Self-concordant Tree and Decomposition Based Interior Point Methods for Stochastic Convex Optimization Problem *

Michael Chen[†]and Sanjay Mehrotra[‡]

May 23, 2007

Technical Report 2007-04
Department of Industrial Engineering and Management Sciences,
Robert R. McCormick School of Engineering,
Northwestern University, Evanston, Illinois 60208.

Abstract

We consider barrier problems associated with two and multistage stochastic convex optimization problems. We show that the barrier recourse functions at any stage form a self-concordant family with respect to the barrier parameter. We also show that the complexity value of the first stage problem increases additively with the number of stages and scenarios. We use these results to propose a prototype primal interior point decomposition algorithm for the two-stage and multistage stochastic convex optimization problems admitting self-concordant barriers.

1 Introduction

We study the self-concordance properties of two and multistage stochastic convex optimization problems in this paper. These properties are then used to analyze prototype decomposition based interior point algorithms for solving these problems. Let us consider a two-stage stochastic convex problem with K scenarios:

$$\begin{array}{ll} min\, c^T x + \sum_{k=1}^{K} \bar{\eta}^k(x) & s.t.\ x \in G \cap L \\ \bar{\eta}^k(x) := min\, d^{k^T} y^k & s.t.\ y^k \in G^k \cap L^k(x), \end{array} \tag{1}$$

where G and $G^k, k = 1, \ldots, K$ are compact convex sets of $x \in \mathbb{R}^n$ and $y^k \in \mathbb{R}^{n^k}, k = 1, \ldots, K$, respectively. We assume that the set G has a non-empty relative interior, and the sets G^k, $k = 1, \ldots, K$, have a non-empty relative interior for any x that is in the relative interior of G

*Supported by NSF grants DMI-0200151, DMI-0522765 and ONR grant N00014-01-1-0048.
†Department of IE/MS, Northwestern University, Evanston, IL 60208, m-chen@northwestern.edu
‡Department of IE/MS, Northwestern University, Evanston, IL 60208, mehrotra@iems.northwestern.edu

(*complete recourse*). L and $L^k, k = 1, \ldots, K$, represent affine spaces, i.e., $L \equiv \{x \mid Ax = b\}$, and $L^k(x) \equiv \{y^k \mid Q^k y^k = q^k + T^k x\}, k = 1, \ldots, K$. Without loss of generality, we also assume that A and $Q^k, k = 1, \ldots, K$ have full row rank. The objective functions of both first and second stage problems are linear, and we assume that d^k, $k = 1, \ldots, K$, has already absorbed the scenario probabilities π^k.

We now discuss the assumptions imposed on the form of (1). The assumption that the feasible sets are compact and have a non-empty interior can be satisfied by introducing artificial variables and constraints. We can reformulate a nonlinear convex objective function $d^k(y^k)$ as $min\, z^k$ and $z^k - d^k(y^k) \geq 0$. We can also redefine a convex set $G^k = \{y^k \mid h_i^k(x, y^k) \geq 0, i = 1, \ldots, l^k\}$ equivalently as $\bar{G}^k = \{(Y^k, \bar{y}) \mid h_i^k(\bar{y}, y^k) \geq 0, \bar{y} - x = 0\}$, where $(y^k, \bar{y})$ are the second stage decision variables and $\bar{y} - x = 0$ becomes part of the definition of $L^k(x)$. This satisfies the assumption that the first and the second stage variables do not interact in the nonlinear constraints. A prior assurance of full recourse assumption is more difficult, however, for a given first stage solution the feasibility of the second stage problems can be ensured by using artificial variables. A discussion of problems without complete recourse assumption is beyond the scope of this paper. We refer interested readers to Birge and Louveaux [2] for some insightful examples and the potential difficulties without this assumption.

The decomposition based primal interior point algorithms developed in this paper for the two-stage stochastic convex problem generalize those proposed by Zhao [18] and Mehrotra and Ozevin [8, 9] for two-stage linear and two-stage semi-definite problems. However, the analysis in the current case requires different proof techniques since we no longer assume specific form of the feasible regions. In particular, this analysis draws from the literature on sensitivity and stability analysis of nonlinear programming [5]. A central step to this analysis is showing that the second stage barrier recourse functions and the first stage barrier functions form self-concordant families.

In the multistage case, any intermediate stage recourse function contains recourse functions of the next stage, and hence imposes extra difficulty in the analysis. Using the proven self-concordant family properties of each last two-stage problem as a starting point, we prove by induction that each objective function and the recourse function of any stage are self-concordant families with respect to the barrier parameter. The complexity values accumulate additively across the scenario tree from the last stage to the first stage. This result leads to showing that the number of the first stage Newton steps required to solve multistage stochastic convex programs are polynomial in the total complexity value, under the assumption that the second and the subsequent stage centering problems are solved exactly in a decomposition setting.

This paper is organized as follows. In Section 2 we consider the two-stage problem and prove that each barrier recourse function forms a self-concordant family. In Section 3 we analyze the first stage barrier objective function and show that it also forms a self-concordant family. In Section 4 we propose both short-step and long-step path following interior point algorithms, and analyze the rate of convergence of the long step algorithm. Section 5 shows the self-concordant family property of any intermediate stage recourse functions and objective functions for the multistage problem. This section also gives a prototype decomposition algorithm for the multistage problem. Section 6 gives some concluding remarks on the results presented in this paper.

2 Two-Stage Stochastic Convex Programs

2.1 Two-Stage Barrier Recourse Problem

We regularize the first and second stage problems by barrier functions $b(x)$ on $intG$ and $B^k(y^k)$ on $intG^k, k = 1, \ldots, K$. Here $intG$ and $intG^k$ represent the relative interiors of sets G and $G^k, k = 1, \ldots, K$, respectively. The two-stage barrier problem is thus:

$$min \quad f(x, \mu) \equiv c^T x + \mu b(x) + \textstyle\sum_{k=1}^{K} \eta^k(x, \mu) \quad s.t.\ x \in L \tag{2}$$

$$\eta^k(x, \mu) \equiv min\, {d^k}^T y^k + \mu B^k(y^k) \qquad s.t.\ y^k \in L^k(x). \tag{3}$$

Here μ is a positive scalar, $b(x)$ and $B^k(y^k)$ are assumed to be non-degenerate and strongly self-concordant-barrier functions with complexity values ϑ^0 and ϑ^k of $intG$ and $intG^k$, $k = 1, \ldots, K$, respectively.

