

Decomposition-Based Interior Point Methods for Two-Stage Stochastic Convex Quadratic Programs with Recourse ^{*}

Sanjay Mehrotra

M. Gokhan Ozevin

July, 2003 (revised January 2004)

*Dept. of Industrial Engineering and Management Sciences
Northwestern University, Evanston, IL
(mehrotra@northwestern.edu, ozevin@northwestern.edu).*

Abstract

Zhao [28] recently showed that the log barrier associated with the recourse function of two-stage stochastic linear programs behaves as a strongly self-concordant barrier and forms a self concordant family on the first stage solutions. In this paper we show that the recourse function is also strongly self-concordant and forms a self concordant family for the two-stage stochastic convex quadratic programs with recourse. This allows us to develop Benders decomposition based linearly convergent interior point algorithms. An analysis of such an algorithm is given in this paper.

^{*}This research was partially supported by NSF grant DMI-0200151, and ONR grant N0014-01-1-0048/P00002

1. Introduction

We study the two-stage stochastic convex quadratic programs (TSQP) with recourse:

$$\begin{aligned} \min \quad & \eta(x) := \frac{1}{2}x^T Gx + c^T x + \rho(x) \\ \text{s.t.} \quad & Ax = b, \\ & x \geq 0, \end{aligned} \tag{1.1}$$

where

$$\rho(x) := E\{\rho(x, \tilde{\xi})\} \tag{1.2}$$

and

$$\begin{aligned} \rho(x, \xi) := \min \quad & \frac{1}{2}y(\xi)^T H(\xi)y(\xi) + d(\xi)^T y(\xi) \\ \text{s.t.} \quad & W(\xi)y(\xi) = h(\xi) - T(\xi)x, \\ & y(\xi) \geq 0. \end{aligned} \tag{1.3}$$

Here E represents the expectation with respect to $\tilde{\xi}$ and Ξ is the support of $\tilde{\xi}$. For each realization ξ of $\tilde{\xi}$; $H(\xi) \in \mathbb{R}^{m \times m}$ is symmetric positive semidefinite, $T(\xi) \in \mathbb{R}^{l \times n}$, $W(\xi) \in \mathbb{R}^{l \times m}$ and $y(\xi) \in \mathbb{R}^m$. $G \in \mathbb{R}^{n \times n}$ is symmetric positive semidefinite, $A \in \mathbb{R}^{s \times n}$, and $b \in \mathbb{R}^s$.

The TSQP was introduced and studied by Rockafellar and Wets [20, 21]. They gave a Lagrangian based approach for solving the case where objectives are strictly convex functions. Rockafellar [18] and Rockafellar and Wets [22] studied applications of this model in deterministic and stochastic optimal control in discrete time. A modified proximal point algorithm was given by Zhu [29], and a primal-dual projected gradient algorithms were given by Zhu and Rockafellar [30]. A finite envelop method was given in Rockafellar [19]. Chen, Qi and Womersley [7] gave a modified Newton based method for solving this problem. An inexact Newton method combined with stochastic decomposition was developed by Birge, *et al.* [3] for the case where Ξ is continuous.

In this paper we develop Benders decomposition based interior point methods for TSQP with finite support. This is accomplished by showing that the log-barrier recourse function is a strongly self-concordant function for this problem. In general, the recourse function $\rho(x)$ is not differentiable everywhere, however, the log-barrier recourse function becomes differentiable. The traditional decomposition algorithms either use the nonsmooth optimization techniques [1, 4, 27], or use alternative techniques to smooth this function [20, 23]. Given the recent success of interior point methods, it is interesting to see if decomposition based interior point algorithms are possible for stochastic programming problems. Recently, Zhao [28] developed such an algorithm for the linear case. Zhao [28] showed that the log barrier associated with the recourse function of two-stage stochastic linear programs behaves as a strongly self-concordant barrier (Nesterov and Nemirovskii [16]) on the first stage solutions.

Zhao's results are extended here for the quadratic case to develop algorithms for TSQP.

In this paper we focus on problems where Ξ is discrete and finite. When the support Ξ is infinite or finite but very large, sample-average approximation methods can be used [12, 24]. Here a random sample ξ_1, \dots, ξ_K of the random vector $\tilde{\xi}$ is used to approximate the expectation $E\{\rho(x, \tilde{\xi})\}$ by $\frac{1}{K} \sum_{i=1}^K \rho(x, \xi_i)$. One then solves the resulting approximate problem and uses its optimal solution as an approximation to the optimal solution of the original problem. Shapiro and Homem-De-Mello [25] showed that when the underlying probability distribution is discrete, under mild conditions such an approximation yields an exact solution to the original problem with sufficiently large sample size. Linderoth, Shapiro and Wright [13] investigated the quality of solutions obtained from sample-average approximations to two-stage linear programs with recourse.

We note that TSQP with finite scenarios can be explicitly formulated as a large-scale QP, which can be solved by directly. In particular, we can use primal-dual interior point methods exploiting its special structure through efficient matrix factorization schemes [5, 6, 8]. However, in order to apply primal-dual methods we need to know upfront, and it is not clear how scenarios can be included adaptively. In contrast, the decomposition method of this paper handles the variables x and $y(\cdot)$ separately, consequently it allows the possibility of including scenarios adaptively. In particular, the emphasis is on obtaining sufficient decrease in the barrier function which may be possible through inexact computations of Newton direction. The gradient and Hessian needed to compute the Newton direction is built from the solutions of second stage centering problems, and its computation decomposes. If information from only a subset of scenarios is used inexact gradients and Hessians are calculated. This may have computational advantages in the single and multi-processor computational environments. In the single processor environment, it may allow for less computations in the early stage of interior point algorithm, particularly when the total number of scenarios is very large. In the multi-processor, and particularly distributed computing environment, where some of the computational nodes may not be reliable it has the advantage that the algorithm need not depend on completely finishing computations with all the scenarios. Furthermore, decomposition may allow use of information from one scenario to save computational efforts at other scenarios. Since a careful computational study requires significant additional research effort, we intend to explore it in the future.

This paper is organized as follows. In Section 2 we state our notation, the problem formulation and our assumptions. In Section 3 we show that $\eta(\mu, \cdot)$ (defined in Section 2) is a self-concordant family. In Section 4 we present our algorithm and give its convergence theorems. The proofs of convergence theorems are given in Section 5

We use the following notation: For any strictly positive vector x in \mathbb{R}^n , we define $\ln x := (\ln x_1, \dots, \ln x_n)^T$, $x^{-1} := (x_1^{-1}, \dots, x_n^{-1})^T$. $X := \text{diag}(x_1, \dots, x_n)$ denote the $n \times n$ diagonal matrix whose diagonal entries are x_1, \dots, x_n . A vector with all entries equal to one is denoted

by e ; its dimension will be clear from the context. We also define $\mathbb{R}_+^n := \{x \in \mathbb{R}^n \mid x \geq 0\}$ and $\mathbb{R}_{++}^n := \{x \in \mathbb{R}^n \mid x > 0\}$. The notation $M \preceq N$ means that $N - M$ is a symmetric positive semidefinite matrix. Throughout the paper we will use “ ∇ ”, “ ∇^2 ”, “ ∇^3 ” to denote the gradient, Hessian and the third order derivative with respect to x and a “ $'$ ” for the derivative with respect to variables other than x . “ ∇ ” is also used to denote the Jacobian of a vector function. For example,

$$[\{\nabla^2 f(\mu, x)\}']_{i,j} = \frac{\partial}{\partial \mu} \left(\frac{\partial^2 f(\mu, x)}{\partial x_i \partial x_j} \right).$$

Let the random variable $\tilde{\xi}$ have a finite discrete support $\Xi = \{\xi_1, \dots, \xi_K\}$ with probabilities $\{\pi_1, \dots, \pi_K\}$. For simplicity of notation we define $\rho_i(x) := \rho(x, \xi_i)$, $H_i := \pi_i H(\xi_i)$, $T_i := T(\xi_i)$, $W_i := W(\xi_i)$, $h_i := h(\xi_i)$, $y_i := y(\xi_i)$ and $d_i := \pi_i d(\xi_i)$.

