LINEAR-QUADRATIC PROGRAMMING PROBLEMS
WITH STOCHASTIC PENALTIES:
THE FINITE GENERATION ALGORITHM

R.T. Rockafellar[1] *and R.J.-B. Wets*[1]

Much of the work on computational methods for solving stochastic programming problems has been focused on the linear case, and with considerable justification. Linear programming techniques for large-scale deterministic problems are highly developed and offer hope for the even larger problems one obtains in certain formulations of stochastic problems. Quadratic programming techniques have not seemed ripe for such a venture, although the ultimate importance of quadratic stochastic programming has been clear enough.

There is another kind of approach, however, in which quadratic stochastic programming problems are no harder to solve than linear ones, and in some respects easier. In this approach, for which the theoretical groundwork has been laid in Rockafellar and Wets [1], the presence of quadratic terms is welcomed because of their stabilizing effect, and such terms are even introduced in iterative fashion. The underlying stochastic problem, whether linear or quadratic, is replaced by a sequence of deterministic quadratic programming problems whose relatively small dimension can be held in check. Among the novel features of the method is its ability to handle more kinds of random coefficients, for instance a random technology matrix.

In this paper we present a particular case of the problem and method in [1] which is especially easy to work with and capable nevertheless of covering many applications. This case falls in the category of stochastic programming with simple recourse. It was described briefly by us in [2], but with the theory in [1] now available, we are able to derive precise results about convergence and the nature of the stopping criterion that can be used. This is also the one case that has been implemented so far and for which numerical experience has gained. For a separate report on the implementation, see King [3].

For the purpose at hand, where duality plays a major role and the constructive use of quadratic terms must be facilitated, the format for stating the problem is crucial. The following deterministic model in linear-quadratic programming serves as the starting point:

$(\mathrm{P_{det}})$
$$\text{maximize } f(x) = \sum_{j=1}^{n}\left[c_j x_j - \tfrac{1}{2} r_j x_j^2\right] - \sum_{k=1}^{\ell} \rho(v_k; p_k, q_k)$$

subject to $0 \le x_j \le s_j$ for $j = 1, \ldots, n$,

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \text{ for } i = 1, \ldots m,$$

$$v_k = \sum_{j=1}^{n} t_{kj} x_j - h_k \text{ for } k = 1, \ldots, \ell,$$

where ρ is a *penalty* function depending on two parameters p_k and q_k and having the form shown in Figure 1, namely

$$\rho(v_k; p_k, q_k) = \begin{cases} 0 & \text{for } v_k \le 0, \\ \tfrac{1}{2} v_k^2 p_k & \text{for } 0 \le v_k \le p_k q_k, \\ q_k v_k - \tfrac{1}{2} p_k q_k^2 & \text{for } v \ge p_k q_k. \end{cases} \qquad (0.1)$$

This is convex in v_k, so the object function f in $(\mathrm{P_{det}})$ is concave; *it is assumed that p_k, q_k, r_j and s_j are nonnegative.* For $p_k = 0$, one takes

$$\rho(v_k; p_k, q_k) = \begin{cases} 0 & \text{for } v_k \le 0, \\ q_k v_k & \text{for } v_k \ge 0. \end{cases} \qquad (0.2)$$
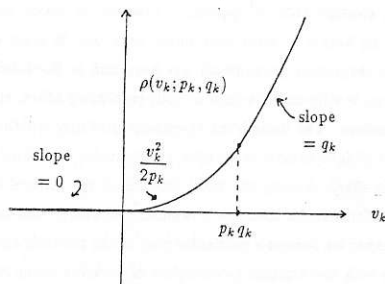

FIGURE 1

The penalty terms in $(\mathrm{P_{det}})$ represent a weakened incorporation of constraints

$$\sum_{j=1}^{n} t_{kj} x_j \le h_k \text{ for } k = 1, \ldots, \ell \qquad (0.3)$$

into the problem. They vanish as long as these constraints are satisfied, but charge a positive cost when they are violated. The cost grows linearly in the special case of (0.2), but otherwise it first passes smoothly through a quadratic phase.

The stochastic programming problem $(\mathrm{P_{sto}})$ that we want to consider is obtained by allowing t_{kj}, h_k, p_k and q_k all to be random variables and replacing each penalty term by its expectation. (In any one application, of course, only a few of these variables might actually be random.) The interpretation is that the x_j's are decision variables whose values must be fixed here and now. The constraints $0 \le x_j \le s_j$ and

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \text{ for } i = 1, \ldots, m \qquad (0.4)$$

are known at the time of this decision, but about the random variables in question there is only statistical information (their distributions). The constraints (0.3) therefore cannot be enforced in the selection of the x_j's without severe consequence. Instead of trying to guard against all possible violations by being extremely conservative, we imagine there is a way of coping with violations of the constraints (0.3), if they should occur. Some recourse action is considered to be possible after the values of the random variables have been realized, and this recourse has an associated cost which depends on the extent of violations. This cost is represented by the penalty terms $\rho(v_k; p_k, q_k)$, and its expectation is subtracted from the here-and-now expression in the x_j's that is being maximized.

Besides the direct applications of this model, we see it as potentially valuable in problems that until now have been formulated deterministically, but in which some of the data may be rather uncertain. By putting such problems in the form of $(\mathrm{P_{sto}})$ it should be possible, even with every crude guesswork about penalty costs and probabilities, to gain some appreciation of how the choice of the x_j's should be modified to hedge against the uncertainties. Certainly this ought to be better than merely assigning specific values to the fuzzy data.

We mention again that although our basic problem is nominally quadratic (a formulation that sidesteps the "piecewise" nature of the penalty terms will be recorded later, in §3), we are also very much concerned with the linear case where $r_j = 0$ and $p_k = 0$. Our plan is first to display a method whose characteristics are most attractive in the strictly quadratic case where $r_j > 0$ and $p_k > 0$, and then apply it to problems lacking in strict quadraticity by means of the proximal point technique [4], [1].

1. OPTIMALITY CONDITIONS AND DUALITY

The approach we are taking depends very much on duality. A subproblem of a certain dual problem will explicitly be solved at every iteration. The Lagrange multipliers in this process will generate the optimizing sequence for the primal problem.