Definition 2.1 *Nesterov and Nemirovskii [11] and Renegar [13]. Let $G \subseteq R^n$ be a closed convex domain, and $intG$ be its non-empty interior. A function $B(x) : intG \to R$ is called non-degenerate and strongly α-self-concordant if $\nabla^2 B(x)$ is positive definite, and*

$$|\nabla^3 B(x)[h, h, h]| \le 2\alpha^{-\frac{1}{2}} (\nabla^2 B(x)[h, h])^{\frac{3}{2}}, \forall x \in intG, h \in R^n. \tag{4}$$

Furthermore, if $B(x) \to \infty$ as x converges to a boundary point of G, then $B(x)$ is called a ϑ-self-concordant barrier if $\alpha = 1$ in (4) and

$$\vartheta \equiv sup\{\nabla B(x)^T [\nabla^2 B(x)]^{-1} \nabla B(x) \mid x \in intG\} < \infty.\square \tag{5}$$

The parameter ϑ is called the complexity value of a self-concordant barrier.

Definition 2.2 *Nesterov and Nemirovskii [11] A family of functions $\{\psi(x, \mu), \mu > 0\}$ is strongly self-concordant on nonempty open convex domain $\Omega \in \mathbb{R}^n$ with positive scalar parameter functions $\alpha(\mu), \gamma(\mu), \nu(\mu), \xi(\mu)$, and $\sigma(\mu)$, where $\alpha(\mu), \gamma(\mu), \nu(\mu)$ are continuously differentiable, if the following properties hold:*

(SCF1) Convexity and differentiability. $\psi(x, \mu)$ is convex in x, continuous in $(x, \mu) \in \Omega \times \mathbb{R}_+$, thrice continuously differentiable in x, and twice continuously differentiable in μ.

(SCF2) Self-concordance of members. For any $\mu > 0$, $\psi(x, \mu)$ is $\alpha(\mu)$-self-concordant on Ω.

(SCF3) Compatibility of neighbors. For every $(x, \mu) \in \Omega \times \mathbb{R}_+$ and $h \in \mathbb{R}^n$, $\nabla_x \psi(x, \mu)$, $\nabla_x^2 \psi(x, \mu)$ are continuously differentiable in μ, and

$$|\{h^T \nabla_x \psi(x, \mu)\}' - \{ln\, \nu(\mu)\}' h^T \nabla_x \psi(x, \mu)| \le \xi(\mu) \alpha(\mu)^{\frac{1}{2}} (h^T \nabla_x^2 \psi(x, \mu) h)^{\frac{1}{2}} \tag{6}$$

$$|\{h^T \nabla_x^2 \psi(x, \mu) h\}' - \{ln\, \gamma(\mu)\}' h^T \nabla_x^2 \psi(x, \mu) h| \le 2\sigma(\mu) h^T \nabla_x^2 \psi(x, \mu) h. \tag{7}$$

We will call the parameter $\xi(\mu)$ the complexity value of a self-concordant family.

□

The definition of a self-concordant function family defined by Nesterov and Nemirovskii [11] is more general than the above definition. For our development, the above simpler definition suffice. We can scale an α-self-concordant function by a factor $\frac{1}{\alpha}$ to get an 1-self-concordant function. In the following we use abbreviation self-concordant for non-degenerate and strongly self-concordant, and barrier for non-degenerate and strongly self-concordant-barrier. Non-degeneracy requires a barrier function to be a strictly convex function. Nesterov and Nemirovskii [11] show the existence of an universal barrier function for any closed convex set. The parameter ϑ is called the complexity value of a self-concordant barrier. Appropriate barriers for the linear, quadratic, semi-definite, and second-order cone problems are well known. In particular, for R^n_+, $B(x) = -ln \sum_{i=1}^n ln\, x_i$ with complexity value $\vartheta = n$; for the second-order cone $K^2_n = \{(t,x) \in R^{n+1} \mid t \geq \|x\|_2\}$, $B(x) = -ln\,(t^2 - |x|^2_2)$ with $\vartheta = 2$; and for the cone of symmetric positive semi-definite $n \times n$ matrices $X \in S^+_n$, $B(x) = -ln(det(X))$ with $\vartheta = n^2$. More recently, Faybusovich [3, 4] has given self-concordant barriers for cones generated by Tchebychev systems and report experiments with such barriers. Recently Ariyawansa and Zhu [1] have given a volumetric center based algorithm for two-stage semi-definite programs. Since the Vaidya volumetric barriers are another example of a self-concordant barrier, the analysis of this paper also applies to the volumetric barrier case.

2.2 Properties of Two-Stage Barrier Recourse Function

In this section we show that the barrier recourse function $\eta^k(x,\mu)$ is differentiable in x and μ, convex in x, and concave in μ; it is self-concordant, and forms a self-concordant family for $\mu > 0$. Since we study common properties of any barrier recourse function $\eta^k(x)$ for $k = 1,\ldots,K$, we drop the superscript k and represent a barrier recourse function as:

$$\rho(x,\mu) \quad = \quad min\,\{r(y) \mid Qy = q + Tx\}, \tag{8}$$

where

$$r(y) = d^T y + \mu B(y). \tag{9}$$

Let y^* represent the optimal solution of (8) and u^* be the corresponding Lagrangian multiplier for a given μ and x. Since $r(y)$ is strictly convex, y^* is unique. The Lagrangian multiplier u^* is also unique because the KKT conditions (11) have a unique solution since Q is a full row rank matrix. Let

$$F(x,y,u,\mu) \equiv \begin{pmatrix} d + \mu \nabla_y B(y) + Q^T u \\ Qy - q - Tx \end{pmatrix}. \tag{10}$$

For a fixed x and μ the optimal solution (y^*, u^*) satisfies the first order KKT conditions:

$$F(x,y,u,\mu) = 0. \tag{11}$$

The Lagrangian function for the second stage problem is

$$L(x,y,u,\mu) \quad = \quad d^T y + \mu B(y) + u^T Q y - u^T q - u^T T x. \tag{12}$$

Throughout the paper y^* and u^* means the optimal primal and dual solutions, and we write $y^*(x)$ to stress that y^* is a function of x for any fixed $\mu > 0$ (we shall prove this later);

similarly $y^*(\mu)$ means y^* is a function of μ for any fixed x; this convention extends to $u^*(x)$, $u^*(\mu)$, $\rho(x)$, and $\rho(\mu)$ as well. We also let

$$g \equiv \nabla_y B(y),\ H \equiv \nabla_y^2 B(y),\ R \equiv QH(y)^{-\frac{1}{2}}, \tag{13}$$

where $H^{-1/2}(y)$ represent the inverse of the square-root of matrix $H(y)$. This square-root exist since $H(y)$ is a positive definite matrix.