2. Problem Formulation and Assumptions

The problem (1.1-1.3) is rewritten as

$$\begin{aligned} \min \quad & \eta(x) := \frac{1}{2}x^T Gx + c^T x + \rho(x) \\ \text{s.t.} \quad & Ax = b, \\ & x \geq 0, \end{aligned} \tag{2.1}$$

where

$$\rho(x) := \sum_{i=1}^K \rho_i(x) \tag{2.2}$$

and for $i = 1, \dots, K$

$$\begin{aligned} \rho_i(x) := \min \quad & \frac{1}{2}y_i^T H_i y_i + d_i^T y_i \\ \text{s.t.} \quad & W_i y_i = h_i - T_i x, \\ & y_i \geq 0. \end{aligned} \tag{2.3}$$

The duals to subproblems (2.3) are:

$$\begin{aligned} \max \quad & -\frac{1}{2}y_i^T H_i y_i + (h_i - T_i x)^T z_i \\ \text{s.t.} \quad & W_i^T z_i - H_i y_i + s_i = d_i, \\ & s_i \geq 0. \end{aligned} \tag{2.4}$$

Let us define the following feasibility sets:

$$\begin{aligned}
\mathcal{F}_i^p(x) &:= \{y_i | W_i y_i = h_i - T_i x, y_i > 0\}, \\
\mathcal{F}_i^1 &:= \{x | \mathcal{F}_i^p(x) \neq \emptyset\}, \\
\mathcal{F}^1 &:= \bigcap_{i=1}^K \mathcal{F}_i^1, \\
\mathcal{F}^0 &:= \mathcal{F}^1 \cap \{x | Ax = b, x > 0\}, \\
\mathcal{F} &:= \{(x, \lambda) \times (y_1, z_1, \dots, y_K, z_K) | Ax = b, x > 0; W_i y_i = h_i - T_i x, y_i > 0, \\
&\quad W_i^T z_i - H_i y_i < d_i, i = 1, \dots, K; -Gx + A^T \lambda + \sum_{i=1}^K T_i^T z_i < c\}.
\end{aligned}$$

Consider the following log-barrier decomposition problem:

$$\begin{aligned}
\min \quad \eta(\mu, x) &:= \frac{1}{2} x^T G x + c^T x - \mu e^T \ln x + \rho(\mu, x) \\
\text{s.t.} \quad &Ax = b, \\
&x > 0,
\end{aligned} \tag{2.5}$$

where

$$\rho(\mu, x) := \sum_{i=1}^K \rho_i(\mu, x) \tag{2.6}$$

and, for $i = 1, \dots, K$,

$$\begin{aligned}
\rho_i(\mu, x) &:= \min \quad \frac{1}{2} y_i^T H_i y_i + d_i^T y_i - \mu e^T \ln y_i \\
\text{s.t.} \quad &W_i y_i = h_i - T_i x, \\
&y_i > 0.
\end{aligned} \tag{2.7}$$

The barrier problem associated with the dual of (2.3) is defined as follows:

$$\begin{aligned}
\max \quad &-\frac{1}{2} y_i^T H_i y_i + (h_i - T_i x)^T z_i + \mu e^T \ln s_i \\
\text{s.t.} \quad &W_i^T z_i - H_i y_i + s_i = d_i, \\
&s_i > 0.
\end{aligned} \tag{2.8}$$

Here we omit a constant term $m\mu(1 - \ln \mu)$. Since the problems (2.7) and (2.8) are respectively convex and concave, (y_i) and (y_i, z_i, s_i) are optimal solutions to (2.7) and (2.8), respectively, if and only if they satisfy the following optimality conditions:

$$\begin{aligned}
Y_i s_i &= \mu e, \\
W_i y_i &= h_i - T_i x, \\
W_i^T z_i - H_i y_i + s_i &= d_i, \\
y_i > 0, \quad s_i > 0.
\end{aligned} \tag{2.9}$$

We make the following assumptions:

A1 Every matrix W_i has full row rank.

A2 $\mathcal{F} \neq \emptyset$.

Assumption A1 is for convenience and Assumption A2 can be ensured by introducing artificial variables. Note that for a given $\mu > 0$, the log-barrier recourse function $\rho(\mu, x) < \infty$ iff $x \in \mathcal{F}^1$. Hence \mathcal{F}^0 is the implicit feasible set of problem (2.5). Throughout the paper we denote the optimal solution to the first stage problem (2.5) by $x(\mu)$ and the solutions to the optimality conditions (2.9) for a given $x \in \mathcal{F}^1$ by $(y_i(\mu, x), z_i(\mu, x), s_i(\mu, x))$.

Assumption A2 requires that the problem (2.10) and its dual have strictly feasible solutions, which in turn implies that problems (2.5) and (2.10) below have unique optimal solutions. Assumption A2 also implies that for every $\mu > 0$ and $x \in \mathcal{F}^1$ the optimality conditions (2.9) have unique solutions. Thus the first stage primal central trajectory $\{x(\mu), \mu > 0\}$ and the second stage primal-dual central trajectories $\{y_i(\mu, x), z_i(\mu, x), s_i(\mu, x), \mu > 0\}$ for any $x \in \mathcal{F}^1$ are well defined.

The optimal solutions of (2.5-2.7) and those of the explicit log-barrier problem:

$$\begin{aligned}
\min \quad & \frac{1}{2}x^T Gx + c^T x + \sum_{i=1}^K \left[\frac{1}{2}y_i^T H_i y_i + d_i^T y_i \right] - \mu \ln x - \mu e^T \ln y_i \\
\text{s.t.} \quad & Ax = b, \\
& W_i y_i = h_i - T_i x, \quad i = 1, \dots, K, \\
& x > 0, \quad y_i > 0, \quad i = 1, \dots, K,
\end{aligned} \tag{2.10}$$

associated with the explicit deterministic equivalent formulation of (2.1-2.3) have the following relationship. For a given $\mu > 0$, if $(x(\mu)^*, y_1(\mu)^*, \dots, y_K(\mu)^*)$ is the optimal solution to (2.10), then $x(\mu)^*$ is the optimal solution to (2.5), and $(y_1(\mu)^*, \dots, y_K(\mu)^*)$ are the optimal solutions to subproblems (2.7) for given μ and $x = x(\mu)^*$. Conversely, if for a given μ , $x(\mu)^*$ is the optimal solution to (2.5) and $(y_1(\mu)^*, \dots, y_K(\mu)^*)$ are the optimal solutions to (2.7) with $x = x(\mu)^*$, then $(x(\mu)^*, y_1(\mu)^*, \dots, y_K(\mu)^*)$ is the optimal solution to (2.10).

3. The Self-Concordance Property of the Family $\{\eta(\mu, \cdot) : \mu > 0\}$

3.1 Computation of $\nabla \eta(\mu, x)$ and $\nabla^2 \eta(\mu, x)$

From (2.9) we can show that $\rho_i(\mu, x)$ is equal to the optimal objective of the dual subproblem (2.8), i.e.:

$$\rho_i(\mu, x) = -\frac{1}{2}y_i(\mu, x)^T H_i y_i(\mu, x) + (h_i - T_i x)^T z_i(\mu, x) + \mu e^T \ln s_i(\mu, x) + m\mu(1 - \ln \mu). \tag{3.1}$$

Differentiating (3.1) and using the optimality conditions (2.9) and (3.6) we can verify that

$$\nabla \rho_i(\mu, x) = -T_i^T z_i(\mu, x). \quad (3.2)$$

Hence,

$$\nabla \eta(\mu, x) = Gx + c - \mu x^{-1} - \sum_{i=1}^K T_i^T z_i(\mu, x), \quad (3.3)$$

$$\nabla^2 \eta(\mu, x) = G + \mu X^{-2} - \sum_{i=1}^K T_i^T \nabla z_i(\mu, x). \quad (3.4)$$

In order to compute $\nabla^2 \eta(\mu, x)$ in (3.4) we need to determine the derivative of $z_i(\mu, x)$ with respect to x . Let $(y_i, z_i, s_i) := (y_i(\mu, x), z_i(\mu, x), s_i(\mu, x))$. Differentiating (2.9) with respect to x we get

$$\begin{aligned} Y_i \nabla s_i + S_i \nabla y_i &= 0, \\ W_i \nabla y_i &= -T_i, \\ W_i^T \nabla z_i - H_i \nabla y_i + \nabla s_i &= 0, \end{aligned} \quad (3.5)$$

where $\nabla s_i, \nabla y_i$ and ∇z_i are Jacobian matrices.