For the deterministic problem $(\mathrm{P_{det}})$, the appropriate dual would be

(D_{det}) minimize $g(y,z) = \sum_{i=1}^{m} b_i y_i + \sum_{k=1}^{\ell} [h_k z_k + \frac{1}{2} p_k z_k^2]$

$$+ \sum_{j=1}^{n} \rho(w_j; r_j, s_j)$$

subject to $0 \le y_i$ for $i = 1, \ldots, m,$

$$0 \le z_k \le q_k \text{ for } k = 1, \ldots, \ell,$$

$$w_j = c_j - \sum_{i=1}^{m} y_i a_{ij} - \sum_{k=1}^{\ell} z_k \ell_{kj} \text{ for } j = 1, \ldots, n.$$

Here ρ is the same function as before (cf. Fig. 1), except that the symbols for the variables have been switched:

$$\rho(w_j; r_j, s_j) = \begin{cases} 0 & \text{for } w_j \le 0 \\ \frac{1}{2} w_j^2 / r_j & \text{for } 0 \le w_j \le r_j s_j \\ s_j w_j - \frac{1}{2} r_j s_j^2 & \text{for } w_j \ge r_j w_j. \end{cases} \quad (1.1)$$

The terms $\rho(w_j; r_j, s_j)$ in (D_{det}) are to be viewed as penalty representation replacements for constraints

$$\sum_{i=1}^{m} y_i a_{ij} + \sum_{k=1}^{\ell} z_k \ell_{kj} \ge c_j \text{ for } j = 1, \ldots, n. \quad (1.2)$$

This form of duality is a special case of the scheme used in monotropic programming [5]. It results from the conjugacy between the convex functions

$$\varphi_k(v_k) = \rho(v_k; p_k, q_k),$$
$$\psi_k(z_k) = \begin{cases} \frac{1}{2} p_k v_k^2 & \text{if } 0 \le z_k \le q_k, \\ \infty & \text{otherwise.} \end{cases} \quad (1.3)$$

One can show that as long as the constraints (P_{det}) are consistent, one has

$$\max (P_{det}) = \min (D_{det}).$$

In the stochastic case we are directly concerned with in this paper, the appropriately modified primal and dual problems are

(P_{sto}) maximize $f(x) = \sum_{j=1}^{n} [c_j x_j - \frac{1}{2} r_j x_j^2] - E\{\sum_{k=1}^{\ell} \rho(v_k; p_k, q_k)\}$

subject to $0 \le x_j \le s_j$ for $j = 1, \ldots, n,$

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \text{ for } i = 1, \ldots, m,$$

$$v_k = \sum_{j=1}^{n} \ell_{kj} x_j - h_k \text{ for } k = 1, \ldots, \ell,$$

(D_{sto}) minimize $g(y,z) = \sum_{i=1}^{m} b_i y_i + E\{\sum_{k=1}^{\ell} [h_k z_k + \frac{1}{2} p_k z_k^2]\}$

$$+ \sum_{j=1}^{n} \rho(w_j; r_j, s_j)$$

subject to $0 \le y_i$ for $i = 1, \ldots, m,$

$$0 \le z_k \le q_k \text{ for } k = 1, \ldots, \ell,$$

$$w_j = c_j - \sum_{i=1}^{m} y_i a_{ij} - E\{\sum_{k=1}^{\ell} z_k \ell_{kj}\}.$$

The random variables in these problems have been indicated by $\sim$; the symbol E denotes mathematical expectation.

In order to avoid minor technical complications that have no real importance in our present task of setting up a computational framework for (P_{sto}), we shall rely henceforth on two assumptions.

(A1) There is at least one vector x satisfying

$$0 \le x_j \le s_j \text{ for } j = 1, \ldots, n, \text{ and } \sum_{j=1}^{n} a_{ij} x_j \le b_i \text{ for } i = 1, \ldots, m.$$

(A2) The given random variables ℓ_{pj}, h_k, p_k, q_k take on only finitely many values.

Only (A2) needs comment. We are assuming that whatever the "true" distribution of these variables might be, we are treating them here in terms of finitely many values to which probability weights have been assigned. Such a discrete distribution might be obtained by approximating a continuous distribution, or by sampling a continuous distribution, or empirically. For now, that need not matter; the question of the source of the discrete distribution and how it might be "improved" is quite separate. The important thing is that we impose no further conditions on the random variables. Aside from (A2), their distribution can be completely arbitrary. In particular a joint distribution is allowed; the variables do not have to be independent.

THEOREM 1. *Under assumptions* (A1) *and* (A2), *problems* (P_{sto}) *and* (D_{sto}) *both have optimal solutions, and*

$$\max (P_{sto}) = \min (D_{sto}).$$

Moreover in the strictly quadratic case where $r_j > 0$ *and* $p_k > 0$, *the following conditions are necessary and sufficient in order that* $\bar{x}$ *be optimal for* (P_{sto}) *and* $(\bar{y}, \bar{z})$ *optimal for* (D_{sto}):

$$\sum_{j=1}^{n} a_{ij} \bar{x}_j - b_i \le 0, \ \bar{y}_i \ge 0, \ and \ (\sum_{j=1}^{n} a_{ij} \bar{x}_j - b_i) \bar{y}_i = 0. \quad (1.4)$$

$$\overline{x}_j = \rho'(\overline{w}_j; r_j, s_j) \quad for \quad \overline{w}_j = c_j - \sum_{i=1}^{m} \overline{y}_i a_{ij} - E\Big\{ \sum_{k=1}^{\ell} \overline{z}_k \underline{t}_{kj} \Big\}, \tag{1.5}$$

$$\overline{z}_k = \rho'(\overline{v}_k; p_k, q_k) \quad for \quad \overline{v}_k = \sum_{j=1}^{n} \underline{t}_{kj} \overline{x}_j - \underline{h}_k. \tag{1.6}$$

In these relations the derivatives ρ' refer to the first argument indicated, not the parameter arguments. Thus

$$\rho'(w_j; r_j, s_j) = \begin{cases} 0 & \text{if } w_j \leq 0 \\ w_j/r_j & \text{if } 0 \leq w_j \leq r_j s_j \\ s_j & \text{if } w_j \geq r_j s_j \end{cases} \tag{1.7}$$

and likewise, with just a change of notation,

$$\rho'(v_k; p_k, q_k) = \begin{cases} 0 & \text{if } v_k \leq 0 \\ v_k/p_k & \text{if } 0 \leq v_k \leq p_k q_k \\ q_k & \text{if } v_k \geq p_k q_k. \end{cases}$$

It is clear then that (1.5) entails $0 \leq \overline{x}_j \leq s_j$, and (1.6) entails $0 \leq \underline{z}_j \leq \underline{q}_j$. This is why these basic requirements do not appear explicitly in the theorem along with the feasibility and complementary slackness conditions (1.4).