From (11) it is easy to see that

$$u^* = -\left(QH^{-1}Q^T\right)^{-1} QH^{-1}\nabla_y r(y^*) = (RR^T)^{-1}RH^{-1/2}\nabla_y r(y^*) \tag{14}$$

2.2.1 Convexity and Self-Concordance of the Barrier Recourse Function of x

In this section we show that $\rho(x)$ is a convex and twice continuously differentiable function for a given $\mu > 0$.

Proposition 2.1 *For any $\mu > 0$, the barrier recourse function $\rho(x)$ is convex in x.*

Proof: Let $\alpha^1 + \alpha^2 = 1$, $\alpha^1, \alpha^2 \geq 0$, and $x^1, x^2 \in intG$. Now,

$$\begin{aligned}
\alpha^1\rho(x^1) + \alpha^2\rho(x^2) &= d^T(\alpha_1 y^*(x^1) + \alpha^2 y^*(x^2)) + \alpha^1\mu B(y^*(x^1)) + \alpha_2\mu B(y^*(x^2)) \\
&\geq d^T(\alpha^1 y^*(x^1) + \alpha^2 y^*(x^2)) + \mu B(\alpha^1 y^*(x^1) + \alpha^2 y^*(x^2)) \\
&\geq \rho(\alpha^1 x^1 + \alpha^2 x^2).
\end{aligned}$$

The first inequality holds because of the convexity of $B(\cdot)$ and the second inequality holds since $\alpha^1 y^*(x^1) + \alpha^2 y^*(x^2)$ is feasible for $Qy = q + T(\alpha^1 x^1 + \alpha^2 x^2)$. □

Lemma 2.1 *For each fixed $\mu > 0$ the optimal solution $y^*(x)$ and the Lagrangian multiplier $u^*(x)$ are differentiable functions of x. In particular,*

$$\begin{pmatrix} \nabla_x y^*(x) \\ \nabla_x u^*(x) \end{pmatrix} = \begin{pmatrix} H(x)^{-\frac{1}{2}} R(x)^T (R(x)R(x)^T)^{-1} T \\ -\mu \left(R(x)R(x)\right)^{-1} T \end{pmatrix}. \tag{15}$$

Proof: Let $H := H(x)$ and $R := R(x)$. From (11) we have

$$\nabla_{(y,u,x)}F(y,u,x) = \begin{pmatrix} \nabla_{(y,u)}F(y,u,x) & \vdots & \nabla_x F(y,u,x) \end{pmatrix} = \begin{pmatrix} \mu H & Q^T & \vdots & 0 \\ Q & 0 & \vdots & -T \end{pmatrix}. \tag{16}$$

The matrix $\nabla_{(y,u)}F(y,u,x)$ is invertible since H is positive definite and Q has full row rank. Its inverse is expressed as:

$$\nabla_{(y,u)}^{-1}F(y,u,x) = \begin{pmatrix} \frac{1}{\mu}H^{-1} - \frac{1}{\mu}H^{-\frac{1}{2}}R^T(RR^T)^{-1}RH^{-\frac{1}{2}} & H^{-\frac{1}{2}}R^T(RR^T)^{-1} \\ (RR^T)^{-1}RH^{-\frac{1}{2}} & -\mu\left(RR^T\right)^{-1} \end{pmatrix}. \tag{17}$$

Since the conditions of Implicit Function Theorem [15] are satisfied at (y^*, u^*), we have

$$\begin{pmatrix} \nabla_x y^*(x) \\ \nabla_x u^*(x) \end{pmatrix} = -[\nabla_{(y,u)}F(y^*,u^*,x)]^{-1}\nabla_x F(y^*,u^*,x),$$

giving us the desired result. □

Lemma 2.2 *Let $x \in intG \subset \mathbb{R}^n, p \in \mathbb{R}^n$, $\varrho(t) := \rho(x+tp)$, then we have*

$$\nabla_x \rho(x) = -T^T u^*(x) = -T^T (R(x)R(x)^T)^{-1} R(x) H(x)^{-1/2} \nabla_y r(y^*) = \nabla_x y^*(x)^T \nabla_y r(y^*),$$
$$\nabla_x^2 \rho(x) = \mu T^T \left(R(x)R^T(x)\right)^{-1} T,$$
$$\varrho'(0) = \nabla_x \rho(x)^T p = \nabla_y r(y^*)[s],$$
$$\varrho''(0) = \nabla^2 \rho(x)[p,p] = \mu \nabla_y^2 B(y^*)[s,s],$$
$$\varrho'''(0) = \nabla_x^3 \rho(x)[p,p,p] = \mu \nabla_y^3 B(y^*)[s,s,s],$$

where $\nabla_x y^*(x) = H(x)^{-1/2} R(x) \left(R(x)R(x)^T\right)^{-1} T$ and $s = \nabla_x y^*(x)p$.