Solving the system (3.5) we obtain

$$\begin{aligned} \nabla z_i &= -R_i^{-1} T_i, \\ \nabla y_i &= -Q_i^2 W_i^T R_i^{-1} T_i, \\ \nabla s_i &= (I - H_i Q_i^2) W_i^T R_i^{-1} T_i, \end{aligned} \quad (3.6)$$

where

$$Q_i := (H_i + Y_i^{-1} S_i)^{-1/2} \quad \text{and} \quad R_i := W_i Q_i^2 W_i^T. \quad (3.7)$$

Then substituting for ∇z_i in (3.4) we get

$$\nabla^2 \eta(\mu, x) = G + \mu X^{-2} + \sum_{i=1}^K T_i^T R_i^{-1} T_i. \quad (3.8)$$

3.2 Self-Concordance of the Recourse Function

Definition 3.1 (Nesterov and Nemirovskii [16]) *Let \mathcal{E} be a finite-dimensional real vector space, \mathcal{Q} be an open nonempty convex subset of \mathcal{E} , $f : \mathcal{Q} \rightarrow \mathbb{R}$ be a function, $\alpha > 0$. f is called α -self-concordant on \mathcal{Q} with the parameter value α , if $f \in C^3$ is a convex function on \mathcal{Q} , and, for all $x \in \mathcal{Q}$ and $h \in \mathcal{E}$ the following inequality holds:*

$$|\nabla^3 f(x)[h, h, h]| \leq 2\alpha^{-1/2} (\nabla^2 f(x)[h, h])^{3/2}.$$

An α -self-concordant on \mathcal{Q} function f is called strongly α -self-concordant on \mathcal{Q} if $f(x_i)$ tends to infinity along every sequence $\{x_i \in \mathcal{Q}\}$ converging to a boundary point of \mathcal{Q} .

We now show that the recourse function $\rho(\mu, x)$ behaves as a strongly self-concordant barrier on \mathcal{F}^1 .

The following lemma can be directly shown using Proposition 5.1.5 in Nesterov and Nemirovskii [16] as noted by Zhao [28, Lemma 1] in the context of linear two-stage stochastic programs. We prefer to present a more explicit proof for the current setting as it uses more elementary arguments.

Lemma 3.1 *For any $\mu > 0$, $\rho_i(\mu, \cdot)$ is strongly μ -self-concordant on $\mathcal{F}_i^1, i = 1, \dots, K$.*

Proof. For any $\mu > 0$, $x \in \mathcal{F}_i^1$, and $h \in \mathbb{R}^n$ we define the univariate function

$$\Phi_i(t) := \nabla^2 \rho_i(\mu, x + th)[h, h].$$

Note that $\Phi_i'(0) = \nabla^3 \rho(\mu, x)[h, h, h]$. Along every sequence $\{x_j \in \mathcal{F}_i^1\}$ converging to the boundary of \mathcal{F}_i^1 , $\rho_i(\mu, x_j)$ tends to infinity. To prove the lemma it suffices to show that

$$|\Phi_i'(0)| \leq \frac{2}{\sqrt{\mu}} |\Phi_i(0)|^{3/2}.$$

Let $(y_i(t), z_i(t), s_i(t)) := (y_i(\mu, x + th), z_i(\mu, x + th), s_i(\mu, x + th))$. For convenience we define

$$Q_i(t) := (H_i + Y_i(t)^{-1} S_i(t))^{-1/2} \quad \text{and} \quad R_i(t) := W_i Q_i(t)^2 W_i^T. \quad (3.9)$$

From (3.8) we have

$$\Phi_i(t) := h^T [T_i^T R_i(t)^{-1} T_i] h.$$

Differentiating $\Phi(t)$ with respect to t , we get

$$\begin{aligned} |\Phi_i'(t)| &= |h^T [-T_i^T R_i(t)^{-1} R_i'(t) R_i(t)^{-1} T_i] h| \\ &\quad \text{(since } \{R_i(t)^{-1}\}' = -R_i(t)^{-1} R_i'(t) R_i(t)^{-1}\text{)} \\ &= |h^T T_i^T R_i(t)^{-1} W_i Q_i(t)^2 \{Q_i(t)^{-2}\}' Q_i(t)^2 W_i^T R_i(t)^{-1} T_i h| \\ &\quad \text{(using the definition of } R_i\text{)} \\ &= \frac{2}{\sqrt{\mu}} |h^T T_i^T R_i(t)^{-1} W_i Q_i(t)^2 \{Q_i(t)^{-2}\}' Q_i(t)^2 W_i^T R_i(t)^{-1} T_i h|. \end{aligned} \quad (3.10)$$

Now we bound the term in the right-hand-side of (3.10).

Since $\{Q_i(t)^{-2}\}'$ is a diagonal matrix, for any $d \in \mathbb{R}^m$ we have

$$\begin{aligned} |d^T Q_i(t) \{Q_i(t)^{-2}\}' Q_i(t) d| &\leq \|Y_i(t)^2 \{Q_i(t)^{-2}\}'\|_2 d^T Q_i(t) Y_i(t)^{-2} Q_i(t) d \\ &= \|Y_i(t)^2 \{Q_i(t)^{-2}\}' e\|_\infty d^T Q_i(t) Y_i(t)^{-2} Q_i(t) d \\ &\leq \mu^{-1} \|Y_i(t)^2 \{Q_i(t)^{-2}\}' e\|_\infty d^T d \\ &\quad \text{(since } Q_i(t) Y_i(t)^{-2} Q_i(t) \preceq \mu^{-1} I\text{)}. \end{aligned} \quad (3.11)$$

Using (3.6) we can write:

$$\begin{aligned} y_i(t)' &= \nabla y_i(\mu, x + th)^T h = -Q_i(t)^2 W_i^T R_i(t)^{-1} T_i h, \\ s_i(t)' &= \nabla s_i(\mu, x + th)^T h = (I - H_i Q_i(t)^2) W_i^T R_i(t)^{-1} T_i h. \end{aligned} \quad (3.12)$$

Using (3.12) we have

$$\begin{aligned} \|Y_i(t)^2 \{Q_i(t)^{-2}\}' e\|_\infty &= \|Y_i(t) s_i(t)' - S_i(t) y_i(t)'\|_\infty \\ &= \|Y_i(t) (I - H_i Q_i(t)^2) W_i^T R_i(t)^{-1} T_i h + S_i(t) Q_i(t)^2 W_i^T R_i(t)^{-1} T_i h\|_\infty \\ &= \|[Y_i(t) Q_i^{-1} + (S_i(t) - Y_i(t) H_i) Q_i(t)] Q_i(t) W_i^T R_i(t)^{-1} T_i h\|_\infty \\ &= 2 \|S_i(t) Q_i^2(t) W_i^T R_i(t)^{-1} T_i h\|_\infty \\ &\leq 2 [h^T T_i^T R_i(t)^{-1} W_i Q_i^2(t) S_i^2(t) Q_i^2(t) W_i^T R_i(t)^{-1} T_i h]^{1/2} \\ &= 2\sqrt{\mu} [h^T T_i^T R_i(t)^{-1} T_i h]^{1/2} \\ &\quad (\text{observing that } Q_i(t) S_i^2(t) Q_i(t) \preceq \mu I \text{ and using the definition of } R_i) \\ &= 2\sqrt{\mu} \Phi_i(t)^{1/2}. \end{aligned} \quad (3.13)$$

Combining (3.11) and (3.13) it follows that for any $d \in \mathbb{R}^m$

$$|d^T Q_i(t) \{Q_i(t)^{-2}\}' Q_i(t) d| \leq \frac{2}{\sqrt{\mu}} \Phi_i(t)^{1/2} d^T d. \quad (3.14)$$

Finally, for setting $t = 0$ and by taking $d = Q_i(0) W_i^T R_i(0)^{-1} T_i h$, (3.10) and (3.14) imply

$$\begin{aligned} |\Phi_i'(0)| &= |h^T T_i^T R_i(0)^{-1} W_i Q_i(0)^2 \{Q_i(0)^{-2}\}' Q_i(0)^2 W_i^T R_i(0)^{-1} T_i h| \\ &\leq \frac{2}{\sqrt{\mu}} \Phi_i(0)^{1/2} h^T T_i^T R_i(0)^{-1} T_i h. \\ &= \frac{2}{\sqrt{\mu}} \Phi_i(0)^{3/2}. \end{aligned} \quad (3.15)$$

Hence the proof is complete ■

We have the following immediate corollary.

