A Quadratic Recourse Function for the Two-Stage Stochastic Program

John R. Birge
Stephen M. Pollock
Department of Industrial and Operations Engineering
The University of Michigan
Ann Arbor MI 48109-2117, USA
and
Liqun Qi
School of Mathematics
The University of New South Wales
Sydney, New South Wales 2052, Australia

Technical Report 95-13

A Quadratic Recourse Function for the Two-Stage Stochastic Program

John R. Birge, [1]
Stephen M. Pollock
Department of Industrial and Operations Engineering
University of Michigan, Ann Arbor, MI 48109, USA

and

Liqun Qi [2]
School of Mathematics
The University of New South Wales
Sydney, New South Wales 2052, Australia

July 28, 1995

Abstract

We present a quadratic recourse representation of the two-stage stochastic linear problem. Unlike the usual linear recourse model, it is differentiable with respect to the first stage decision variables. This offers the possibility of applying high convergence rate methods to solve the two-stage problem. We show that the quadratic recourse function approximates the linear recourse function (and the corresponding solution of the two-stage problem with quadratic recourse converges to the solution of the two-stage problem with linear recourse) as a parameter $k \to \infty$ and another parameter $\epsilon_k \to 0$. We also give a bound for this approximation.

[1]This author's work was supported in part by the National Science Foundation under Award Number DDM-9215921.

[2]This author's work was supported in part by the Australian Research Council.

1 Introduction

One of the main paradigms of stochastic programming is represented by the two-stage stochastic linear program formulated as a master problem and a recourse problem [8, 9]. The *master problem* is

$$\begin{array}{ll} \min & f(x) = c^t x + \Phi(x) \\ \text{s.t.} & Ax = b \\ & x \geq 0, \end{array} \tag{1}$$

where $x \in \Re^n$ is the first-stage decision vector, $c \in \Re^n$ is the cost coefficient vector for x, $Ax = b$ and $x \geq 0$ are the linear constraints on x, with $b \in \Re^m$ and $A \in \Re^{m \times n}$.

$\Phi(x)$ is the expected value of the linear recourse function, $\Phi(x) = E(\phi(x, \xi))$, where $\phi(x, \xi)$ is defined by the *recourse problem*

$$\begin{array}{rllll} \phi(x,\xi) = & \min & (q(\xi))^t y & & \\ & \text{s.t.} & W(\xi)y & = & T(\xi)x - h(\xi) \\ & & y & \geq & 0. \end{array} \tag{2}$$

Here ξ is a random variable of dimension r with distribution $P(\cdot)$, so by definition,

$$\Phi(x) = E(\phi(x, \xi)) = \int \phi(x, \xi) P(d\xi). \tag{3}$$

The vector, $y \in \Re^s$, is the second-stage decision vector, $h \in \Re^r$ is the demand or resource vector, and $T \in \Re^{r \times n}$ is a technology matrix.

The usual origin of the constraint equations in (2) is the desire to satisfy the condition $T(\xi)x - h(\xi) = 0$. However, because of the random nature of both T and h, $T(\xi)x - h(\xi)$ does not equal zero in general. Thus Wy is introduced to represent the "discrepancy" (where $W \in \Re^{r \times s}$) and $q \in \Re^s$, $q > 0$, is the associated cost coefficient vector for non-zero y.

To simplify our discussion, we only discuss in detail the *fixed* recourse problem: when W does not depend upon ξ. We also assume that T and q are independent of ξ and $h = \xi$, conditions corresponding to uncertain resource levels but known prices and technologies. Equation (2) then becomes

$$\begin{array}{rllll} \phi(x,\xi) = & \min & q^t y & & \\ & \text{s.t.} & Wy & = & Tx - \xi \\ & & y & \geq & 0. \end{array} \tag{4}$$

For convenience, we let ξ be a discrete random vector, so that

$$\Phi(x) = \sum_{j=1}^{l} \phi(x, \xi_j) p_j, \tag{5}$$

where $p_j \geq 0$ and $\sum_{j=1}^{l} p_j = 1$. Our results can be readily extended to continuous ξ.

There are two disadvantages associated with the two-stage stochastic program with fixed recourse represented by equations (1), (5) and (4). First, generally Φ is nondifferentiable with respect to x, which prevents the use of algorithms with high rates of convergence. Secondly, (4) may not be feasible for some x. (A problem with (4) feasible for all x is called a problem with *complete recourse*. If (4) is feasible for all x satisfying the linear constraints of (1), then it is called a problem with *relatively complete* recourse.)