Proof: Because of strong duality we have $\rho(x) = L(x, y^*(x), u^*(x), \mu)$. Hence, by using chain rule in (12) and first order KKT conditions (11) we have,

$$\begin{aligned}
\nabla_x \rho(x) &= \nabla_x L(x, y^*(x), u^*(x)) \\
&= (d^T + u^*(x)^T Q) \nabla_x y^*(x) + \mu \nabla_x B(y^*(x)) + \nabla_x u^*(x)^T (Qy^*(x) - q - Tx) - T^T u^*(x) \\
&= (d^T + \mu \nabla_y B(y^*(x)) + u^*(x)^T Q) \nabla_x y^*(x) - T^T u^*(x) \\
&= -T^T u^*(x) \qquad (18) \\
&= \nabla_x y^*(x) \nabla_y r(y^*) \ \text{(using (14) and (15))} \qquad (19)
\end{aligned}$$

By differentiating (18), and using (15) we have

$$\begin{aligned}
\nabla_x^2 \rho(x) &= -T^T \nabla_x u^*(x) \qquad (20) \\
&= \mu T^T \left(R(x)R^T(x)\right)^{-1} T. \qquad (21)
\end{aligned}$$

Hence, $\varrho''(t) = p^T \nabla_x^2 \rho(x+tp) p = \mu T^T \left(R(t)R^T(t)\right)^{-1} T$, where $y^*(t) := y^*(x+tp)$, $H(t) := H(y^*(t))$, and $R(t) := QH^{\frac{1}{2}}(t)$. In particular, at $t=0$ we have $\varrho''(0) = \mu p^T T^T \left(RR^T\right)^{-1} Tp = \mu \nabla_y^2 B(y^*)[\nabla y^*(x)p, \nabla y^*(x)p]$, where the last equality uses Lemma 2.1. Now recall that $RR^T = QH^{-1}Q^T = Q\left[\nabla_y^2 B\left(y^*\right)\right]^{-1} Q^T$. Hence, we have

$$\begin{aligned}
\varrho'''(t) &= \mu p^T T^T \left(R(t)R^T(t)\right)^{-1} (R(t)R^T(t))' \left(R(t)R^T(t)\right)^{-1} Tp \\
&= \mu p^T T^T \left(R(t)R^T(t)\right)^{-1} QH^{-1}(t) H(t)' H^{-1}(t) Q^T \left(R(t)R^T(t)\right)^{-1} Tp.
\end{aligned}$$

Now, at $t=0$ by using Lemma 2.1 we have

$$\begin{aligned}
\varrho'''(0) &= \mu p^T \nabla_x y^*(x)^T H(t)'|_{t=0} \nabla_x y^*(x) p \\
&= \mu \nabla_x^3 B(y^*)[\nabla_x y^*(x)p, \nabla_x y^*(x)p, \nabla_x y^*(x)p]. \square \qquad (22)
\end{aligned}$$

We now establish that for a fixed $\mu > 0$, $\rho(x, \mu)$ is self-concordant.

Theorem 2.1 *For each $\mu > 0$, the barrier recourse function $\rho(x)$ is μ-self-concordant.*

Proof: By Lemma 2.2 and the fact that B is a self-concordant barrier we have

$$\begin{aligned}
\left|\nabla^3 \rho(x)[p,p,p]\right| &= \left|\mu \nabla_y^3 B(y^*)[\nabla y^*(x)p, \nabla y^*(x)p, \nabla y^*(x)p]\right| \\
&\leq 2\mu \left(\nabla_y^2 B(y^*)[\nabla y^*(x)p, \nabla y^*(x)p]\right)^{\frac{3}{2}} \\
&= 2\mu \left(p^T T^T \left(R(x)R^T(x)\right)^{-1} Tp\right)^{\frac{3}{2}} \\
&= 2\mu^{-\frac{1}{2}} \left(\nabla_x^2 \rho(x)[p,p]\right)^{\frac{3}{2}}. \square
\end{aligned}$$

2.2.2 Differentiability and Concavity of the Recourse Function of μ

In this section we bound the second derivative of $\rho(\mu)$ for a given x .

Lemma 2.3 *For an $x \in intG$, optimal second stage solution $y^*(\mu)$, and Lagrangian multiplier $u^*(\mu)$ are well defined differentiable functions of μ, $\mu > 0$. Furthermore,*

$$\begin{pmatrix} y^*(\mu)' \\ u^*(\mu)' \end{pmatrix} = -\begin{pmatrix} \frac{1}{\mu}H(\mu)^{-\frac{1}{2}}\left(I - R(\mu)^T(R(\mu)R(\mu)^T)^{-1}R(\mu)\right)H(\mu)^{-\frac{1}{2}}g(\mu) \\ (R(\mu)R(\mu)^T)^{-1}R(\mu)H(\mu)^{-\frac{1}{2}}g(\mu) \end{pmatrix},$$

and

$$\{\nabla_x \rho(x,\mu)\}_\mu^{'} = T^T(R(x,\mu)R^T(x,\mu))^{-1}R(x,\mu)H^{-\frac{1}{2}}(x,\mu)g(x,\mu).$$

Proof: Similar to the proof of Lemma 2.1, we have an equation system

$$F(y,u,\mu) = 0.$$

Differentiating $F(y,u,\mu)$ with respective to μ we get

$$\nabla_{(y,u,\mu)}F(y,u,\mu) = \begin{pmatrix} \nabla_{(y,\mu)}F(y,u,\mu) & \vdots & \nabla_x F(y,u,\mu) \end{pmatrix} = \begin{pmatrix} \mu H(\mu) & Q^T & \vdots & g(\mu) \\ Q & 0 & \vdots & 0 \end{pmatrix}. \quad (23)$$

Now by applying the Implicit Function Theorem we have

$$\begin{pmatrix} y^*(\mu)' \\ u^*(\mu)' \end{pmatrix} = -\begin{pmatrix} \frac{1}{\mu}H(\mu)^{-\frac{1}{2}}\left(I - R(\mu)^T(R(\mu)R(\mu)^T)^{-1}R(\mu)\right)H(\mu)^{-\frac{1}{2}}g(\mu) \\ (R(\mu)R(\mu)^T)^{-1}R(\mu)H(\mu)^{-\frac{1}{2}}g(\mu) \end{pmatrix}. \quad (24)$$

From (18) we have $\{\nabla_x \rho(x,\mu)\}_\mu^{'} = \left\{-T^T u^*(\mu)\right\}_\mu^{'}$ which gives the desired result from using (24). □