Formula (1.5) serves as a means of obtaining the optimal solution to (P_{sto}) from the optimal solution to (D_{sto}), or an approximately optimal solution to (P_{sto}) from an approximately optimal one for (D_{sto}), the mapping being continuous. Formula (1.6), on the other hand, says that the component $\overline{z}$ of an optimal solution to (D_{sto}) is a random variable expressible in terms of the known random variables $\underline{t}_{kj}, \underline{h}_k, \underline{p}_k, \underline{q}_k$, and the (nonrandom) optimal solution $\overline{x}$ to (P_{sto}). More generally, by means of this formula as applied to various nonoptimal vectors x that arise in the solution process, it is possible economically to represent (and store in a computer) some of the elements $\underline{z}$ that will be needed in the solution process.

PROOF OF THEOREM 1. The duality will be obtained from a minimax representation in terms of the sets

$$X = \{ x = (x_1, \ldots, x_n) \mid 0 \leq x_j \leq s_j \}, \tag{1.9}$$

$$Y = \{ y = (y_1, \ldots, y_m) \mid 0 \leq y_j \}, \tag{1.10}$$

$$Z = \{ \underline{z} = (\underline{z}_1, \ldots, \underline{z}_\ell) \mid 0 \leq \underline{z}_k \leq \underline{q}_k \}, \tag{1.11}$$

and the function L on $X \times Y \times Z$ defined by

$$L(x, y, \underline{z}) = \sum_{j=1}^{n} [c_j x_j - \tfrac{1}{2} r_j x_j^2] + \sum_{i=1}^{m} y_i [b_i - \sum_{j=1}^{n} a_{ij} x_j] + \sum_{k=1}^{\ell} E\Big\{ \underline{z}_k [\underline{h}_k - \sum_{j=1}^{n} \underline{t}_{kj} x_j] + \tfrac{1}{2} \underline{p}_k \underline{z}_k^2 \Big\}. \tag{1.12}$$

Here because of assumption (A2) we could think of each of the random variables as functions on a single finite probability space Ω, or equivalently as vectors indexed by $\omega \in \Omega$. Then in (1.1) we could write $0 \leq z_{\omega k} \leq q_{\omega k}$ for all ω and k, while in (1.12) we could write

$$E\Big\{ \underline{z}_k [\underline{h}_k - \sum_{j=1}^{n} \underline{t}_{kj} x_j] + \tfrac{1}{2} \underline{p}_k \underline{z}_k^2 \Big\} = \sum_{\omega \in \Omega} (z_{\omega k} [\pi_\omega h_{\omega k} - \sum_{j=1}^{n} \pi_\omega t_{\omega k j} x_j] + \tfrac{1}{2} \pi_\omega p_{\omega k} z_{\omega k}^2),$$

where $\pi_\omega > 0$ is the probability weight assigned to the element ω of Ω. This makes it plain that Z, like X and Y, is simply a finite-dimensional convex polyhedron, although the dimension may be very large, and L is a quadratic function which is concave in x and convex in $(y, \underline{z})$.

It is easily verified that

$$\inf_{(y, \underline{z}) \in Y \times Z} L(x, y, \underline{z}) = \begin{cases} f(x) & \text{if } x \text{ is feasible in } (P_{sto}), \\ -\infty & \text{otherwise}, \end{cases} \tag{1.14}$$

$$\sup_{x \in X} L(x, y, \underline{z}) = \begin{cases} g(y, \underline{z}) & \text{if } (y, \underline{z}) \text{ is feasible in })(D_{sto}), \\ \infty & \text{otherwise}, \end{cases} \tag{1.15}$$

where $f(x)$ and $g(y, \underline{z})$ are the objective functions specified for (P_{sto}) and (D_{sto}). Thus (P_{sto}) and (D_{sto}) are the primal and dual problems associated with the minimax problem for L on $X \times (Y \times Z)$. Because L is quadratic concave-convex, and the sets X and $Y \times Z$ are convex polyhedra, we may conclude from generalized quadratic programming theory (see [1, Theorem 1]) that if the optimal value in either problem is finite, or if both problems have feasible solutions, then both problems have optimal solutions and max(P_{sto}) = min (D_{sto}). This is indeed the case here. because (D_{sto}) trivially has feasible solutions, and our assumption (A1) guarantees that (P_{sto}) has feasible solutions.

The optimality conditions (1.4), (1.5), (1.6), are just a restatement of the requirement that $(\overline{x}, \overline{y}, \overline{z})$ be a saddlepoint of L on $X \times (Y \times Z)$. For instance, the part of the saddlepoint property that corresponds to maximization in $\underline{z}$ decomposes into

$$\overline{z}_{\omega k} \in \operatorname*{argmin}_{0 \leq z_{\omega k} \leq q_{\omega k}} \Big\{ z_{\omega k} [h_{\omega k} - \sum_{j=1}^{n} t_{\omega k j} \overline{x}_j] + \tfrac{1}{2} p_{\omega k} z_{\omega k}^2 \Big\}.$$