To address these difficulties we consider below a problem with a quadratic recourse function $\phi_k(x, \xi)$, where $k > 0$ is a parameter. This quadratic recourse function is always continuously differentiable, which paves the way for using algorithms with high rates of convergence to solve the two-stage stochastic program.

Two-stage quadratic recourse models have been proposed before. In particular, Rockafellar and Wets discuss the extended linear-quadratic problem (ELQP) [18]. ELQP introduces *additional* coefficient matrices but there are no direct links between its solutions and the solutions of the two-stage stochastic program with linear recourse. On the other hand, our quadratic recourse function will have a direct link with (4) in that it is proposed as an alternative way to model the original linear recourse problem. We show that the quadratic recourse function $\phi_k(x, \xi)$ converges to the linear recourse function $\phi(x, \xi)$, and the solution of the two-stage stochastic program with quadratic recourse converges to the solution of the two-stage stochastic program with linear recourse respectively, as $k \to \infty$. An error bound is also given for this convergence.

The remaining part of this paper is as follows. In Section 2 we present the quadratic recourse function $\phi_k(x, \xi)$, discuss its meanings and point out its differences from the ELQP model. We discuss its differentiability properties and algorithmic meanings in Section 3. In Section 4 we prove that it converges to the linear recourse function as $k \to \infty$. An error bound is given for this convergence and a simple example is presented to illustrate the use of this bound. We also show that the solution of the two-stage stochastic program with recourse $\phi_k(x, \xi)$ converges to the solution of the two-stage stochastic program with recourse $\phi(x, \xi)$. In Section 5 we make some concluding remarks.

2 The Recourse Function

We now consider an alternative to (1), (4) and (5) that has the same second-stage optimal solution under our nonnegative objective assumptions. We do this by minimizing the square of the recourse function; i.e., we solve

$$\begin{array}{rlrcl} \phi^2(x, \xi) = & \min & (q^t y)^2 & & \\ & \text{s.t.} & Wy & = & Tx - \xi \\ & & y & \geq & 0, \end{array} \qquad (6)$$

for $q > 0$.

We could use this function directly in an optimization procedure, but more than first-order differentiability is needed for high convergence rate methods. We therefore approximate the problem represented in (6) by the following parametrized quadratic recourse function:

$$\begin{aligned} \phi_k^2(x,\xi) = \quad & \min \quad (q^t y)^2 + k\|Wy - Tx + \xi\|^2 + \epsilon_k \\ & \text{s.t.} \quad y \geq 0, \end{aligned} \tag{7}$$

where $\|v\|^2 = v^t v$ is the 2-norm of the vector v, and k and ϵ_k are two positive parameters. We also assume $\phi_k(x,\xi) \geq 0$. Thus, there is a clearly defined ϕ_k that satisfies equation (7). Correspondingly, (1) and (5) are replaced by

$$\begin{aligned} \min \quad & f_k(x) = c^t x + \Phi_k(x) \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0, \end{aligned} \tag{8}$$

where

$$\Phi_k(x) \equiv E(\phi_k(x,\xi)) = \sum_{j=1}^{l} \phi_k(x,\xi_j) p_j. \tag{9}$$

We may think of (7) as an alternative representation of the situation expressed by the equations of (6). Instead of minimizing $(q^t y)^2$ with the constraint $Wy = Tx - \xi$, we minimize the weighted sum of $(q^t y)^2$ and $\|Wy - Tx + \xi\|^2$, with k reflecting the relative importance of satisfying the $Tx - \xi = Wy$ constraints compared to minimizing $(q^t y)^2$. The additional parameter ϵ_k is added to ensure that $\phi_k^2 \geq \epsilon_k > 0$. This will be useful for establishing differentiability properties of ϕ_k. [We could also replace k by an $n \times n$ diagonal matrix K such that each diagonal element of K "weights" a component of $Wy - Tx + \xi$. In this case, (7) would have the form

$$\begin{aligned} \phi_k^2(x,\xi) = \quad & \min \quad (q^t y)^2 + (Wy - Tx + \xi)^t K (Wy - Tx + \xi) + \epsilon_k \\ & \text{s.t.} \quad y \geq 0. \end{aligned} \tag{10}$$

Again, for simplicity, we restrict our discussion to the form (7), although the results below can be shown to hold under the conditions $\|K\| \to \infty$ and $\epsilon_k \to 0$.]