Lemma 2.4 *For an $x \in intG$ the barrier recourse function $\rho(\mu)$ is a concave function and it satisfies*

$$-\rho(\mu)'' \leq \frac{\vartheta}{\mu}.$$

Proof: Because of strong duality we have $\rho(\mu) = L(x, y^*(\mu), u^*(\mu), \mu)$ for all $\mu > 0$. Hence,

$$\begin{aligned} \rho(\mu)' &= L(x, y^*(\mu), u^*(\mu))' \\ &= (d^T + u^*(\mu)^T Q + \mu g(\mu))y^*(\mu)' + B(y^*(\mu)) + u^*(\mu)'^T(Qy^*(\mu) - q - Tx) \\ &= B(y^*(\mu)). \end{aligned} \quad (25)$$

By differentiating (25) and applying (24) we get

$$\begin{aligned} -\rho(\mu)'' &= -\left(\nabla_y B(y^*)\right)^T y^*(\mu)' \\ &= g(\mu)^T\left(\frac{1}{\mu}H(\mu)^{-1}g(\mu) - \frac{1}{\mu}H(\mu)^{-\frac{1}{2}}R(\mu)^T(R(\mu)R(\mu)^T)^{-1}R(\mu)H(\mu)^{-\frac{1}{2}}g(\mu)\right) \\ &= \frac{1}{\mu}g(\mu)^T H(\mu)^{-\frac{1}{2}}\left(I - R(\mu)^T(R(\mu)R(\mu)^T)^{-1}R(\mu)\right)H(\mu)^{-\frac{1}{2}}g(\mu) \quad (26) \\ &\leq \frac{1}{\mu}g(\mu)^T H(\mu)^{-1}g(\mu) \leq \frac{\vartheta}{\mu}. \quad (27) \end{aligned}$$

The first inequality above holds since $(I - R^T(RR^T)^{-1}R)$ is a projection matrix. The second inequality holds because $B(y)$ is a self-concordant barrier with complexity value ϑ. The concavity of $\rho(\mu)$ follows from (26) because $H(\mu)$ is a positive definite matrix. □

2.3 Self-Concordant Family of the Barrier Recourse Function

In this section we show that when both x and μ vary, $\{\rho(x,\mu), \mu > 0\}$ is a self-concordant family as defined below.

Proposition 2.2 *The second stage barrier objective function $r(y,\mu), \mu > 0$ forms a self-concordant family with parameters $\alpha(\mu) = \mu, \gamma(\mu) = \nu(\mu) = 1, \xi(\mu) = \frac{\sqrt{\vartheta}}{\mu}, \sigma(\mu) = \frac{1}{2\mu}$.* □

A proof of Proposition 2.2 is straightforward. Next we show that the barrier recourse function is self-concordant family.

Theorem 2.2 *The barrier recourse function $\{\rho(x,\mu), \mu > 0\}$ is a self-concordant family with parameters: $\alpha(\mu) = \mu, \gamma(\mu) = \nu(\mu) = 1, \xi(\mu) = \frac{\sqrt{\vartheta}}{\mu}, \sigma(\mu) = \frac{1}{2\mu}$. In particular, for any h,*

$$\left|h^T \nabla_x \rho(x,\mu)'\right| \leq \sqrt{\frac{\vartheta}{\mu}}\sqrt{h^T \nabla_x^2 \rho(x,\mu) h}, \text{ and}$$

$$\left|\{h^T \nabla_x^2 \rho(x,\mu) h\}'_\mu\right| \leq \frac{1}{\mu} h^T \nabla_x^2 \rho(x,\mu) h.$$

Proof: The conditions (SCF1) and (SCF2) follow from Theorem 2.1. We now show (SCF3). For simplicity we use $R := R(x,\mu)$, $H := H(x,\mu)$ and $g := g(x,\mu)$. From Lemma 2.3 we have

$$\begin{aligned}
\left|h^T \left\{\nabla_x \rho(x,\mu)\right\}'_\mu\right| &= \left|h^T T^T (RR^T)^{-1} R H^{-\frac{1}{2}} g\right| \\
&\leq \sqrt{h^T T^T (RR^T)^{-1} T h}\sqrt{g^T H^{-\frac{1}{2}} R^T (RR^T)^{-1} R H^{-\frac{1}{2}} g} \\
&\leq \sqrt{\frac{1}{\mu} h^T \nabla_x \rho^2(x) h}\sqrt{g^T H^{-1} g} \\
&\leq \sqrt{\frac{\vartheta}{\mu}}\sqrt{h^T \nabla_x \rho^2(x) h}. \qquad (28)
\end{aligned}$$

The second inequality above uses (21) and the fact that $R(RR^T)^{-1}R$ is an orthogonal projection matrix; the last inequality holds since $B(\cdot)$ is a self-concordant barrier of complexity value ϑ.

This proves (6) for $\nu(\mu)=1, \xi(\mu)=\frac{\sqrt{\vartheta}}{\mu}$, and $\alpha(\mu)=\mu$. Now,

$$
\begin{aligned}
& \left|\left\{h^T \nabla_x^2 \rho(x,\mu) h\right\}_\mu^{'}\right| \\
= & \left|h^T\left\{\mu T^T\left(RR^T\right)^{-1} T\right\}_\mu^{'} h\right| \\
= & \left|h^T\left\{T^T\left(\frac{1}{\mu} RR^T\right)^{-1} T\right\}_\mu^{'} h\right| \\
= & \left|h^T T^T\left(\frac{1}{\mu} RR^T\right)^{-1}\left(\frac{1}{\mu} RR^T\right)_\mu^{'}\left(\frac{1}{\mu} RR^T\right)^{-1} T h\right| \\
= & \left|h^T T^T\left(\frac{1}{\mu} RR^T\right)^{-1}\left\{Q(\mu H)^{-1} Q^T\right\}_\mu^{'}\left(\frac{1}{\mu} RR^T\right)^{-1} T h\right| \\
= & \left|h^T T^T\left(\frac{1}{\mu} RR^T\right)^{-1} Q(\mu H)^{-1}\{\mu H\}_\mu^{'}(\mu H)^{-1} Q^T\left(\frac{1}{\mu} RR^T\right)^{-1} T h\right| \\
= & \left|\bar{h}^T\left\{\nabla_y^2 r(y^*)\right\}_\mu^{'} \bar{h}\right| \leq \frac{1}{\mu} \bar{h}^T \nabla_y^2 r(y^*) \bar{h}=\frac{1}{\mu} h^T \nabla_x^2 \rho(x,\mu) h,
\end{aligned} \tag{29}
$$

where $\bar{h}=H^{-1}Q^T\left(RR^T\right)^{-1}Th$. The inequality above follows from Proposition 2.2. The last equality follows by substituting for $\bar{h}$ and $\nabla_x^2\rho(x,\mu)$. Hence, the inequality (7) holds for $\rho(x,\mu)$ for parameter functions $\gamma(\mu)=1, \sigma(\mu)=1/2\mu$.□