In terms of the conjugate convex functions in (1.3) and the notation

$$\overline{v}_{\omega k} = \sum_{j=1}^{n} t_{\omega k j} \overline{x}_j - h_{\omega k},$$

this can be written as

$$\overline{z}_{\omega k} \in \operatorname*{argmin}_{z_{\omega k} \in R} \{ \psi_{\omega k}(z_{\omega k}) - \overline{v}_{\omega k} z_{\omega k} \},$$

or $0 \in \partial \psi_{\omega k}(\overline{z}_{\omega k}) - \overline{v}_{\omega k}$, and then equivalently as $\overline{v}_{\omega k} \in \partial \psi_{\omega k}(\overline{z}_{\omega k})$ or $\overline{z}_{\omega k} \in \partial \varphi_{\omega k}(\overline{v}_{\omega k})$. The latter reduces to $\overline{z}_{\omega k} = \varphi'_{\omega k}(\overline{v}_{\omega k})$ and condition (1.6) when $\varphi_{\omega k}$ is differentiable, as is the case when $p_{\omega k} > 0$. The derivation of (1.5) from the saddle point property is similar. □

This formulation of stochastic programming duality differs somewhat from the one in our basic paper [1]. In order to facilitate the application of the results in [1] to the present context, an explanation of the connection is needed. In [1], problem (P_{sto}) is associated with a different minimax problem, namely for

$$L_0(x,\underset{\sim}{z}) = \sum_{j=1}^{n} [c_j x_j - \tfrac{1}{2} r_j x_j^2] + \sum_{k=1}^{\ell} E\{\underset{\sim}{z}_k [\underset{\sim}{h}_k - \sum_{j=1}^{n} \underset{\sim}{t}_{kj} x_j] + \tfrac{1}{2} \underset{\sim}{p}_k \underset{\sim}{z}_k^2\} \qquad (1.16)$$

on $X_0 \times Z$, where Z is still the set in (1.11) but X_0 is the set of feasible solutions to (P_{sto}):

$$X_0 = \{x = (x_1, \ldots, x_n) \mid 0 \le x_j \le s_j, \sum_{j=1}^{n} a_{ij} x_j \le b_i\}. \qquad (1.17)$$

This leads to the dual problem

$$\text{minimize} \quad g_0(\underset{\sim}{z}) \quad \text{over all} \quad \underset{\sim}{z} \in Z, \qquad (D_{sto}^0)$$

where

$$g_0(\underset{\sim}{z}) = \min_{y \in Y} g(y, \underset{\sim}{z}). \qquad (1.18)$$

Indeed, one has in parallel to (1.14), (1.15), that

$$\min_{\underset{\sim}{z} \in Z} L_0(x, \underset{\sim}{z}) = f(x) \quad \text{for all} \quad x \in X_0, \qquad (1.19)$$

and by quadratic programming duality (using (A1))

$$\begin{aligned}
\max_{z \in X_0} L_0(x, \underset{\sim}{z}) &= \max_{x \in X} \inf_{y \in Y} L(x, y, \underset{\sim}{z}) \\
&= \min_{y \in Y} \max_{x \in X} L(x, y, \underset{\sim}{z}) = \min_{y \in Y} g(y, \underset{\sim}{z}) \quad \text{for all} \quad \underset{\sim}{z} \in Z.
\end{aligned} \qquad (1.20)$$

(Actually in [1] one has minimization in the primal problem and maximization in the dual, but that calls for only a minor adjustment.) Obviously, then, the pairs $(\overline{y}, \overline{\underset{\sim}{z}})$ that solve (D_{sto}) are the ones such that $\overline{\underset{\sim}{z}}$ solves (D_{sto}^0) and $\overline{y}$ provides the corresponding minimum (1.18).

2. FINITE GENERATION ALGORITHM IN THE STRICTLY QUADRATIC CASE

The basic idea of our computational procedure is easy to describe. We limit attention for the time being to the strictly quadratic case where $r_j > 0$ and $p_k > 0$, because we will be able to show in section 4 that problems that are not strictly quadratic can be made so as part of an additional iterative process. This limitation also simplifies the exposition and helps us focus on the results we believe to be the most significant. It is not truly necessary, however. A more general version of what follows could likewise be deduced from the fundamental theory in [1].

In problem (D_{sto}) we minimize a certain convex function $g(y, \underset{\sim}{z})$ over $Y \times Z$, where Y and Z are the convex polyhedra in (1.10) and (1.11). As we have seen in the proof of Theorem 1, this corresponds to finding a saddlepoint $(\overline{x}, \overline{y}, \overline{\underset{\sim}{z}})$ of the function L in (1.12) relative to $X \times (Y \times Z)$, where X is the polyhedron in (1.9). Indeed, if $(\overline{y}, \overline{\underset{\sim}{z}})$ is optimal for (D_{sto}), then the $\overline{x}$ obtained from formula (1.5) gives us the saddlepoint. This $\overline{x}$ is the unique optimal solution to (P_{sto}).

The trouble is, however, that because of the potentially very high dimensionality of Z (whose elements $\underset{\sim}{z}$ have components $z_{\omega k}$ for $k = 1, \ldots, \ell$ and all $\omega \in \Omega$, with Ω possibly very large), we cannot hope to solve (D_{sto}) directly, even though it is reducible in principle to a quadratic programming problem. What we do instead is develop a method of descent which produces a minimizing sequence $\{(\overline{y}^{\nu}, \overline{\underset{\sim}{z}}^{\nu})\}_{\nu=1}^{\infty}$ in (D_{sto}) and at the same time, by formula (1.5), a maximizing sequence $\{\overline{x}^{\nu}\}_{\nu=1}^{\infty}$ in (P_{sto}).

In this method we "generate Z finitely from within". Let Z be expressed as

$$Z = Z_1 \times \ldots \times Z_{\ell} \quad \text{with} \quad Z_k = \{\underset{\sim}{z}_k \mid 0 \le \underset{\sim}{z}_k \le \underset{\sim}{q}_k\}. \qquad (2.1)$$

At iteration ν we take a finite subset $\widehat{Z}_k^{\nu}$ of Z_k, and instead of minimizing $g(y, \underset{\sim}{z})$ over $Y \times Z$ we minimize it over $Y \times Z^{\nu}$, where

$$Z^{\nu} = Z_1^{\nu} \times \ldots \times Z_{\ell}^{\nu} \quad \text{with} \quad Z_k^{\nu} = \text{co}\{0, \widehat{Z}_k^{\nu}\}. \qquad (2.2)$$

By employing a parametric representation of the convex hull $\text{co}\{0, \widehat{Z}_k^{\nu}\}$ and keeping the number of elements in $\widehat{Z}_k^{\nu}$ small, which turns out always to be possible, we are able to express this subproblem as one of quadratic programming in a relatively small number of variables. This subproblem is deterministic in character; the coefficients are certain expectations in terms of the given random variables $\underset{\sim}{t}_{kj}, \underset{\sim}{h}_k, \underset{\sim}{p}_k$ and the chosen random variables in $\widehat{Z}_k^{\nu}$.