The idea of converting the objective function of (2) by a quadratic form is not new. For example, the ELQP model considers the dual problem of (4):

$$\begin{aligned} \phi(x,\xi) = \quad & \max \quad (Tx - \xi)^t u \\ & \text{s.t.} \quad W^T u \leq q, \end{aligned} \tag{11}$$

where u is the dual variable vector. ELQP adds a quadratic term $\frac{1}{2}u^t H u$ to (11), to obtain

$$\begin{aligned} \phi_H(x,\xi) = \quad & \max \quad (Tx - \xi)^t u + \tfrac{1}{2} u^t H u \\ & \text{s.t.} \quad W^T u \leq q, \end{aligned} \tag{12}$$

where H is an $n \times n$ positive definite symmetric monitor matrix.

As a result of formulations such as (12), superlinearly convergent algorithms have been developed by Qi and Womersley [16], Chen, Qi and Womersley [5], and Birge, Chen and Qi [2]. However, the model represented by (12) has no immediate relation to the original linear recourse model (4), and, in particular, it is hard to interpret the "meaning" of the matrix H. Moreover, although we may let $\|H\| \to 0$ so that $\phi_H \to \phi$, this results in computational instability in the second-order algorithms used to solve the two-stage recourse problem with recourse ϕ_H.

On the other hand, as we show below, when we solve the quadratic version of the original problem, not only does $\phi_k \to \phi$ as $k \to \infty$, but the algorithms used to solve the problem represented by (8), (9) and (7) are stable as $k \to \infty$.

3 Differentiability of the Quadratic Recourse Function

By the theory of linear programming, $\phi(x, \xi)$, defined by (4), is not in general differentiable with respect to x. This makes it impossible to apply superlinearly convergent methods, such as the Newton method, to solve the stochastic program defined by (1), (4) and (5). Classically, superlinear convergence of a Newton method for solving a nonlinear optimization problem requires that the objective and the constraint functions of the nonlinear optimization problem are twice continuously differentiable. In [14], based upon the superlinear convergence theory for nonsmooth equations [11, 13, 15], Qi developed superlinearly convergent generalized Newton methods for solving a nonlinear optimization problem, whose objective and constraint functions are SC^1, (i.e., they are continuously differentiable and their derivatives are semismooth [12]).

In general, an SC^1 function is not twice differentiable. A nonlinear optimization problem with an SC^1 objective function and linear constraints is called an SC^1 problem [12]. It has been shown that the ELQP is an SC^1 problem [14, 16]. The superlinearly convergent generalized Newton method proposed in [14] was globalized by using a line search and the trust region strategy in [12] and [7] respectively. These methods were applied to the ELQP in [2, 5, 16].

In this section, we will show that the two-stage stochastic program defined by (8), (9) and (7) is also an SC^1 problem. This opens the way to apply the superlinearly and globally convergent generalized Newton method [14, 12, 7, 2] to solve this problem. Before doing this, we briefly review the definition of semismoothness of a vector function and related concepts of generalized Jacobians of vector functions.

Suppose that $F : \Re^n \to \Re^m$ is a Lipschitz vector function. By Rademacher's theorem, F is differentiable almost everywhere. Let D_F be the set where F is differentiable. At any point $x \in \Re^n$, the B-differential [11, 13] of F at x is defined by

$$\partial_B F(x) = \{ \lim_{\substack{y \to x \\ y \in D_F}} \nabla F(y) \},$$

which is a nonempty set. The Clarke subdifferential [6] of F at x is

$$\partial F(x) = \text{conv } \partial_B F(x),$$

which is a nonempty convex set. If for any $h \in \Re^n$,

$$\lim_{\substack{V \in \partial F(x+th) \\ t \downarrow 0}} \{Vh\} \tag{13}$$

exists, then we say that F is semismooth at x. In this case, F is also directionally differentiable at x and $F'(x; h)$ equals the limit in (13). The concept of semismooth functions was introduced for functionals by Mifflin in [10] and extended to vector functions by Qi and Sun in [15].

The following theorem establishes the SC^1 property of the stochastic two-stage problem with quadratic recourse.