Corollary 3.1 *The recourse function $\rho(\mu, x)$ is as a strongly μ -self-concordant barrier on \mathcal{F}^1 and the first stage objective function, $\eta(\mu, x) := \frac{1}{2} x^T G x + c^T x - \mu e^T \ln x + \rho(\mu, x)$, is strongly μ -self-concordant on $\mathcal{F}^1 \cap R_{++}^n$.*

Proof. It is easy to verify that $\mu e^T \ln x$ is strongly μ -self-concordant on R_{++}^n . The corollary follows from Proposition 2.1.1 (ii) in [16] ■

3.3 Parameters of the Self-Concordance Family

The self-concordant family with appropriate parameters is defined in Nesterov and Nemirovskii [16]. They showed that given such a family, the parameters defining the family allow us to relate the rate at which the barrier parameter μ is varied and the number of Newton steps required to maintain the proximity to the central path. Below is the definition of a strongly self-concordant family adapted to the current setting from the original definition in Nesterov and Nemirovskii [16]. These conditions might look rather technical; nevertheless they simplify our convergence analysis and the accompanying proofs in the sequel explicitly reveal some essential properties of the log-barrier recourse function $\rho(\mu, x)$.

Definition 3.2 *The family of functions $\{\eta(\mu, \cdot) : \mu > 0\}$ is strongly self-concordant on \mathcal{F}^0 with parameter functions $\alpha(\mu), \gamma(\mu), \nu(\mu), \xi(\mu)$ and $\sigma(\mu)$ if*

1. *If $\eta(\mu, x)$ is concave in x , continuous in $(\mu, x) \in \mathbb{R}_{++}^n \times (\mathcal{F}^1 \cap \mathbb{R}_{++}^n)$ and has three derivatives in x , continuous in $(\mu, x) \in \mathbb{R}_{++} \times (\mathcal{F}^1 \cap \mathbb{R}_{++}^n)$.*
2. *$\nabla\eta(\mu, x)$ and $\nabla^2\eta(\mu, x)$ are continuously differentiable in μ ,*
3. *For any $\mu \in \mathbb{R}_{++}$, $\eta(\mu, x)$ is strongly $\alpha(\mu)$ -self-concordant on $\mathcal{F}^1 \cap \mathbb{R}_{++}^n$,*
4. *The parameter functions $\alpha(\mu), \gamma(\mu), \xi(\mu)$ and $\sigma(\mu)$ are continuous positive scalar functions on $\mu \in \mathbb{R}_{++}$,*
5. *For every $(\mu, x) \in \mathbb{R}_{++} \times (\mathcal{F}^1 \cap \mathbb{R}_{++}^n)$ and $h \in \mathbb{R}^n$,*

$$|\{\nabla\eta(\mu, x)h\}' - \{\ln \nu(\mu)\}'\{\nabla\eta(\mu, x)h\}| \leq \xi(\mu)\alpha(\mu)^{1/2}(-h^T\nabla^2\eta(\mu, x)h)^{1/2},$$

6. *For every $(\mu, x) \in \mathbb{R}_{++} \times (\mathcal{F}^1 \cap \mathbb{R}_{++}^n)$ and $h \in \mathbb{R}^n$,*

$$|\{h^T\nabla^2\eta(\mu, x)h\}' - \{\ln \gamma(\mu)\}'h^T\nabla^2\eta(\mu, x)h| \leq -2\sigma(\mu)h^T\nabla^2\eta(\mu, x)h.$$

We refer the reader to Nesterov and Nemirovskii [16] for the original definition of self-concordant families and their properties. The essence of the above definition is in conditions 5 and 6.

Theorem 3.1 *The family of functions $\eta : \mathbb{R}_{++} \times \mathcal{F} \mapsto \mathbb{R}$ is a strongly self-concordant family with parameters $\alpha(\mu) = \mu$, $\gamma(\mu) = \nu(\mu) = 1$, $\xi(\mu) = \frac{\sqrt{n+mK}}{\mu}$ and $\sigma(\mu) = \frac{\sqrt{m}}{2\mu}$.*

Proof. It is easy to verify that conditions 1 through 4 of Definition 3.2 hold. Lemma 3.2 and Lemma 3.3 show that conditions 5 and 6 are satisfied ■

In Lemmas 3.2 and 3.3 we bound the changes of $\nabla\eta(\mu, x)$ and $\nabla^2\eta(\mu, x)$ as μ changes. This requires us to calculate (y'_i, z'_i, s'_i) , which are the derivatives of $(y_i(\mu, x), z_i(\mu, x), s_i(\mu, x))$ with respect to μ .

Differentiating (2.9) we get

$$\begin{aligned} Y_i s'_i + S_i y'_i &= e, \\ W_i y'_i &= 0, \\ W_i^T z'_i - H_i y'_i + s'_i &= 0. \end{aligned} \tag{3.16}$$

Solving (3.16) we obtain

$$\begin{aligned} z'_i &= -R_i^{-1} W_i Q_i^2 y_i^{-1} \\ y'_i &= Q_i^2 (I - W_i^T R_i^{-1} W_i Q_i^2) y_i^{-1} \\ s'_i &= H_i Q_i^2 y_i^{-1} + (I - H_i Q_i^2) W_i^T R_i^{-1} W_i Q_i^2 y_i^{-1}. \end{aligned} \tag{3.17}$$

Lemma 3.2 For any $\mu > 0$, $x \in \mathcal{F}^1 \cap \mathbb{R}_{++}^n$ and $h \in \mathbb{R}^n$ we have

$$|\{\nabla \eta(\mu, x)^T h\}'| \leq \left[\frac{n + mK}{\mu} \nabla^2 \eta(\mu, x)^T [h, h] \right]^{1/2}. \tag{3.18}$$

Proof. Differentiating (3.3) with the respect to μ and applying (3.17) we get

$$\{\nabla \eta(\mu, x)\}' = -x^{-1} + \sum_{i=1}^K T_i^T R_i^{-1} W_i Q_i^2 y_i^{-1}. \tag{3.19}$$

We define

$$B := (X^{-1}, T_1^T R_1^{-1} W_1 Q_1^2 Y_1^{-1}, \dots, T_K^T R_K^{-1} W_K Q_K^2 Y_K^{-1}).$$

Let us first show that $BB^T \preceq \mu^{-1} \nabla^2 \eta(\mu, x)$.

Note that

$$\{\nabla \eta(\mu, x)\}' = Be, \tag{3.20}$$

where $e \in \mathbb{R}^{(n+Km)}$.

Now,

$$\begin{aligned} BB^T &= X^{-2} + \sum_{i=1}^K T_i^T R_i^{-1} W_i Q_i^2 Y_i^{-2} Q_i^2 W_i^T R_i^{-1} T_i \\ &= X^{-2} + \sum_{i=1}^K T_i^T R_i^{-1/2} Q_i Y_i^{-2} Q_i R_i^{-T/2} T_i. \end{aligned} \tag{3.21}$$

Observing that $Q_i Y_i^{-2} Q_i \preceq \mu^{-1} I$, in view of (3.21) and the definition of $\nabla^2 \eta(\mu, x)$ in (3.8) we have

$$BB^T \preceq X^{-2} + \mu^{-1} \sum_{i=1}^K T_i^T R_i^{-1} T_i \preceq \mu^{-1} \nabla^2 \eta(\mu, x). \quad (3.22)$$

Using (3.22) and (3.20) we get

$$\begin{aligned} \{\nabla \eta(\mu, x)^T\}' [\nabla^2 \eta(\mu, x)]^{-1} \{\nabla \eta(\mu, x)\}' &\leq \mu^{-1} e^T B^T (BB^T)^{-1} B e \\ &\leq \mu^{-1} e^T e \\ &= \mu^{-1} (n + mK). \end{aligned} \quad (3.23)$$