3 Properties of the First Stage Objective Function

In Section 2 we studied properties of the barrier recourse function $\rho(x,\mu)$. For a K scenario problem, each of the K barrier recourse functions satisfy: (i) the optimal solution y^{k^*} and optimal Lagrangian multiplier u^{k^*} of (3) are continuously differentiable functions of x or μ; (ii) the functions $\eta^k(x,\mu), k=1,\ldots,K$ are strictly convex in x, and concave in μ; (iii) the functions $\eta^k(x,\mu), k=1,\ldots,K$ are twice continuously differentiable in x or μ, and continuously differentiably in (x,μ); (iv) the functions $\eta^k(x,\mu), k=1,\ldots,K$ are a self-concordant family. We now analyze the self-concordance properties of the first stage function:

$$
f(x,\mu)=c^T x+\mu b(x)+\sum_{k=1}^{K} \eta^k(x,\mu). \tag{30}
$$

In particular, we show that this first stage objective function is also a self-concordant family.

We use the following notations addressing scenarios consistently with our discussion of the properties of $\rho(\cdot)$ in Section 2.2. $\eta^k(\cdot)$,$u^{k^*}(\cdot)$,y^{k^*},$g^k(\cdot)$,$H^k(\cdot)$,$R^k(\cdot)$,$F^k(\cdot)$ and $L^k(\cdot)$ are functions of x and/or $\mu, \forall k=1,\ldots,K$. We let $g^0(\cdot)=\nabla b(\cdot)$ and $H^0(\cdot)=\nabla^2 b(\cdot)$, $g(\cdot)=\nabla f(\cdot)$, $H(\cdot)=\nabla^2 f(\cdot)$, and $R(\cdot)=AH(\cdot)^{-\frac{1}{2}}$. We see immediately that $g(x)=\mu\nabla_x b(x)+\sum_{k=1}^{K}\nabla_x\eta(x)$, and $H(x)=\mu\nabla_x^2 b(x)+\sum_{k=1}^{K}\nabla_x^2\eta(x)$. It is useful to represent these summation in matrix notation, and for this purpose we define $\hat{I}=[I,\ldots,I]$, $\bar{h}^T=h^T\hat{I}$, $\bar{g}^T(x,\mu)=\left[g^{0^T}(x),\ldots,g^{K^T}(x,\mu)\right]$,

$D^0(x) = \sqrt{\mu} H^{0^{\frac{1}{2}}}(x)$, $\tilde{D}^0 = \frac{1}{\sqrt{\mu}} H^{0^{-\frac{1}{2}}}(x)$, $D^i(x,\mu) = \sqrt{\mu} T^{i^T} \left(R^i(x,\mu) R^{i^T}(x,\mu) \right)^{-\frac{1}{2}}$, $\tilde{D}^i(x,\mu) = \frac{1}{\sqrt{\mu}} (R^i(x,\mu) R^{i^T}(x,\mu))^{-\frac{1}{2}} R^i(x,\mu) H^{i^{-\frac{1}{2}}}(x,\mu)$, $\forall i = 1,\ldots,K$, and

$$D(x,\mu) = blkdiag(D^0(x,\mu),\ldots,D^K(x,\mu)), \tilde{D}(x,\mu) = blkdiag(\tilde{D}^0(x,\mu),\ldots,\tilde{D}^K(x,\mu)). \quad (31)$$

Here *blkdiag* represents a block diagonal diagonal matrix. Let

$$\tilde{\vartheta} = \sum_{i=0}^{K} \vartheta^i.$$

Corollary 3.1 *The first stage objective function $f(x,\mu)$ is strictly convex in x, concave in μ, thrice differentiable in x, twice continuously differentiable in μ, and continuously differentiable in (x,μ). Furthermore, the optimal solutions $x^*(\mu)$ and $u^*(\mu)$ are continuously differentiable in μ. Specifically, we have*

$$H(x,\mu) = \mu \nabla_x^2 b(x) + \sum_{k=1}^{K} \nabla_x^2 \eta(x) = \mu \nabla_x^2 b(x) + \mu \sum_{k=1}^{K} T^{k^T} \left(R^k(x,\mu) R^{k^T}(x,\mu) \right)^{-1} T^k = \hat{I} D D^T \hat{I}^T,$$

$$\{\nabla_x f(x,\mu)\}_\mu' = \mu g^0(x) + \sum_{k=1}^{K} T^{k^T} (R(x,\mu) R(x,\mu)^T)^{-1} R(x,\mu) H(x,\mu)^{-1/2} g^k(x,\mu) = \hat{I} D \tilde{D} \bar{g},$$

$$\{f(x,\mu)\}_\mu'' \in [-\frac{\tilde{\vartheta} - \vartheta^0}{\mu}, 0]$$

$$\begin{pmatrix} x^*(\mu)' \\ u^*(\mu)' \end{pmatrix} = - \begin{pmatrix} \frac{1}{\mu} H(\mu)^{-\frac{1}{2}} (I - R(\mu)^T (R(\mu) R(\mu)^T)^{-1} R(\mu)) H(\mu)^{-\frac{1}{2}} g(\mu) \\ (R(\mu) R(\mu)^T)^{-1} R(\mu) H(\mu)^{-\frac{1}{2}} g(\mu) \end{pmatrix}. \quad (32)$$