Theorem 1 *The stochastic program (8), where Φ_k is defined by (9) and (7), is an SC^1 problem.*

Proof. Since $\phi_k^2 \geq \epsilon_k > 0$, it suffices to prove that ϕ_k^2 is an SC^1 function with respect to x. Define $z \equiv Tx - \xi$ and $\psi_k(z) \equiv \phi_k^2(x, \xi)$. Then rewriting (7) gives

$$\begin{array}{rll} \psi_k(z) = & \min & (q^t y)^2 + k\|Wy - z\|^2 + \epsilon_k \\ & \text{s.t.} & y \geq 0. \end{array} \tag{14}$$

If ψ_k is differentiable with respect to z, then ϕ_k^2 is differentiable with respect to x and

$$\nabla_x \phi_k^2(x, \xi) = T^t \nabla_z \psi_k(z).$$

Let $\bar{z} = (2z^t, 0)^t \in \Re^{r+1}$, $\overline{W} = \binom{kW}{q^t}$, and $u = \overline{W} y \in \Re^{r+1}$. Then $U = \{u = \overline{W} y : y \geq 0\}$ is a polyhedron in $\Re^{r+1}$.

Define $g : \Re^{r+1} \rightarrow \Re \cup \{+\infty\}$ by

$$g(u) = \begin{cases} u^t u, & \text{if } u \in U, \\ +\infty, & \text{otherwise.} \end{cases}$$

Then g is a closed proper extended-valued strongly convex function.

We now can write (14) as

$$\psi_k(z) = kz^t z - \max_{u \in U}\{\bar{z}^t u - u^t u\} + \epsilon_k \tag{15}$$

$$= kz^t z - g^*(2z, 0) + \epsilon_k, \tag{16}$$

where g^* is the conjugate function of g. By Theorem 23.5 of [17], since g is strongly convex, g^* is finite and continuously differentiable everywhere and its derivative at $\bar{z}$ is the unique maximum point in the maximum operation in (15). Furthermore, the derivative of g^* is Lipschitz. Actually, it is not difficult to see that the unique maximum of the maximum operation in (15) is piecewise linear with respect to $\bar{z}$. Hence, the derivative mapping of g^* is semismooth [15]. This shows that ψ_k, hence Φ_k, is an SC^1 function. Therefore, the stochastic programming problem (8), (9) and (7) is an SC^1 optimization problem. □

Now we can apply the generalized Newton (SQP) method proposed in [14, 12, 7] to solve (8). With an adequate nonsingularity condition, this method is superlinearly and globally convergent. In the next section, we will show that $\psi_k(z)$ converges to $\psi(z) \equiv \phi^2(x, \xi)$ as $k \rightarrow \infty$, and give an error bound. This shows that the generalized Newton method is also stable for this problem.

It is noted that, although g^* is convex, ψ_k and Φ_k are not convex in general. In fact, by (15), if W is a nonnegative matrix, then ψ_k is the difference of two convex functions of $\bar{z}$.

If P is continuous, with an argument similar to that in [3] we can show that Φ_k is twice differentiable. Then superlinear convergence can be established for quasi-Newton methods solving (8).

4 Approximation to the Linear Recourse Function

In this section, we show that $\psi_k(z)$ approximates $\psi(z)$ monotonically from below when $k \to \infty$, and give an error bound for this approximation.

Theorem 2 *Suppose that (4) is feasible for $z = Tx - \xi$. Then for any $0 < k \le K$, we have*

$$\sqrt{\phi_k^2(x,\xi) - \epsilon_k} \le \sqrt{\phi_K^2(x,\xi) - \epsilon_K} \le \phi(x,\xi).$$

Proof. Let y_k and y_K be solutions to the minimum operations in the definitions of $\psi_k(z)$ and $\psi_K(z)$ by (14) respectively. Let y^* be any feasible solution of (4). Then we have

$$\begin{aligned}
\psi_k(z) - \epsilon_k &= (q^t y_k)^2 + k\|Wy_k - z\|^2 \\
&\le (q^t y_K)^2 + k\|Wy_K - z\|^2 \text{ by optimality of } y_k \\
&\le (q^t y_K)^2 + K\|Wy_K - z\|^2 \text{ since } k \le K \\
&= \psi_K(z) - \epsilon_K \\
&\le (q^t y^*)^2 + K\|Wy^* - z\|^2 \text{ by optimality of } y_K \\
&= (q^t y^*)^2 \text{ by feasibility of } y^* \\
&= \psi(z). \qquad (17)
\end{aligned}$$

The conclusion of the theorem follows by taking square roots (and noting that $q^t y > 0$). □

Corollary 1 *In Theorem 2, as $k \to \infty$,*

$$\|Wy_k - z\| \to 0,$$

where y_k is the unique solution of the minimum operation in the definition of $\psi_k(z)$.