Hence for any $h \in \mathbb{R}^n$,

$$\begin{aligned} |\{\nabla \eta(\mu, x)^T h\}'| &\leq [\{\nabla \eta(\mu, x)^T\}' [\nabla^2 \eta(\mu, x)]^{-1} \{\nabla \eta(\mu, x)\}']^{1/2} [\nabla^2 \eta(\mu, x)[h, h]]^{1/2} \\ &\leq \left[\frac{n + mK}{\mu} \nabla^2 \eta(\mu, x)^T [h, h] \right]^{1/2} \blacksquare \end{aligned} \quad (3.24)$$

Lemma 3.3 For any $\mu > 0$, $x \in \mathcal{F}^1 \cap \mathbb{R}_{++}^n$ and $h \in \mathbb{R}^n$ we have

$$|\{\nabla^2 \eta(\mu, x)[h, h]\}'| \leq \frac{\sqrt{m}}{\mu} \nabla^2 \eta(\mu, x)[h, h]. \quad (3.25)$$

Proof. We have

$$\begin{aligned} \{Q_i^{-2} e\}' &= (H_i + Y_i^{-1} S_i)' e \\ &= (Y_i^{-1} S_i)' e \\ &= Y_i^{-2} (Y_i s_i' - S_i y_i') \\ &= Y_i^{-2} [I - 2S_i Q_i^2 (I - W_i^T R_i^{-1} W_i Q_i^2) Y_i^{-1}] e \\ &\quad \text{(substituting for } s_i' \text{ and } y_i' \text{ from (3.17))} \end{aligned} \quad (3.26)$$

Now

$$\begin{aligned} S_i Q_i^2 (I - W_i^T R_i^{-1} W_i Q_i^2) Y_i^{-1} &= \mu^{-1} S_i Q_i (I - Q_i W_i^T R_i^{-1} W_i Q_i) Q_i S_i \\ &\preceq \mu^{-1} S_i Q_i^2 S_i \\ &= \mu^{-1} S_i (H_i + Y_i^{-1} S_i)^{-1} S_i \\ &\preceq \mu^{-1} S_i (Y_i^{-1} S_i)^{-1} S_i \\ &\preceq I. \end{aligned} \quad (3.27)$$

From (3.27) and noting that $S_i Q_i^2 (I - W_i^T R_i^{-1} W_i Q_i^2) Y_i^{-1}$ is a symmetric positive semidefinite matrix we conclude that

$$\|I - 2S_i Q_i^2 (I - W_i^T R_i^{-1} W_i Q_i^2) Y_i^{-1}\|_2 \leq 1. \quad (3.28)$$

Therefore, in view of (3.26) and (3.28), for any $h \in \mathbb{R}^n$ we have

$$\begin{aligned} |h^T \{Q_i^{-2}\}' h| &\leq \|Y_i^2 \{Q_i^{-2}\}' e\|_\infty h^T Y_i^{-2} h \quad (\text{noting that } Y_i^2 \{Q_i^{-2}\}' \text{ is diagonal}) \\ &= \|[I - 2S_i Q_i^2 (I - W_i^T R_i^{-1} W_i Q_i^2) Y_i^{-1}] e\|_\infty h^T Y_i^{-2} h \\ &\leq \mu^{-1} \|I - 2S_i Q_i^2 (I - W_i^T R_i^{-1} W_i Q_i^2) Y_i^{-1}\|_2 \|e\|_2 h^T (H_i + Y_i^{-1} S_i) h \\ &\leq \frac{\sqrt{m}}{\mu} h^T Q_i^{-2} h. \end{aligned} \quad (3.29)$$

Differentiating (3.8) with respect to μ and using (3.29), for any $h \in \mathbb{R}^n$ we have

$$\begin{aligned} |\{\nabla^2 \eta(\mu, x)[h, h]\}'| &= |h^T X^{-2} h - \sum_{i=1}^K h^T T_i^T R_i^{-1} R_i' R_i^{-1} T_i h| \\ &= |h^T X^{-2} h + \sum_{i=1}^K h^T T_i^T R_i^{-1} W_i Q_i^2 \{Q_i^{-2}\}' Q_i^2 W_i^T R_i^{-1} T_i h| \\ &\leq |h^T X^{-2} h + \frac{\sqrt{m}}{\mu} \sum_{i=1}^K h^T T_i^T R_i^{-1} W_i Q_i^2 W_i^T R_i^{-1} T_i h| \\ &\leq \sqrt{m} h^T X^{-2} h + \frac{\sqrt{m}}{\mu} \sum_{i=1}^K h^T T_i^T R_i^{-1} T_i h \\ &= \frac{\sqrt{m}}{\mu} \nabla^2 \eta(\mu, x)[h, h] \quad \blacksquare \end{aligned} \quad (3.30)$$

4. The Two-Stage Stochastic Convex QP Algorithm

The algorithm is a standard primal interior point method, which reduces μ by a factor at each iteration and seeks to approximate the minimizer $x(\mu)$ for each μ by taking one or more Newton steps. The novelty is in computing the Newton direction from the solutions of the decomposed second stage problems. As μ varies, the minimizers $x(\mu)$ form the central path. The limit $\lim_{\mu \rightarrow 0} x(\mu)$ exists and it is an optimal solution to the original problem (2.1). Therefore by tracing the central path as $\mu \rightarrow 0$ this procedure will generate a strictly feasible ϵ -solution to (2.5).

For a given μ the optimality conditions for the first stage problem (2.5) are:

$$\begin{aligned} \nabla \eta(\mu, x) - A^T \lambda &= 0, \\ Ax &= b. \end{aligned} \quad (4.1)$$

Let

$$g := \nabla\eta(\mu, x) - A^T\lambda \quad \text{and} \quad \Omega := \nabla^2\eta(\mu, x).$$

The Newton system takes the following form:

$$\begin{aligned} \Omega\Delta x - A^T\Delta\lambda &= -g, \\ A\Delta x &= 0. \end{aligned}$$

Solving the above system we get the Newton direction:

$$\begin{aligned} \Delta x &= -(\Omega^{-1} - \Omega^{-1}A^T(A\Omega^{-1}A^T)^{-1}A\Omega^{-1})g, \\ \Delta\lambda &= (A\Omega^{-1}A^T)^{-1}A\Omega^{-1}g. \end{aligned} \tag{4.2}$$

Note that although problems (2.5- 2.7) and (2.10) share the same central path, the associated Newton directions are not identical and lead to different ways of path following. A generic primal path following algorithm is given below.

4.1 Algorithm

Here $\beta > 0, \gamma \in (0, 1)$ and $\theta > 0$ are suitable scalars. We make their values more precise in Theorems 4.1 and 4.2. The desired precision ϵ , an initial point $x^0 \in \mathcal{F}^0$ and μ^0 are given as inputs.

Initialization.

$$x = x^0; \mu = \mu^0.$$

Step 1.

- 1.1. For all i solve the optimality conditions (2.9) to find $(y_i(\mu, x), z_i(\mu, x), s_i(\mu, x))$.
- 1.2. Compute the Newton direction $(\Delta x, \Delta\lambda)$ from (4.2).
- 1.3. Let $\delta(\mu, x) = \sqrt{\frac{1}{\mu}\Delta x^T\Omega\Delta x}$. If $\delta \leq \beta$ go to Step2.
- 1.4. Set $x = x + \theta\Delta x$ and $\lambda = \lambda + \theta\Delta\lambda$. Then go to Step 1.1.

Step 2. If $\mu \leq \epsilon$ stop, otherwise set $\mu = \gamma\mu$ and go to Step 1.1.

A practical approach to initialize the algorithm will introduce artificial variables with large costs so that an initial point $x^0 \in \mathcal{F}^0$ is readily available, while ensuring that the artificial variables go to zero as solutions converge. A formal approach to find an initial point $x^0 \in \mathcal{F}^0$ is discussed in Zhao [28].

In the above algorithm we assume that we can find exact solutions of the optimality conditions (2.9). This assumption considerably simplifies the complexity analysis. A practical

implementation of this algorithm will use an approximate solution of the optimality conditions (2.9) to construct the Newton direction (4.2). Theorems 4.1 and 4.2 give two standard complexity results for the generic primal interior point method. In the short-step version of the algorithm barrier parameter μ is decreased by a factor $1 - \sigma/\sqrt{n+m}$ ($\sigma > 0$) in each iteration.