The details of the subproblem will be explained in due course (§3). First we state the algorithm more formally and establish its convergence properties.

FINITE GENERATION ALGORITHM (*version under the strict quadraticity assumption that $r_j > 0$ and $p_k > 0$.*)

Step 0 (Initialization). *Choose finite subsets $\widehat{Z}_k^1 \subset Z_k$ for $k = 1, \ldots, \ell$. Set $\nu = 1$.*

Step 1 ((Quadratic Programming Subproblem). *Calculate an optimal solution $(\overline{y}^{\nu}, \overline{\underset{\sim}{z}}^{\nu})$ to the problem of minimizing $g(y, \underset{\sim}{z})$ over $Y \times Z^{\nu}$, where Z^{ν} is given by (2.2). Denote the minimum value by $\overline{\alpha}_{\nu}$. Define $\overline{x}^{\nu}$ from $(\overline{y}^{\nu}, \overline{\underset{\sim}{z}}^{\nu})$ by formula (1.5).*

Step 2 (Generation of Test Data). *Define $\underset{\sim}{z}^{\nu}$ from $\overline{x}^{\nu}$ by formula (1.6). Set $\alpha_{\nu} = L_0(\overline{x}^{\nu}, z^{\nu})$ in (1.16).*

Step 3 (Optimality Test). Define $\varepsilon_\nu = \overline{\alpha}_\nu - \alpha_\nu \geq 0$. Then $\overline{x}^\nu$ is an ε_ν-optimal solution to (P_{sto}), $(\overline{y}^\nu, \overline{\underline{z}}^\nu)$ is an ε_ν-optimal solution to (D_{sto}), and

$$\overline{\alpha}_\nu \geq \max (P_{sto}) = \min (D_{sto}) \geq \alpha_\nu.$$

(Stop if this is good enough.)

Step 4 (Polytope Modification). For each $k = 1, \ldots, \ell$, choose a finite set $\widehat{Z}_k^{\nu+1} \subset Z_k$ whose convex hull contains both $\overline{z}^\nu$ and $\underline{z}^\nu$. Replace ν by $\nu + 1$ and return to Step 1.

Note the very mild condition in Step 4 on the choice of $\widehat{Z}_k^{\nu+1}$. One could simply take

$$\widehat{Z}_k^{\nu+1} = \{\overline{\underline{z}}_k^\nu, \underline{z}_k^\nu\}, \tag{2.3}$$

or at the opposite extreme,

$$\widehat{Z}_k^{\nu+1} = \widehat{Z}_k^\nu \cup \{\underline{z}_k^\nu\}. \tag{2.4}$$

Another possibility would be

$$\widehat{Z}_k^{\nu+1} = \widehat{Z}_k^1 \cup \{\overline{\underline{z}}_k^\nu, \underline{z}_k^\nu\} \tag{2.5}$$

in all iterations, with $\widehat{Z}_k^1$ selected initially to provide a certain richness of representation. Although the number of elements of $\widehat{Z}_k^\nu$ (which determines the dimensionality of the quadratic programming subproblem in Step 1) would continue to grow indefinitely under (2.4), it stays fixed under (2.3) or (2.5).

For the statement of our convergence result we introduce the vector norms

$$\|x\|_r = \Big[\sum_{j=1}^n r_j x_j^2\Big]^{1/2}, \tag{2.6}$$

$$\|z\|_p = \Big[\sum_{k=1}^\ell p_k z_k^2\Big]^{1/2}, \tag{2.7}$$

and matrix norm

$$\|T\|_{p,r} = \max\{z \cdot Tx \mid \|z\|_p \leq 1, \|x\|_r \leq 1\}. \tag{2.8}$$

THEOREM 2. *Under the strict quadraticity assumption that $r_j > 0$ and $p_k > 0$, the sequence $\{\overline{x}^\nu\}_{\nu=1}^\infty$ produced by the finite generation algorithm converges to the unique optimal solution $\overline{x}$ to (P_{sto}). Moreover it does so at a linear rate, in the following sense.*

Let σ be an upper bound to the range of the (finitely discrete) random variable $\|\underline{T}\|_{p,r}$ in (2.8), where $\underline{T}$ is the matrix with entries $\underline{t}_{kj}$. Let $\tau \in [0,1)$ be the factor defined by

$$\tau = \begin{cases} \sigma^2 & \text{if } \sigma^2 \leq \frac{1}{2}, \\ 1 - (1/4\sigma^2) & \text{if } \sigma^2 \geq \frac{1}{2}. \end{cases} \tag{2.9}$$

Then in terms of the values

$$\overline{\alpha} = \max (P_{sto}) = \min (D_{sto}) \quad and \quad \overline{\varepsilon}_\nu = \overline{\alpha} - \overline{\alpha}_\nu \leq \varepsilon_\nu \tag{2.10}$$

one has

$$\overline{\varepsilon}_{\nu+\mu} \leq \tau^\mu \varepsilon_\nu \quad \text{for all} \quad \nu = 1, 2, \ldots, \quad \text{and} \quad \mu = 1, 2, \ldots, \tag{2.11}$$

$$\|\overline{x} - \overline{x}^{\nu+\mu}\|_r \leq [2\tau^\mu \varepsilon_\nu]^{1/2} \quad \text{for all} \quad \nu = 1, 2, \ldots, \quad \text{and} \quad \mu = 1, 2, \ldots. \tag{2.12}$$

Observe well that in (2.11) and (2.12) the estimates are claimed for *all* ν and μ, *not* just when ν is sufficiently large. Most convergence results are not of such type, so this is rather surprising, especially in view of the fact that the factor $\tau \in [0, 1)$ can in principle, at least, be estimated in advance of computation, right from the given data. Moreover τ does not depend on any data in the problem other than $\underline{t}_{kj}, p_k$ and r_j. In the special case of nonrandom t_{kj} and p_k (the only random variables in the problem being $\underline{h}_k$ and $\underline{q}_k$), one can simply take $\sigma = \|T\|_{p,r}$.