Proof. By step 2 of the proof of Theorem 2,

$$0 \le k\|Wy_k - z\|^2 \le \psi(z) - (q^t y_k)^2 \le \psi(z) < \infty.$$

The conclusion follows since $\psi(z)$ is finite. □

The next theorem shows that as $k \to \infty$, $\phi_k(x,\xi)$ converges to $\phi(x,\xi)$ and gives an error bound.

Theorem 3 *Suppose that (4) is feasible for $z = Tx - \xi$ and α is the maximum value of 2-norms of optimal dual solutions of (4). Then, $\sqrt{\phi_k^2(x,\xi) - \epsilon_k}$ monotonically converges to $\phi(x,\xi)$ from below and, for k large enough,*

$$0 \le \frac{\phi(x,\xi) - \sqrt{\phi_k^2(x,\xi) - \epsilon_k}}{\phi(x,\xi)} \le \frac{\alpha}{\sqrt{k}}. \qquad (18)$$

Proof. Let y_k and y^* be the same as in Theorem 2. Then, y_k solves

$$\sqrt{\phi_k^2(x,\xi)-\epsilon_k} = \begin{array}{lrcl} \min & q^t y & & \\ \text{s.t.} & Wy & = & Wy_k \\ & y & \geq & 0, \end{array} \tag{19}$$

while y^* solves

$$\phi(x,\xi) = \begin{array}{lrcl} \min & q^t y & & \\ \text{s.t.} & Wy & = & z \\ & y & \geq & 0. \end{array} \tag{20}$$

By Corollary 1, (19) can be regarded as a perturbation of (20) with a small change of $Wy_k - z$ on the right hand side of the constraint of the linear program (20). By the perturbation theory of linear programming, we have

$$|\sqrt{\phi_k^2(x,\xi)-\epsilon_k} - \phi(x,\xi)| \leq \alpha\|Wy_k - z\|.$$

By the proof of Corollary 1,

$$\|Wy_k - z\| \leq \frac{q^t y^*}{\sqrt{k}}.$$

By Theorem 2,

$$\sqrt{\phi_k^2(x,\xi)-\epsilon_k} \leq \phi(x,\xi).$$

Combining these three inequalities, we have (18). The last conclusion follows. □

Corollary 2 *If $\epsilon_k \to 0$, then $\phi_k(x,\xi)$ converges to $\phi(x,\xi)$.*

Proof. The conclusion follows mfrom (18) and Theorem 2. □

Example. To see how these bounds appear in practice, consider the following simple example (where $\epsilon_k = 0$ for illustration).

$$\begin{array}{lrcl} \phi(x,\xi) = \min & y_1 + y_2 + y_3 & & \\ \text{s.t.} & y_1 + y_3 & = & z_1 \\ & y_2 + y_3 & = & z_2 \\ & y & \geq & 0. \end{array} \tag{21}$$

Figures 1 and 2 show convergence of ϕ_k for $z = (1, 0.5)$ and $z = (1, 1)$. Note that the convergence is somewhat faster for $z = (1, 1)$. In this case, the dual solution to (21) is not unique and has the same maximum norm, $\alpha = 1$, as for $z = (1, 0.5)$. The convergence behavior for $z = (1, 1)$ may benefit from a smaller norm for the average of the norms of dual solutions.

Theorem 3 also gives us the condition for the optimal solutions of (7) to converge to an optimal solution of (2). We use the following two results from the theory of *epi-convergence* ([1]). A sequences of functions, $\{f^\nu\}$, *epi-converges* to f, if the epi-graphs of f^ν, converge as sets to the epi-graph of f.