An iteration of the short-step algorithm is performed as follows. At the beginning of iteration k , x^k is close to the central path, i.e. $\delta(\mu^k, x^k) \leq \beta$. After reducing the parameter from μ^k to $\mu^{k+1} = \gamma\mu^k$, we will have $\delta(\mu^{k+1}, x^k) \leq 2\beta$. Then a Newton step with step size $\theta = 1$ is taken resulting in a new point x^{k+1} with $\delta(\mu^{k+1}, x^{k+1}) \leq \beta$. In the long-step version we decrease the barrier parameter μ by an arbitrarily constant factor ($\lambda \in (0, 1)$). It has potential for much faster progress, however, several damped Newton steps might be needed for restoring the proximity to the central path. We have the following theorems for the short and long step versions of the algorithm. The proves of these theorems are given in the next section.

Theorem 4.1 *Let μ^0 be the initial barrier parameter, $\epsilon > 0$ the stopping criterion and $\beta = (2 - \sqrt{3})/2$. If the starting point x^0 is sufficiently close to the central path, i.e. $\delta(\mu^0, x^0) \leq \beta$, then the short-step algorithm reduces the barrier parameter μ at a linear rate and terminates within $O(\sqrt{n+m} \ln \mu^0/\epsilon)$ iterations.*

Theorem 4.2 *Let μ^0 be the initial barrier parameter and $\epsilon > 0$ be the stopping criterion and $\beta = 1/6$. If the starting point x^0 is sufficiently close to the central path, i.e. $\delta(\mu^0, x^0) \leq \beta$, then the long-step algorithm reduces the barrier parameter μ at a linear rate and terminates within $O((n+m) \ln \mu^0/\epsilon)$ iterations.*

5. Convergence Analysis for the Short and Long Step Algorithms

The following proposition follows directly from the definition of self-concordance and is due Nesterov and Nemirovskii [16, Theorem 2.1.1].

Proposition 5.1 *For any $\mu > 0$, $x \in \mathcal{F}$ and Δx let $\delta := \sqrt{\frac{1}{\mu} \Delta x^T \nabla^2 \eta(\mu, x) \Delta x}$. Then for $\delta < 1$, $\tau \in [0, 1]$ and any $h \in \mathbb{R}^n$ we have*

$$(1 - \tau\delta)^2 h^T \nabla^2 \eta(\mu, x) h \leq h^T \nabla^2 \eta(\mu, x + \tau\Delta x) h \leq (1 - \tau\delta)^{-2} h^T \nabla^2 \eta(\mu, x) h. \quad (5.1)$$

For the estimation of number of Newton steps needed for recentering we use two different merit functions to measure the speed of Newton's method. We use $\delta(\mu, x)$ for the short-step algorithm and the first stage objective $\eta(\mu, x)$ (defined in Step 1) for the long-step algorithm. The following lemma is due to Theorem 2.2.3 in [16] and describes the behavior of the Newton method as applied to $\eta(\mu, \cdot)$.

Lemma 5.1 For any $\mu > 0$ and $x \in \mathcal{F}$, let Δx be the Newton direction calculated by (4.2) and let $\delta(\mu, x) := \sqrt{\frac{1}{\mu} \Delta x^T \nabla^2 \eta(\mu, x) \Delta x}$. Then the following relations hold:

(i) If $\delta < 2 - \sqrt{3}$ then

$$\delta(\mu, x + \Delta x) \leq \left(\frac{\delta}{1 - \delta} \right)^2 \leq \frac{\delta}{2}.$$

(ii) If $\delta \geq 2 - \sqrt{3}$ then

$$\eta(\mu, x) - \eta(\mu, x + \bar{\theta} \Delta x) \geq \mu[\delta - \ln(1 + \delta)],$$

where $\bar{\theta} = (1 + \delta)^{-1}$.

5.1 Complexity of the Short-Step Algorithm

We will show that in this version of the algorithm a single Newton step is sufficient for recentering after updating barrier parameter μ . To this end we will make use of Theorem 3.1.1 in [16], which can be restated for the present context as in the next proposition.

Proposition 5.2 Let $\varphi_\kappa(\eta; \mu, \mu^+) := \left(\frac{1 + \sqrt{m}}{2} + \frac{\sqrt{n + mK}}{\kappa} \right) \ln \gamma^{-1}$. Assume that $\delta(\mu, x) < \kappa$ and $\mu^+ := \gamma \mu$ satisfies

$$\varphi_\kappa(\eta; \mu, \mu^+) \leq 1 - \frac{\delta(\mu, x)}{\kappa}.$$

Then $\delta(\mu^+, x) < \kappa$.

Lemma 5.2 Let $\mu^+ = \gamma \mu$ where $\gamma = 1 - \sigma / \sqrt{n + mK}$ and $\sigma \leq 0.1$. Furthermore let $\beta = (2 - \sqrt{3})/2$. If $\delta(\mu, x) \leq \beta$ then $\delta(\mu^+, x) \leq 2\beta$.

Proof. Let $\kappa = 2\beta = 2 - \sqrt{3}$. It is easy to verify that for $\sigma \leq 0.1$ μ^+ satisfies

$$\begin{aligned} \varphi_\kappa(\eta; \mu, \mu^+) &= \left(\frac{1 + \sqrt{m}}{2} + \frac{\sqrt{n + mK}}{\kappa} \right) \ln(1 - \sigma / \sqrt{n + mK})^{-1} \\ &\leq \frac{1}{2} \leq 1 - \frac{\delta(\mu, x)}{\kappa}. \end{aligned}$$

Now Proposition 5.2 implies

$$\delta(\mu^+, x) \leq \kappa = 2\beta \quad \blacksquare$$

From Lemma 5.1 and Lemma 5.2 it is clear that we can reduce μ by the factor $\gamma = 1 - \sigma / \sqrt{n + mK}$, $\sigma < 0.1$ at each iteration and a single Newton step is sufficient for recentering.

Hence Theorem 4.1 follows.

5.2 Complexity of the Long-Step Algorithm

For the analysis of the long-step algorithm we use η as the merit function since the iterates generated by the less conservative long-step algorithm may violate the condition, $\delta < 2 - \sqrt{3}$, required in part (i) of Lemma 5.1. Our analysis follows the template in Zhao [28].

Assume that we have a point x^{k-1} sufficiently close to $x(\mu^{k-1})$. Then we reduce the barrier parameter from μ^{k-1} to $\mu^k = \gamma\mu^{k-1}$, where $\gamma \in (0, 1)$. While searching for a point x^k that is sufficiently close to $x(\mu^k)$ our algorithm will generate a finite sequence of points $p_1, \dots, p_N \in \mathcal{F}$ and we finally set $x^k = p_N$. We need to determine an upper bound on N , the number of Newton iteration needed for recentering. We begin by determining an upper bound on the difference

$$\phi(\mu^k, x^{k-1}) := \eta(\mu^k, x^{k-1}) - \eta(\mu^k, x(\mu^k)). \quad (5.2)$$

Then by part (ii) of Lemma 5.1 we know that at $p_i \in F$, independent of i , a Newton step with step size $\bar{\theta} = (1 + \delta)^{-1}$ decreases $\eta(\mu^k, p_i)$ at least by a certain amount which depends on the current value of δ and μ . A line search might yield even a larger decrease. Consequently we will get an upper bound on N .

The next proposition [28, Lemma 7] and lemma give upper bounds on $\phi(\mu^{k-1})$ and $\phi'(\mu^{k-1})$, respectively. They facilitate us bounding $\phi(\mu^k)$.