PROOF OF THEOREM 2. The procedure specified here is a special case of the algorithm presented in [1], as can be seen in the following way. In calculating a pair $(\overline{y}^\nu, \overline{\underline{z}}^\nu)$ that minimizes $g(y, \underline{z})$ over $Y \times Z^\nu$ in the subproblem in Step 1, we obtain a solution $\overline{\underline{z}}^\nu$ to the different subproblem of [1], in which $g_0(\underline{z})$ is minimized over Z^ν (with g_0 the function in (1.18)). The number $\overline{\alpha}_\nu$ is the optimal value in both subproblems, and $\overline{x}^\nu$ furnishes the saddle point $\overline{x}^\nu, \overline{y}^\nu, \overline{\underline{z}}^\nu$ to L on $X \times (Y \times Z^\nu)$ in the present formulation, but also the saddlepoint $(\overline{x}^\nu, \overline{\underline{z}}^\nu)$ to L_0 on $X \times Z^\nu$, as required by Step 1 of the algorithm as formulated in [1].

The elements $\underline{z}^\nu$ and α_ν calculated in Step 2 satisfy

$$\underline{z}^\nu = \underset{\underline{z} \in Z}{\operatorname{argmin}} L_0(\overline{x}^\nu, \underline{z}), \qquad \alpha_\nu = \min_{\underline{z} \in Z} L_0(\overline{x}^\nu, \underline{z}). \tag{2.13}$$

Thus these are the same as the elements calculated in the version of Step 2 in [1] (except for a notational switch between maximization and minimization). Of course they are given here by closed formulas, whereas in the far more general setting of [1] they might have to be calculated by solving a large collection of quadratic programming subproblems in the random components $z_{\omega k}$.

The updated polyhedron $Z^{\nu+1}$ does contain $\overline{\underline{z}}^\nu$ and $\underline{z}^\nu$ under the conditions in Step 4, as required by the conditions in the more general version of Step 4 in [1].

Thus all the conditions in Theorem 5 of [1] are fulfilled, and the stated convergence properties follow, provided that we reconcile the choice of σ given here with the corresponding one in [1]. The condition specified in [1, Theorem 5] is that

$$\sigma^2 \|z\|_p^2 \geq \|\underline{T}^* z\|_{r^{-1}}^2 \tag{2.14}$$

for all realizations of the random vector p and matrix T and all possible choices of the vector z. Here we are using the notation $r^{-1} = (r_1^{-1}, \ldots, r_n^{-1})$. The norm $\|\cdot\|_{r^{-1}}$ is the dual of the norm $\|\cdot\|_r$ in (2.6). so

$$\|T^* z\|_{r^{-1}} = \max\{(T^* z) \cdot x \mid \|x\|_r \leq 1\}.$$

($T^* =$ transpose of T.) Therefore one has

$$\max\{\|T^* z\|_{r^{-1}} \mid \|z\|_p \leq 1\} = \|T\|_{p,r}$$

as defined in (2.8). This shows that (2.14) is equivalent to

$$\sigma \geq \|T\|_{p,r}$$

and the proof of Theorem 2 is thereby completed. □

3. SOLVING THE QUADRATIC PROGRAMMING SUBPROBLEM.

Returning now to the elucidation of the finite generation algorithm and how it may be implemented. we demonstrate that the subproblem in Step 2 can be represented easily as an ordinary quadratic programming problem of relatively low dimension and thereby solved using standard codes. Explicit notation for the elements of the finite sets $\hat{Z}_k^\nu$ selected from Z_k is now needed. Let us suppose that

$$\hat{Z}_k^\nu = \{\hat{z}_{k\alpha}^\nu \mid \alpha = 1, \ldots, m_\nu\}. \tag{3.1}$$

This yields

$$Z_k^\nu = \mathrm{co}\{0, \hat{Z}_k^\nu\} = \{z_k = \sum_{\alpha=1}^{m_\nu} \lambda_{k\alpha} \hat{z}_{k\alpha}^\nu \mid \lambda_{k\alpha} \geq 0, \sum_{\alpha=1}^{m_\nu} \lambda_{k\alpha} \leq 1\}. \tag{3.2}$$

In Step 2 we want to minimize the objective $g(y, z)$ in (D_{sto}) not over all of $Y \times Z$ (the variables w_j standing for linear expressions in y and z), but only over $Y \times Z^\nu$. By virtue of (3.2) we can substitute for the elements z of interest in this subproblem certain linear expressions in the parameters $\lambda_{k\mu}$. In this way we get the function

$$g^\nu(y, \lambda) = \sum_{i=1}^m b_i y_i + \sum_{j=1}^n \rho(w_j; r_j, s_j)$$
$$+ E\{\sum_{k=1}^\ell [h_k(\sum_{\alpha=1}^{m_\nu} \lambda_{k\alpha} \hat{z}_{k\alpha}^\nu) + \frac{1}{2} p_k(\sum_{\alpha=1}^{m_\nu} \lambda_{k\alpha} \hat{z}_{k\alpha}^\nu)^2]\}, \tag{3.3}$$

where

$$w_j = c_j - \sum_{i=1}^m y_i a_{ij} - E\{\sum_{k=1}^\ell (\sum_{\alpha=1}^{m_\nu} \lambda_{k\alpha} \hat{z}_{k\alpha}^\nu) t_{kj}\}. \tag{3.4}$$