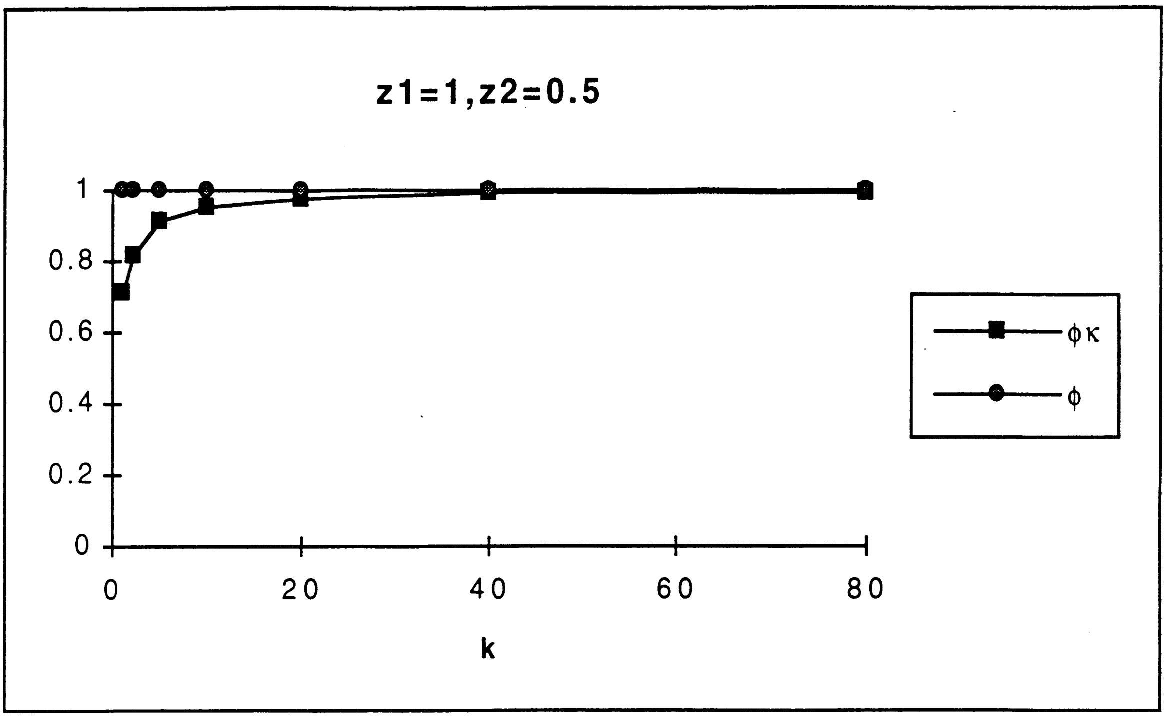

Figure 1: Convergence of ϕ_k to ϕ for $z = (1, 0.5)$.

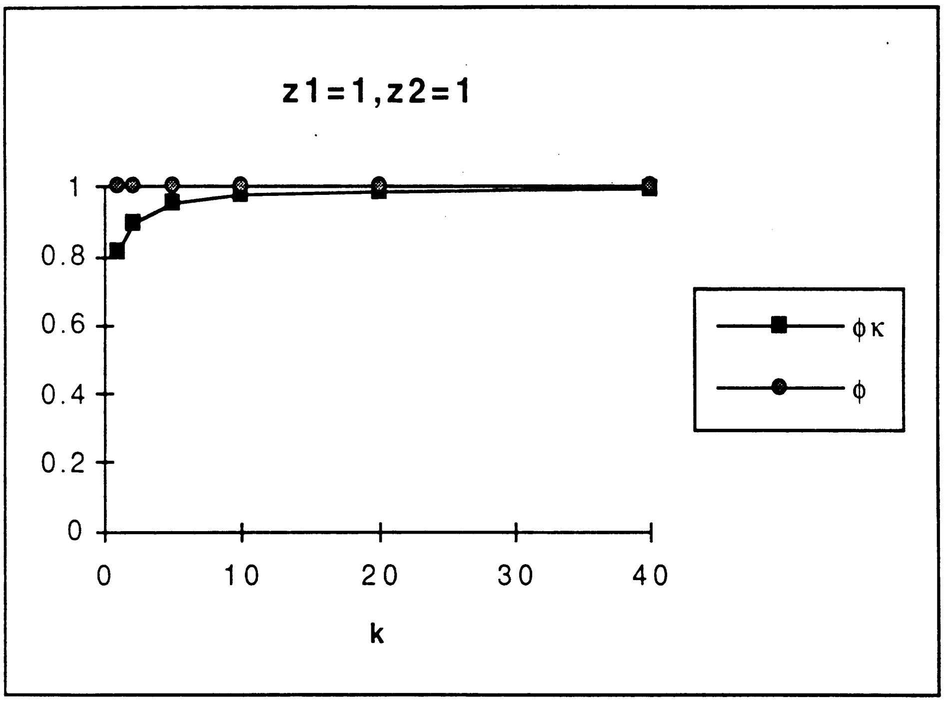

Figure 2: Convergence of ϕ_k to ϕ for $z = (1, 1)$.