Proposition 5.3 *For any $\mu > 0$ and $x \in \mathcal{F}^1 \cap \mathbb{R}_{++}^n$, we denote $\tilde{\Delta}x := x - x(\mu)$ and define*

$$\tilde{\delta}(\mu, x) := \sqrt{\frac{1}{\mu} \tilde{\Delta}x^T \nabla^2 \eta(\mu, x) \tilde{\Delta}x}.$$

If $\tilde{\delta} < 1$, then

$$\phi(\mu, x) \leq \mu \left[\frac{\tilde{\delta}}{1 - \tilde{\delta}} + \ln(1 - \tilde{\delta}) \right]. \quad (5.3)$$

Lemma 5.3 *Let $\tilde{\Delta}x$ and $\tilde{\delta}$ be as defined in Proposition 5.3. For any $\mu > 0$ and $x \in \mathcal{F}^1 \cap \mathbb{R}_{++}^n$, if $\tilde{\delta} < 1$, then*

$$|\phi'(\mu, x)| \leq -\sqrt{n + mK} \ln(1 - \tilde{\delta}).$$

Proof. For any $\mu > 0$, applying chain rule we can write

$$\phi'(\mu, x) = \eta'(\mu, x) - \eta'(\mu, x(\mu)) - \nabla \eta(\mu, x(\mu))^T x'(\mu).$$

The optimality conditions (4.1) imply that $\nabla\eta(\mu, x(\mu))^T x'(\mu) = 0$.

Therefore

$$\phi'(\mu, x) = \eta'(\mu, x) - \eta'(\mu, x(\mu)). \quad (5.4)$$

From (3.23) in the proof of Lemma 3.2 we have

$$\{\nabla\eta(\mu, x)^T\}'[\nabla^2\eta(\mu, x)]^{-1}\{\nabla\eta(\mu, x)\}' \leq \mu^{-1}(n + mK). \quad (5.5)$$

From (5.4) applying the mean-value theorem we get

$$\begin{aligned} |\phi'(\mu, x)| &= \left| \int_0^1 \{\nabla\eta(\mu, x(\mu) + \tau\tilde{\Delta}x)^T\}'\tilde{\Delta}x \, d\tau \right| \\ &\leq \int_0^1 \left[\tilde{\Delta}x^T \nabla^2\eta(\mu, x(\mu) + \tau\tilde{\Delta}x)\tilde{\Delta}x \right]^{1/2} \\ &\quad \left[\{\nabla\eta(\mu, x(\mu) + \tau\tilde{\Delta}x)^T\}'[\nabla^2\eta(\mu, x(\mu) + \tau\tilde{\Delta}x)]^{-1}\{\nabla\eta(\mu, x(\mu) + \tau\tilde{\Delta}x)^T\}' \right]^{1/2} d\tau \end{aligned} \quad (5.6)$$

From (5.6), using (5.5) and Proposition 5.1 and noting $x(\mu) + \tau\tilde{\Delta}x = x - (1 - \tau)\tilde{\Delta}x$ we get

$$\begin{aligned} |\phi'(\mu, x)| &\leq \int_0^1 \frac{\sqrt{\tilde{\Delta}x^T \nabla^2\eta(\mu, x)\tilde{\Delta}x}}{1 - \tilde{\delta} + \tau\tilde{\delta}} \sqrt{\frac{n + mK}{\mu}} \, d\tau \\ &\leq \int_0^1 \frac{\sqrt{\mu}\tilde{\delta}}{1 - \tilde{\delta} + \tau\tilde{\delta}} \sqrt{\frac{n + mK}{\mu}} \, d\tau \\ &= -\sqrt{n + mK} \ln(1 - \tilde{\delta}) \blacksquare \end{aligned} \quad (5.7)$$

Lemma 5.4 *Let $\mu > 0$ and $x \in \mathcal{F}^1 \cap \mathbb{R}_{++}^n$ be such that $\tilde{\delta} < 1$, where $\tilde{\delta}$ is as defined in Proposition 5.3. Let $\mu^+ = \gamma\mu$ with $\gamma \in (0, 1)$. Then*

$$\eta(\mu^+, x) - \eta(\mu^+, x(\mu^+)) \leq O(n + mK)\mu^+.$$

Proof. Differentiating (5.4) we obtain

$$\phi''(\mu, x) = \eta''(\mu, x) - \eta''(\mu, x(\mu)) - \{\nabla\eta(\mu, x)^T\}'x'(\mu). \quad (5.8)$$

To obtain $x'(\mu)$ we take the derivative of the optimality conditions (4.1) of the first stage problem (2.5):

$$\begin{aligned} \{\nabla\eta(\mu, x(\mu))\}' + \{\nabla^2\eta(\mu, x(\mu))\}'x'(\mu) - A^T\lambda'(\mu) &= 0 \\ Ax'(\mu) &= 0. \end{aligned} \quad (5.9)$$

Solving (5.9) we get

$$x'(\mu) = -[\Omega^{-1} - \Omega^{-1}A^T(A\Omega^{-1}A^T)^{-1}A\Omega^{-1}]\{\nabla\eta(\mu, x(\mu))\},$$

where $\Omega = \nabla^2\eta(\mu, x(\mu))$.

Now we have

$$\begin{aligned} -\{\nabla\eta(\mu, x)^T\}'x'(\mu) &= \{\nabla\eta(\mu, x(\mu))^T\}'[\Omega^{-1} - \Omega^{-1}A^T(A\Omega^{-1}A^T)^{-1}A\Omega^{-1}]\{\nabla\eta(\mu, x(\mu))\}' \\ &\leq \{\nabla\eta(\mu, x(\mu))^T\}'\Omega^{-1}\{\nabla\eta(\mu, x(\mu))\}' \\ &\leq \mu^{-1}(n + mK). \end{aligned} \quad (5.10)$$

The last inequality above follows using (3.23). We still need to bound the first two terms in the right-hand-side of the inequality (5.8).

From the definition of $\eta(\mu, x)$, see (2.5), it can be easily seen that $\eta''(\mu, x) = \sum_{i=1}^K \rho_i''(\mu, x)$.

Now differentiating (3.1) and using (3.17) we obtain

$$\begin{aligned} \rho_i'(\mu, x) &= (H_i y_i + d_i)^T y_i' - e^T \ln y_i - \mu e^T Y_i^{-1} y_i' \\ &= (H_i y_i + d_i - s_i)^T y_i' - e^T \ln y_i \\ &= z_i^T W_i y_i' - e^T \ln y_i \\ &= -e^T \ln y_i \quad (\text{noting that } W_i y_i' = 0). \end{aligned}$$

Hence,

$$\begin{aligned} \rho_i''(\mu, x) &= -e^T Y_i^{-1}(\mu, x) y_i'(\mu, x) \\ &= -e^T Y_i^{-1} Q_i^2 (I - W_i^T R_i^{-1} W_i Q_i^2) y_i^{-1} \\ &= -e^T Y_i^{-1} Q_i (I - Q_i W_i^T R_i^{-1} W_i Q_i) Q_i Y_i^{-1} e \\ &\geq -\|e\|_2^2 \|Y_i^{-1} Q_i\|_2^2 \\ &= -m \|Q_i Y_i^{-2} Q_i\|_2 \\ &\geq -\frac{m}{\mu} \quad (\text{noting } Q_i Y_i^{-2} Q_i \preceq \mu^{-1} I). \end{aligned} \quad (5.11)$$

Hence for any $x \in F$ the following holds

$$\eta''(\mu, x) \geq -\frac{mK}{\mu}. \quad (5.12)$$

Since $\eta(\mu, x)$ is strictly concave in μ for any $x \in F$ we also have

$$\eta''(\mu, x) \leq 0. \quad (5.13)$$

Now using the bounds given in (5.10), (5.12) and (5.13) from (5.8) we get

$$\phi''(\mu, x) \leq \frac{n + 2mK}{\mu}. \quad (5.14)$$

Using Proposition 5.3, Lemma 5.3 and (5.14) we have

$$\begin{aligned} \phi(\mu^+, x) &= \phi(\mu, x) + \phi'(\mu, x)(\mu^+ - \mu) + \int_{\mu}^{\mu^+} \int_{\mu}^{\tau} \phi''(\mu) d\mu d\tau \\ &\leq \mu \left[\frac{\tilde{\delta}}{1 - \tilde{\delta}} + \ln(1 - \tilde{\delta}) \right] - \sqrt{n + mK} \ln(1 - \tilde{\delta})(\mu - \mu^+) \\ &\quad + (n + 2mK) \int_{\mu}^{\mu^+} \int_{\mu}^{\tau} \mu^{-1} d\mu d\tau \\ &\leq \mu \left[\frac{\tilde{\delta}}{1 - \tilde{\delta}} + \ln(1 - \tilde{\delta}) \right] - \sqrt{n + mK} \ln(1 - \tilde{\delta})(\mu - \mu^+) \\ &\quad + (n + 2mK) \ln \gamma^{-1} (\mu - \mu^+). \end{aligned} \quad (5.15)$$

Since γ and $\tilde{\delta}$ are absolute constants (5.15) implies that $\eta(\mu^+, x) - \eta(\mu^+, x(\mu^+)) \leq O(n + mK)\mu^+$ and proves the lemma ■

Note that Proposition 5.3, Lemma 5.3 and Lemma 5.4 all require $\tilde{\delta}$ be less than one. However, we cannot evaluate $\tilde{\delta}$ since we do not explicitly know the points $x(\mu)$ forming the central path. Nonetheless we can evaluate δ and $\tilde{\delta}$ is proportional to δ , as shown in the following proposition which is due Zhao [28, Lemma 9].