But these complicated expressions can greatly be reduced by carrying the expectation operation through the sums to get explicit coefficients for the parameters $\lambda_{k\mu}$. Specifically. let

$$\bar{h}_{k\alpha}^\nu = E\{h_k \hat{z}_{k\mu}^\nu\}, \tag{3.5}$$

$$\bar{p}_{k\alpha\beta}^\nu = E\{p_k \hat{z}_{k\alpha}^\nu \hat{z}_{k\beta}^\nu\}, \tag{3.6}$$

$$\bar{t}_{k\alpha j}^\nu = E\{\hat{z}_{k\alpha}^\nu t_{kj}\}. \tag{3.7}$$

Then

$$g^\nu(y, \lambda) = \sum_{i=1}^m b_i y_i + \sum_{j=1}^n \rho(w_j; r_j, s_j)$$
$$+ \sum_{k=1}^\ell [\sum_{\alpha=1}^{m_\nu} \bar{h}_{k\alpha}^\nu \lambda_{k\alpha} - \frac{1}{2} \sum_{\alpha,\beta=1}^{m_\nu, m_\nu} \bar{p}_{k\alpha\beta}^\nu \lambda_{k\alpha} \lambda_{k\beta}], \tag{3.8}$$

where

$$w_j = c_j - \sum_{i=1}^m y_i a_{ij} - \sum_{k=1}^\ell \sum_{\alpha=1}^{m_\nu} \lambda_{k\alpha} \bar{t}_{k\alpha j}^\nu. \tag{3.9}$$

Finally let us observe that the penalty expression $\rho(w_j; r_j, s_j)$ in these formulas, as given by (1.1), satisfies

$$\rho(w_j; r_j, s_j) = \text{minimum of } s_j w_{1j} + \frac{1}{2} w_{2j}^2/r_j$$
$$\text{subject to } w_{1j} \geq 0, w_{1j} + w_{2j} \geq w_j. \tag{3.10}$$

Moreover

$$\rho'(w_j; r_j, s_j) = \text{Lagrange multiplier } (\geq 0) \text{ for the constraint}$$
$$w_{1j} + w_{2j} \geq w_j \text{ in (3.10)}. \tag{3.11}$$

With these facts in mind we pose the quadratic programming problem

$$\text{minimize} \quad \sum_{i=1}^m b_i y_i + \sum_{j=1}^n [s_j w_{1j} + \frac{1}{2} w_{2j}^2/r_j]$$

(D$^\nu$)
$$+ \sum_{k=1}^\ell [\sum_{\alpha=1}^{m_\nu} \bar{h}_{k\alpha}^\nu \lambda_{k\alpha} - \frac{1}{2} \sum_{\alpha,\beta=1}^{m_\nu, m_\nu} \bar{p}_{k\alpha\beta}^\nu \lambda_{k\alpha} \lambda_{k\beta}]$$

$$\text{subject to } y_i \geq 0, w_{1j} \geq 0, \lambda_{k\alpha} \geq 0,$$

$$\sum_{\alpha=1}^{m_\nu} \lambda_{k\alpha} \leq 1 \text{ for } k = 1, \ldots, \ell,$$

$$\sum_{i=1}^m y_i a_{ij} + \sum_{k=1}^\ell \sum_{\alpha=1}^{m_\nu} \lambda_{k\alpha} \bar{t}_{k\alpha j}^\nu + w_{1j} + w_{2j} \geq c_j \text{ for } j = 1, \ldots, n. \tag{3.12}$$

We then have the following implementation.

SUBALGORITHM (for Step 2). *Given the sets* $\widehat{Z}_k^\nu$ *in the notation (3.1), calculate the co-efficients (3.5), (3.6), (3.7), for the quadratic programming problem* (D^ν). *Solve* (D^ν) *by any method, getting from the optimal solution values* $\overline{y}_i^\nu, \overline{w}_{1j}^\nu, \overline{w}_{2j}^\nu$, *and* $\overline{\lambda}_{k\alpha}^\nu$ *the elements*

$$\overline{z}_k = \sum_{\alpha=1}^{m_\nu} \overline{\lambda}_{k\alpha}^\nu \widehat{z}_{k\alpha}^\nu.$$

The minimum value in (D^ν) is the desired α_ν, and the Lagrange multiplier vector obtained for the constraints (3.12) in (D^ν) is the desired approximate solution $\overline{x}^\nu$ to (P_{sto}).

Thus it is not actually necessary in Step 2 to invoke formula (1.5) to get $\overline{x}^\nu$. Instead, $\overline{x}^\nu$ can be obtained as a byproduct of the solution procedure used for the minimization.

4. APPLICATION TO PROBLEMS THAT ARE NOT STRICTLY QUADRATIC.

If in the given problem (P_{sto}) it is not true that $r_j > 0$ and $p_k > 0$ for all j and k, we use the proximal point technique [4] (as adapted to the Lagrangian $L_0(x,z)$ in (1.16)) to replace (P_{sto}) by a sequence of problems (P_{sto}^μ), $\mu = 1, 2, \ldots$, that do have the desired character. To each problem (P_{sto}^μ) we apply the finite generation algorithm as above, but with a certain stopping criterion in Step 3 that ensures finite termination. This is done in such a way that the overall doubly iterative procedure still converges at a linear rate.

To obtain the problems (P_{sto}^μ), we introduce alongside the given values r_j and p_k some other values $\overline{r}_j > 0$, $\overline{p}_k > 0$ and set

$$r_{*j} = r_j + \eta \overline{r}_j, \qquad p_{*k} = p_k + \eta \overline{p}_k, \tag{4.1}$$

where $\eta > 0$ is a parameter value that will play a role in theory but can be held fixed for the purpose of computation. We also introduce elements

$$\overline{x}_*^\mu = (\overline{x}_{*1}^\mu, \ldots, \overline{x}_{*n}^\mu) \quad \text{and} \quad \overline{z}_*^\mu = (\overline{z}_{*1}^\mu, \ldots, \overline{z}_{*l}^\mu),$$

which are to be thought of as estimates for the optimal solution values in (P_{sto}) and (D_{sto}). In terms of these we set

$$c_{*j}^\mu = c_j - \eta \overline{r}_j \overline{x}_{*j}^\mu \quad \text{and} \quad h_{*k}^\mu = h_k - \eta p_k \overline{z}_{*k}^\mu. \tag{4.2}$$

Then

(P_{sto}^μ), (D_{sto}^μ) are the problems obtained by replacing

$$r_j, p_{*k}, c_j \text{ and } h_k \text{ in } (P_{sto}), (D_{sto}) \text{ by } r_{*j}, p_{*k}, c_{*j}^\mu \text{ and } h_{*k}^\mu. \tag{4.3}$$

These modified problems are, of course, strictly quadratic: one has $r_{*j} > 0$ and $p_{*k} > 0$.