Proof. Conclusions about $H(x,\mu)$, $\{\nabla_x f(x,\mu)\}_\mu'$, and $\{f(x,\mu)\}_\mu''$ follow from Lemma 2.2, 2.3, and 2.4, and straightforward linear algebra. Let $F(x,u,\mu) = \begin{pmatrix} \nabla_x f(x,\mu) + A^T u \\ Ax - b \end{pmatrix}$. At the optimum solution $x^*(x,\mu), u^*(x,\mu)$ from KKT conditions of (2) we know that $F(x^*(x,\mu), u^*(x,\mu), \mu) = 0$. Since,

$$\nabla_{(x,u,\mu)} F = \begin{pmatrix} \nabla_{(x,u)} F & \vdots & \nabla_\mu F \end{pmatrix} = \begin{pmatrix} \nabla_x^2 f(x,\mu) & A^T & \vdots & \{\nabla_x f(x,\mu)\}' \\ A & 0 & \vdots & 0 \end{pmatrix},$$

and the matrix $\nabla_{(x,\mu)} F$ is invertible, (32) follows from using the Implicit Function Theorem and straightforward computation. □

Lemma 3.1 *The function $f(x,\mu)$ is μ-self-concordant.*

Proof. Using Lemma 2.2, Theorem 2.1, and the assumption that the first stage barrier is 1-self-concordant we have

$$\begin{aligned}
\left|\nabla_x^3 f(x)[p,p,p]\right|^{\frac{1}{3}} &= \left|\mu\nabla_x^3 b(x)[p,p,p] + \mu\sum_{k=1}^{K}\nabla_y^3 B(y^{k^*})[s,s,s]\right|^{\frac{1}{3}} \quad (s := \nabla_x y^{k^*}(x)p) \\
&\le \mu^{\frac{1}{3}}\left(\left(\nabla_x^3 b(x)[p,p,p]\right)^{\frac{2}{3}} + \sum_{k=1}^{K}\left(\nabla_y^3 B(y^{k^*})[s,s,s]\right)^{\frac{2}{3}}\right)^{\frac{1}{2}} \quad (\text{using } \|\cdot\|_3 \le \|\cdot\|_2) \\
&\le \mu^{\frac{1}{3}}\left(\left(2\left(\nabla_x^2 b(x)[p,p]\right)^{\frac{3}{2}}\right)^{\frac{2}{3}} + \sum_{k=1}^{K}\left(2\left(\nabla_y^2 B(y^{k^*})[s,s]\right)^{\frac{3}{2}}\right)^{\frac{2}{3}}\right)^{\frac{1}{2}} \\
&= (2\mu)^{\frac{1}{3}}\left(\nabla_x^2 b(x)[p,p] + \sum_{k=1}^{K}\nabla_y^2 B(y^{k^*})[s,s]\right)^{\frac{1}{2}} = \left(\frac{2}{\sqrt{\mu}}\right)^{\frac{1}{3}}\left(\nabla_x^2 f(x)[p,p]\right)^{\frac{1}{2}}.
\end{aligned}$$

The last inequality above uses that the barriers are self-concordant functions. □

We now show that the first stage objective functions forms a self-concordant family for $\mu > 0$.

Theorem 3.1 *The first stage barrier recourse functions* $\{f(x,\mu), \mu > 0\}$ *form a self-concordant family with parameters:* $\alpha(\mu) = \mu, \gamma(\mu) = \nu(\mu) = 1, \xi(\mu) = \frac{\sqrt{\tilde{\vartheta}}}{\mu}, \sigma(\mu) = \frac{1}{2\mu}$.

Proof: (SCF1) and (SCF2) are shown in Corollary 3.1 and Lemma 3.1. We now prove the two inequalities (6–7) of (SCF3). By Corollary 3.1 we have

$$\begin{aligned}
\left|h^T\left\{\nabla_x f(x)\right\}'_\mu\right| &= \left|\bar{h}^T D\tilde{D}\bar{g}\right| \\
&\le \sqrt{\bar{h}^T D D^T \bar{h}}\sqrt{\bar{g}^T\tilde{D}^T\tilde{D}\bar{g}} \\
&= \sqrt{h^T\left(\mu H^0 + \sum_{k=1}^{K}\mu T^{k^T}\left(R^k R^{k^T}\right)^{-1}T^k\right)h}\;\cdot \\
&\quad \sqrt{\frac{1}{\mu}g^{0^T}H^{0^{-1}}g^0 + \frac{1}{\mu}\sum_{k=1}^{K}g^{k^T}H^{k^{-\frac{1}{2}}}R^{k^T}\left(R^k R^{k^T}\right)^{-1}R^k H^{k^{-\frac{1}{2}}}g^k} \\
&\le \sqrt{h^T\nabla_x^2 f(x)h}\sqrt{\frac{1}{\mu}g^{0^T}H^{0^{-1}}g^0 + \frac{1}{\mu}\sum_{k=1}^{K}g^{k^T}H^{k^{-1}}g^k} \qquad (33) \\
&\le \sqrt{h^T\nabla_x^2 f(x)h}\sqrt{\frac{1}{\mu}\left(\vartheta^0 + \sum_{k=1}^{K}\vartheta^k\right)} = \sqrt{\frac{\tilde{\vartheta}}{\mu}}\sqrt{h^T\nabla_x^2 f(x)h} \qquad (34)
\end{aligned}$$

Here (33) used Corollary 3.1 and the fact that $R^{k^T}\left(R^k R^{k^T}\right)^{-1}R^k$ is an orthogonal projection matrix. The inequality in (34) uses the definition of the complexity value of self-concordant barriers $b(x)$, $B^k(y^k), k = 1, \dots K$. This proves (6) for our choice of parameters. Now by using (29) from Theorem 2.2 for each $\eta^k(x)$ we obtain

$$
\begin{aligned}
\left| h^T \left\{ \nabla_x^2 f(x) \right\}_\mu' h \right| &= \left| h^T H^0 h + \sum_{k=1}^K h^T \left\{ \nabla_x^2 \eta^k(x,\mu) \right\}_\mu' h \right| \\
&\leq \left| h^T H^0 h \right| + \sum_{k=1}^K \left| h^T \left\{ \nabla_x^2 \eta^k(x,\mu) \right\}_\mu' h \right| \\
&\leq h^T H^0 h + \sum_{k=1}^K \frac{1}{\mu} h^T \nabla_x^2 \eta^k(x,\mu) h \\
&= \frac{1}{\mu} h^T \nabla_x^2 f(x) h.
\end{aligned}
$$

This proves the second inequality (7) in the comparability with neighbors condition (SCF3) in Defintion 2.2. □

4 Interior Decomposition Algorithms For Two-Stage Convex Stochastic Programs

We now analyze short step and long step primal path following interior point methods. We show that both of these methods obtain an ϵ-optimal solution in polynomial number of first stage Newton iterations. This analysis assumes that $\nabla_x f(x,\mu)$ and $\nabla_x^2 f(x,\mu)$ are computed exactly, i.e., the second stage problems (3) are solved exactly. As a consequence the algorithms presented in this section are only prototype algorithms, and their complexity is presented in terms of first stage Newton iterations. Similar assumptions were made in [18, 8, 9], since this assumption considerably simplifies the analysis.