Proposition 5.4 *For any $\mu > 0$, $x \in \mathcal{F}^1 \cap \mathbb{R}_{++}^n$ and $\lambda \in \mathbb{R}^s$, let $(\Delta x, \Delta \lambda)$ be the Newton direction defined in (4.2) and $(\tilde{\Delta} x, \tilde{\Delta} \lambda) := (x - x(\mu), \lambda - \lambda(\mu))$. We denote*

$$\delta(\mu, x) := \sqrt{\frac{1}{\mu} \Delta x^T \Omega \Delta x} \quad \tilde{\delta}(\mu, x) := \sqrt{\frac{1}{\mu} \tilde{\Delta} x^T \Omega \tilde{\Delta} x},$$

where $\Omega := \nabla^2 \eta(\mu, x)$. If $\delta \leq 1/6$ then

$$\frac{2}{3} \delta \leq \tilde{\delta} \leq 2\delta. \quad (5.16)$$

Lemma 5.1 implies that each line search should decrease the value of η by at least $\mu[\delta - \ln(1 + \delta)]$. Therefore, in view of Lemma 5.1 and Lemma 5.4 it is clear that after reducing μ by a factor $\gamma \in (0, 1)$, at most $O(n + mK)$ Newton iterations will be needed for recentering. In the long-step version of our algorithm we need to update barrier parameter μ no more

than $O(\ln \mu^0/\epsilon)$ times.

Theorem 4.2 follows from Lemma 5.1 (ii), Lemma 5.4 and Proposition 5.4.

6. Concluding Remarks

The algorithm we introduced follows the primal central trajectory in the first stage. At each iteration using optimal dual solutions of the second stage problems, it generates gradient and Hessian information for the first stage problem and takes a Newton step in the primal space. Our theoretical analysis assumes exact solutions of the second stage problems. Practically it is neither possible nor necessary to find exact optimal solutions. It is well known that exact Newton directions are not necessary for the linear convergence of interior point methods. A computational study of the effect of inexact search directions, and techniques to take advantage of the increased algorithmic flexibility offered by the decomposition method are subjects of further investigation.

It is also possible to extend the results of this paper to symmetric conic problems with general convex objectives. Such results will be presented in our future work.

Bibliography

- [1] O. Bahn , O. du Merle, J.-L. Goffin and J.P. Vial, "A cutting plane method from analytic centers for stochastic programming," *Mathematical Programming* 69, (1995) 45-73.
- [2] A. B. Berkelaar, C. Dert, B. Oldenkamp and S. Zhang, "A primal-dual decomposition based interior point approach to two-stage stochastic linear programming," *Operations Research* 50-5, (2002) 904-915.
- [3] J. R. Birge, X. Chen, L. Qi and Z. Wei, "A stochastic Newton method for stochastic quadratic programs with recourse," Working Paper, (1995).
- [4] J. R. Birge, C. J. Donohue, D. F. Holmes, and O. G. Svintsiski, "A parallel implementation of the nested decomposition algorithm for multistage stochastic linear programs," *Mathematical Programming* 75, (1996) 327- 352.
- [5] J. R. Birge and D. F. Holmes, "Computing block-angular Karmarkar projections with applications to stochastic programming," *Computational Optimization and Applications* 1, (1992) 245-276.
- [6] T. Carpenter, I. Lustig and J. Mulvey, "Formulating stochastic programs for interior point methods," *Operations Research* 39, (1991) 757-770.
- [7] X. Chen, L. Qi and R. Womersley, "Newton's method for quadratic stochastic programs with recourse," *J. Comput. Appl. Math.*, 60, 29-46, 1995.
- [8] J. Czyzyk, R. Fourer and S. Mehrotra, "A study of the augmented system and column splitting approaches for solving two-stage stochastic linear programs by interior point methods," *ORSA Journal on Computing* 7, (1995) 474-490.
- [9] A. V. Fiacco, G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, (John Wiley, New York, 1968).
- [10] J-L Goffin, A. Haurie and J-P Vial, "Decomposition and nondifferentiable optimization with the projective algorithm," *Management Science* 38, (1992) 284-302.
- [11] J-L Goffin and J-P Vial, "Convex nondifferentiable optimization: A survey focused on the analytic center cutting plane method," HEC/Logilab Technical Report (1999).
- [12] A.J. King and R.T. Rockafellar, "Asymptotic theory for solutions in statistical estimation and stochastic programming", *Mathematics of Operations Research* 18, (1993) 148-162.
- [13] J. Linderoth, A. Shapiro, S. Wright, "The empirical behavior of sampling methods for stochastic programming", Published electronically in: Optimization Online.
- [14] O. L. Mangasarian, *Nonlinear Programming*, Classics in Applied Mathematics (SIAM, 1994).
- [15] J. E. Nesterov and A. S. Nemirovsky, "Polynomial-time barrier methods in convex programming," *Ekonomika i Matem. Metody* 24,7 (1988) 1084-1091 (In Russian; English trans. Matekon: Translations of Russian and East European Math. Economics.)

- [16] J. E. Nesterov and A. S. Nemirovsky, *Interior-Point Polynomial Algorithms in Convex Programming*, Studies in Applied Mathematics (SIAM, 1994).
- [17] J. Renegar, *A Mathematical view of Interior-Point Methods in Convex Programming*, MPS-SIAM Series on Optimization (SIAM, 2001).
- [18] R. T. Rockafellar, "Linear-quadratic programming and optimal control," *SIAM Journal on Control and Optimization*, 25, 781-814, 1987.
- [19] R. T. Rockafellar, "Computational schemes for solving large-scale problems in extended linear-quadratic programming," *Mathematical Programming*, 48, 447-474, 1990.
- [20] 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.
- [21] R. T. Rockafellar and R. J-B. Wets, "Linear quadratic problems with stochastic penalties: the finite generation algorithm," in: *Numerical Techniques for Stochastic Optimization Problems* (Y. Ermoliev and R. J-B Wets (eds)), Springer-Verlag Lecture Notes in Control and Information Sciences 81, (1987) 545-560.
- [22] R. T. Rockafellar and R. J-B. Wets, "Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time," *SIAM J. Control Opt.*, 25, 781-814, 1990.
- [23] R. T. Rockafellar, R. J-B. Wets, "Scenarios and policy aggregation in optimization under uncertainty," *Math. of Oper. Res.* 16, (1991) 119-147.
- [24] A. Shapiro, "Asymptotic behavior of optimal solutions in stochastic programming," *Mathematics of Operations Research* 18, (1993) 829-845.
- [25] A. Shapiro, T. Homem-de-Mello, "On rate of convergence of Monte Carlo approximations of stochastic programs", *SIAM J. Optimization* 11, (2000) 70-86.
- [26] A. Shapiro, "Stochastic Programming by Monte Carlo Simulation Methods", Published electronically in: Stochastic Programming E-Print Series (2000).
- [27] R. M. Van Slyke and R. J. Wets, "L-Shaped linear programs with applications to optimal control and stochastic linear programming," *SIAM Journal of Applied Mathematics* 17, (1969) 638-663.
- [28] G. Zhao, "A log-barrier method with Benders decomposition for solving two-stage stochastic linear programs," *Mathematical Programming Ser. A* 90, (2001) 507-536.
- [29] C. Zhu, "Modified proximal point algorithm for extended linear-quadratic programming," Johns Hopkins University, technical report, 1992.
- [30] C. Zhu and R. T. Rockafellar, "Primal-dual projected gradient algorithms for extended linear-quadratic programming," University of Washington, technical report, 1992.