MASTER ALGORITHM.

Step 0 (Initialization). *Choose* $\overline{x}_*^1 \in X$ *and* $\overline{z}_*^1 \in Z$. *Set* $\mu = 1$.

Step 1 (Finite Generation Algorithm). *Apply the finite generation algorithm in the manner already described to the strictly quadratic problems* (P_{sto}^μ) *and* (D_{sto}^μ) *in (4.3). Terminate in Step 3 when the stopping criterion given below is satisfied.*

Step 2 (Update). *For the elements* $\overline{x}^\nu$ *and* $\overline{z}^\nu$ *with which Step 1 terminated, set* $\overline{x}_*^{\mu+1} = \overline{x}^\nu$ *and* $\overline{z}_*^{\mu+1} = \overline{z}^\nu$. *Replace* μ *by* $\mu + 1$ *and return to Step 1.*

The stopping criterion is as follows. In terms of the norm

$$\|(x, z)\| = [\|x\|_r^2 + E\{\|z\|_p^2\}]^{1/2} \tag{4.4}$$

and a sequence of values θ_μ with

$$\theta_\mu > 0, \quad \sum_{\mu=1}^{\infty} \theta_\mu < \infty, \tag{4.5}$$

we define the function

$$\varepsilon_{*\mu}(x, z) = \theta_\mu^2 \min\{1, (\eta/2)\|(x, z) - (\overline{x}_*^\mu, \overline{z}_*^\mu)\|^2\}. \tag{4.6}$$

We stop in Step 3 of the finite generation algorithm when the computed elements $\varepsilon_\nu, \overline{x}^\nu$ and $\overline{z}^\nu$ satisfy

$$\varepsilon_\nu \leq \varepsilon_{*\mu}(\overline{x}^\nu, \overline{z}^\nu). \tag{4.7}$$

This stopping criterion will eventually be satisfied, when ν is high enough; the only exception is the case where $\overline{x}_*^\mu$ happens already to be an optimal solution $\overline{x}$ to (P_{sto}) and $\overline{z}_*^\mu$ the $\overline{z}$-component of an optimal solution $(\overline{y}, \overline{z})$ to (D_{sto}). (See [1, §6] for details.)

THEOREM 3. *If the master algorithm is executed with the specified stopping criterion (4.7), then the sequences* $\{\overline{x}_*^\mu\}_{\mu=1}^\infty$ *and* $\{\overline{z}_*^\mu\}_{\mu=1}^\infty$ *converge to particular elements* $\overline{x}$ *and* $\overline{z}$, *where* $\overline{x}$ *is an optimal solution to* (P_{sto}) *and, for some* $\overline{y}$, *the pair* $(\overline{y}, \overline{z})$ *is an optimal solution to* (D_{sto}). *Mooveover there is a number* $\beta(\eta) \in [0, 1)$ *such that* $(\overline{x}_*^\mu, \overline{z}_*^\mu)$ *converges to* $(\overline{x}, \overline{z})$ *at a linear rate with modulus* $\beta(\eta)$.

PROOF. This is an immediate specialization of Theorem 6 of [1] to the case at hand, the path of specialization having been established already in the proof of Theorems 1 and 2. □

The theory of proximal point technique in [4], as applied in the derivation of Theorem 3, shows actually that linear convergence is obtained at the rate

$$\beta(\eta) = \gamma\eta/[1 + (\gamma\eta)^2]^{1/2} \in [0, 1) \tag{4.8}$$

where $\gamma \geq 0$ is a number depending only on the data in the original problems (P_{sto}) and (D_{sto}), not on $\eta, \bar{r}, $ or $\bar{p}_k$. In particular $\beta(\eta) \to 0$ as $\eta \to 0$. Thus an arbitrarily good rate of convergence can be obtained (in principle) for the outer algorithm (master algorithm) simply by choosing the parameter value η *small* enough.

At the same time, however, the choice of η affects the convergence rate in the inner algorithm (finite generation algorithm). That rate corresponds by (2.12) to a number $\tau(\eta)^{1/2} \in [0, 1)$ defined by (2.9) in terms of an upper bound $\sigma(\eta)$ for $\| \underaccent{\tilde}{T} \|_{\underaccent{\tilde}{p}_*, r_*}$, where $\underaccent{\tilde}{p}_*$ and r_* are vectors consisting of the parameters in (4.1). Thus $\sigma(\eta)^2$ is an upper bound for the expression

$$\frac{[z \cdot \underaccent{\tilde}{T} \, x]^2}{[\|z\|_{\underaccent{\tilde}{p}}^2 + \eta \|z\|_{\underaccent{\tilde}{\bar{p}}}^2] \, [\|x\|_r^2 + \eta \|x\|_{\bar{r}}^2]}$$

over all possible choices of the vectors $x \in R^n$ and $z \in R^l$ and all possible values taken on by the random variables $\underaccent{\tilde}{T}, \underaccent{\tilde}{p}$ and $\bar{\underaccent{\tilde}{p}}$. It follows that $\tau(\eta) \to 0$ as $\eta \to \infty$ but $\tau(\eta) \to 1$ as $\eta \to 0$. Thus an arbitrarily good rate of convergence can be obtained (in principle) for the inner algorithm by choosing η *large* enough, but too small a choice could do damage.

This trade-off between the outer and inner algorithms in the choice of η could be a source of difficulty in practice, although we have not had much trouble with the problems tried so far. (See King [3].)

References

1. R.T. Rockafellar and R.J.-B. Wets, "A Lagrangian finite generation technique for solving linear-quadratic problems in stochastic programming," Math. Programming Studies (1985), to appear.

2. 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, 1983, 252-265.

3. A. King, "An implementation of the Lagrangian finite generation method," in: Y. Ermoliev and R.J.-B. Wets, eds.. *Numerical Techniques for Stochastic Programming Problems*, Springer Verlag, to appear.

4. R.T. Rockafellar, "Monotone operators and the proximal point algorithm," SIAM Journal on Control and Optimization 14(1976), 877-898.

5. R.T. Rockafellar, *Network Flows and Monotropic Optimization*, Wiley-Interscience, New York, 1984.