \leq \eta^{-1}\bar{C}^{-1}$. But also, from $H_{*\omega} = H_\omega + \eta\bar{H}_\omega$ we have $\eta\bar{H}_\omega \leq H_{*\omega}$, or in other words $\bar{H}_\omega \leq \eta^{-1}H_{*\omega}$. It follows that

$$T_\omega C_*^{-1} T_\omega^* \leq \eta^{-1} T_\omega \bar{C}^1 T_\omega^* \leq \eta^{-1}\bar{\rho}\bar{H}_\omega \leq \eta^{-2}\bar{\rho}H_{*\omega}$$

as claimed in (5.14). □

This result reveals a trade-off between the rates of linear convergence that can be achieved in the finite generation algortihm and in the master algorithm. The modulus $\beta(\eta)$ for the latter can be improved by making η smaller. But one cannot at the same time make ρ smaller, as would be desirable for the finite generation algorithm in the light of Theorem 5.

References

[1] R.W. Cottle, "Symmetric dual quadratic programs", *Quarterly of Applied Mathematics* 21 (1983) 237–243.

[2] W.S. Dorn, "Duality in quadratic programming", *Quarterly of Applied Mathematics* 18 (1960) 155–162.

[3] M. Frank and P. Wolfe, "An algorithm for quadratic programming", *Naval Research Logistics Quarterly* 3 (1956) 95–110.

[4] A. King, "An implementation of the Lagrangian finite-generation method", in: Y. Ermoliev, R. Wets, eds., *Numerical techniques for stochastic optimaization problems.*

[5] H.W. Kuhn and A.W. Tucker, "Nonlinear programming", in: J. Neyman, ed., *Second Berkeley symposium on mathematical statistics* (University of California Press, Berkeley, 1951) pp. 481–492.

[6] S.M. Robinson, "Some continuity properties of polyhedral multifunctions", *Mathematical Programming Study* 14 (1981) 206–214.

[7] R.T. Rockafellar, "Monotone operators and the proximal point algorithm", *SIAM Journal on Control and Optimization* 14 (1976) 877–898.

[8] R.T. Rockafellar, *Conjugate duality and optimization,* No. 16 in Conference Board of the Mathematical Sciences Series, (SIAM Publications, Philadelphia, 1974).

[9] R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, NJ, 1970).

[10] R.T. Rockafellar and R.J.-B. Wets, "Stochastic convex programming: Relatively complete recourse and induced feasibility", *SIAM Journal on Control and Control and Optimization* 14 (1976) 574–589.

[11] R.T. Rockafellar and R.J.-B. Wets, "A dual solution procedure for quadratic stochastic programs with simple recourse", in: V. Pereyra and A. Reinoza, eds., *Numerical methods,* Lecture notes in mathematics 1005 (Springer-Verlag, Berlin, 1938) pp. 252–265.

[12] R.J.-B. Wets, "Stochastic programming: Solution techniques and approximation schemes", in: A. Bachem et al., eds., *Mathematical programming: The state-of-the-art* (Springer-Verlag, Berlin, 1983) pp. 566–603.