Theorem 4 *([4], Theorem 2.3.) Suppose a sequence of functions,* $\{g^\nu, \nu = 1, \ldots,\}$, *epi-converges to* g, *then* $\limsup_{\nu\to\infty}(\inf g^\nu) \leq \inf g$, *and, if* $x^k \in \text{argming}^{\nu_k}$ *for some subsequence,* $\{g^\nu\}$, *of functions from* $\{g^{\nu_k}\}$, *and* $x = \lim_k x^k$, *then*

$$x \in \text{argming}, \text{ and } \lim_k(\inf g^{\nu_k}) = \inf g. \tag{22}$$

The next result shows that the sequence $\{\phi_k\}$ converges to ϕ.

Theorem 5 *([1], Proposition 3.12). Suppose a sequence of functions,* $\{g^\nu, \nu = 1, \ldots,\}$, *pointwise converge everywhere to* g, *i.e.,* $\lim_\nu g^\nu(x) = g(x)$. *If the* g^ν *are monotone increasing, or monotone decreasing, and* g *is lower semi-continuous, then* g^ν *epi-converges to* g.

The following theorem follows immediately from these observations.

Theorem 6 *([4], Theorem 2.7.) Suppose* $\{g, g^\nu, \nu = 1, \ldots\}$ *form a collection of functions defined on* $\Re^n \times \Xi$, *with value in* $\Re \cup \infty$, *for all* x, $\xi \to g(x,\xi)$ *is measurable, if* $P[\xi | g(x,\xi)) < \infty] = 1$, *then* $\int_\Xi g(x,\xi)P(d\xi) < \infty$, *and* $g^\nu(x,\xi)$ *epi-converges to* $g(x,\xi)$ *for all all* $\xi \in \Xi$, *and all* g^ν *are absolutely bounded by integrable functions, then the expectation functionals,* $\int_\Xi g^\nu(x,\xi)P(d\xi) < \infty$, *epi- and pointwise converge to* $\int_\Xi g(x,\xi)P(d\xi) < \infty$.

Now, we can state our result for convergence of the optimal solutions.

Theorem 7 *Suppose that* ξ *has finite second moments, let* $\{x^k, \pi^k\}$ *be a sequence of optimal primal-dual solution pairs to (8), where the multipliers* π^k *are associated with the linear constraints.* $Ax = b$. *If* $\{x^k, \pi^k\}$ *has a limit point,* (x^*, π^*), *with* $\epsilon_k \to 0$, *then* (x^*, π^*) *is an optimal primal-dual solution pair of (1).*

Proof. If ξ has second moments, then it can be shown easily that ϕ is absolutely integrable (see, for example, ([8])) and that, if $P[\xi|\phi(x,\xi)) < \infty] = 1$, then $\int_\Xi \phi(x,\xi)P(d\xi) < \infty$. The extensions of these results to ϕ_k follow, for example, from Theorem 3.

To invoke Theorem 6, let $\tilde{\phi}_k(x,\xi) = \sqrt{\phi_k^2(x,\xi) - \epsilon_k}$ and define $g^k(x,\xi) = \phi_k(x,\xi) + \delta_{\{Ax=b, x\geq 0\}}(x)$, where the indicator function, $\delta_S(x) = 0$, if $x \in S$, and $+\infty$ if $x \notin S$. Using the bound in Theorem 3 and the result from Theorem 5, we then meet the conditions for Theorem 6. If we replace f_k in (8) with $\tilde{f}_k$, where ϕ_k is replaced by $\tilde{\phi}_k$, then the expectation functionals $\tilde{f}_k$ converge. Using Theorem 4, we have that limit points of minimizers of $\tilde{f}_k$ are optimal in (1).

Next, we must relate $\tilde{f}_k$ to f_k. Note that the optimality conditions for $\tilde{f}_k$ yield $\tilde{x}_k$ and $\tilde{\pi}_k$ such that

$$c + \nabla\tilde{\Phi}_k(\tilde{x}_k) - \tilde{\pi}_k^t A = \tilde{\rho}_k \geq 0, \tilde{\rho}_k^t \tilde{x}_k = 0. \tag{23}$$

For f_k, the optimality conditions are:

$$c + \nabla\Phi_k(x_k) - \pi_k^t A = \rho_k \geq 0, \rho_k^t x_k = 0. \tag{24}$$

Note that for any x such that ϕ_k is finite for all ξ, we have $\nabla \phi_k$ continuous and, thus,

$$\begin{aligned} \nabla \Phi_k(x) &= \int \nabla \phi_k(x,\xi) P(d\xi) \\ &= \int \frac{\tilde{\phi}_k(x,\xi)}{\phi_k(x,\xi)} \nabla \tilde{\phi}_k(x,\xi) P(d\xi) \\ &= \Delta_k \nabla \tilde{\Phi}_k(x), \end{aligned} \tag{25}$$

where $\Delta_k \to 1$ as $k \to \infty$. For any limit point, $(\bar{x}, \bar{\pi})$ of (x_k, π_k), we must have, from (24):

$$c + \nabla \tilde{\Phi}(\bar{x}) - \bar{\pi}^t A = \bar{\rho} \geq 0, \bar{\rho}^t \bar{x} = 0, \tag{26}$$

where $\tilde{\Phi}(x) = \lim_k \tilde{\Phi}_k(x)$. Thus $(\bar{x}, \bar{\pi})$ forms an optimal primal-dual pair in (8) with $\tilde{\Phi}$ replacing Φ_k, and $(\bar{x}, \bar{\pi})$ must be optimal in (1) by the argument above. □

5 Conclusions

We present a quadratic approximation of the stochastic program with linear recourse. The approximation uses a linear-quadratic loss function similar to that used in decision theory. It also allows the use of superlinearly convergent methods that cannot be directly to the linear recourse problem. We showed that the approximation converges and that solutions of the approximation also converge to those of the original problem.

Our interest in this form of quadratic approximation extends beyond an advantage for fast computational methods. We believe that the approach may also enable further approximations using known moments of the relevant random variables.

References

[1] H. Attouch and R. J-B Wets, Approximation and convergence in nonlinear optimization, in: O.L. Mangasarian, R.R. Meyer and S.M. Robinson, **Nonlinear programming, 4**, Academic Press, New York-London, 1981.

[2] J.R. Birge, X. Chen and L. Qi, A stochastic Newton method for solving stochastic quadratic programs with recourse, AMR 94/9, Applied Mathematics Report, University of New South Wales, 1994.

[3] J.R. Birge and L. Qi, Subdifferentials in approximation for stochastic programming, *SIAM Journal on Optimization* 5 (1995) 436-453.

[4] J.R. Birge and R. J–B. Wets, " Designing approximation schemes for stochastic optimization problems, in particular, for stochastic programs with recourse," *Mathematical Programming Study* 27 (1986) 54-102.

[5] X. Chen, L. Qi, and R. Womersley, Newton's method for quadratic stochastic programs with recourse, *Journal of Computational and Applied Mathematics* forthcoming.

[6] F.H. Clarke, *Optimization and Nonsmooth Analysis* Wiley, New York, 1983.

[7] H. Jiang and L. Qi, A globally and superlinearly convergent trust region algorithm for convex SC^1 minimization problems and its application to stochastic programs, AMR 94/7, Applied Mathematics Report, University of New South Wales, 1994.

[8] P. Kall, *Stochastic Linear Programming*, Springer Verlag, Berlin, 1976.

[9] P. Kall and S.T. Wallace, *Stochastic Programming*, Wiley, New York, 1994.

[10] R. Mifflin, Semismooth and semiconvex functions in constrained optimization, *SIAM Journal on Control and Optimization* 15 (1977) 957-972.

[11] J.-S. Pang and L. Qi, Nonsmooth equations: Motivation and algorithms, *SIAM Journal on Optimization* 3 (1993) 443–465.

[12] J.-S. Pang and L. Qi, A globally convergent Newton method for convex SC^1 minimization problems, *Journal of Optimization Theory and Applications* forthcoming.

[13] L. Qi, Convergence analysis of some algorithms for solving nonsmooth equations, *Mathematics of Operations Research* 18 (1993) 227–244.

[14] L. Qi, Superlinearly convergent approximate Newton methods for LC^1 optimization problems, *Mathematical Programming* 64 (1994) 277–294.

[15] L. Qi and J. Sun, A nonsmooth version of Newton's method, *Mathematical Programming* 58 (1993) 353–367.

[16] L. Qi and R.S. Womersley, An SQP algorithm for extended linear problems in stochastic programming, *Annals of Operations Research* forthcoming.

[17] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, N.J., 1970.

[18] R.T. Rockafellar and R.J.-B. Wets, "A Lagrangian finite-generation technique for solving linear-quadratic problems in stochastic programming", *Mathematical Programming Study* 28 (1986) 63-93.