The Newton step Δx is defined as a solution of the system

$$
\begin{aligned}
\nabla_x^2 f(x,\mu)\Delta x - A^T \Delta u &= -\nabla_x f(x,\mu) \qquad (35) \\
A\Delta x &= 0. \qquad (36)
\end{aligned}
$$

Let $x^*(\mu)$ denote the optimal solution of (2) for a give μ. The set of solutions $\{x^*(\mu), \mu > 0\}$ is called the central path. The complexity analysis focuses on the upper bound of the number of inner loops per updating of μ. For this purpose we define:

$$
\begin{aligned}
\delta(x,\mu) &:= \sqrt{\frac{1}{\mu}\Delta x^T \nabla_x^2 f(x) \triangle x} = \sqrt{\frac{1}{\mu}\nabla_x f(x)^T [\nabla_x^2 f(x)]^{-1} \nabla_x^2 f(x)}. \qquad (37) \\
\phi(x,\mu) &:= f(x,\mu) - f(x^*(\mu),\mu) \\
\tilde{\triangle} x &:= x - x^*(\mu) \\
\tilde{\delta}(x,\mu) &:= \sqrt{\frac{1}{\mu}\tilde{\triangle} x^T \nabla^2 f(x) \tilde{\triangle} x}. \qquad (38)
\end{aligned}
$$

At a major iteration k both short and long-step algorithms generate suitable approximations of $x^*(\mu^{k+1})$ starting from a suitable approximation of $x^*(\mu^k)$. We assume that an initial solution x^0 is available that is a suitable approximation of $x^*(\mu^0)$ for an initial μ^0. The short and long step algorithms differ in the rate of decrement of μ. In the short step algorithm $\gamma = 1 - \sigma/\sqrt{\tilde{\vartheta}}$, $\sigma \leq .1$, while in long step algorithms γ is taken to be a small constant, say .1. The short step algorithm follows the central path more closely, and requires only one Newton iteration to obtain a suitable approximation of $x^*(\mu^{k+1})$. While the long-step algorithm in worst case requires $O(\tilde{\vartheta})$ Newton iterations to obtain the next suitable approximation, where $O(\tilde{\vartheta})$ depends on γ, in practice a significantly fewer number of iterations are observed [7]. Lemmas 4.1 and 4.2 below are from [11] Theorem 2.1.1 (i) and Theorem 2.3.3, which are central to the analysis of short and long-step primal interior point methods applied to functions forming a self-concordant family.

Lemma 4.1 *If $\delta(x,\mu) < 1$, $|\tau| \leq 1$, then for any $h, h_1, h_2 \in \mathbb{R}^n$,*

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

$$(ii) \quad |h_1^T [\nabla^2 f(x+\tau\Delta x,\mu) - \nabla^2 f(x,\mu)] h_2| \leq [(1-\tau\delta)^{-2} - 1] \sqrt{h_1^T \nabla^2 f(x,\mu) h_1} \sqrt{h_2^T \nabla^2 f(x,\mu) h_2}.$$

□

Lemma 4.2 *Let $\kappa = \frac{2-\sqrt{3}}{2}$, if $\delta(x,\mu) \geq 2\kappa$, then $f(x,\mu) - f(x + \frac{1}{1+\delta(x,\mu)}\Delta x, \mu) \geq \mu(\delta(x,\mu) - ln(1+\delta(x,\mu)))$; if $\delta(x,\mu) \leq 2\kappa$, then $\delta(x+\Delta x,\mu) \leq \left(\frac{\delta(x,\mu)}{1-\delta(x+\Delta x,\mu)}\right)^2 \leq \frac{\delta(x,\mu)}{2}$.* □

Since in Theorem 3.1 we have shown that $f(x,\mu)$ is a self-concordant family, the analysis of short step algorithm immediately follows from [11]. We state this result without proof.

Theorem 4.1 *Let μ^0 be the initial barrier parameter, and ϵ be the target precision. If $\delta(x^0,\mu^0) \leq \kappa = (2-\sqrt{3})/2$, and a short step algorithm reduces μ at a constant rate $\gamma = 1 - \sigma/\sqrt{\tilde{\vartheta}}$, $\sigma \leq 0.1$, then the short step algorithm terminates with a (x^k,μ) satisfying $\delta(x^k,\mu) \leq \kappa$, $\mu \leq \epsilon$ in $O(\sqrt{\tilde{\vartheta}} ln \mu^0/\epsilon)$ number of first stage Newton iterations.* □

Renegar [13, Section 2.4.1 (equation 2.14) and 2.4.2 (page 46 last paragraph)] shows that for a point satisfying the conditions of Theorem 4.1 and Theorem 4.2 the objective of the first stage problem $c^T x$ satisfies $c^T x - z^* \leq 1.2\tilde{\vartheta}\mu$, where z^* is the optimum objective value of the two-stage stochastic convex programming problem. Hence, it is sufficient to specify the termination criterion in these theorems using a value of μ.

4.1 Prototype Long Step Algorithm and Complexity

A prototype long-step algorithm for the two-stage stochastic convex program is given as Algorithm 1. In order to show the convergence of this algorithm we establish an upper bound on $\phi(x^k,\mu^{k+1})$, and a lower bound on the reduction of $f(x^k,\mu^{k+1})$ using a damped Newton step. This allows us to bound the number of Newton steps after each update of μ. These results are given in Lemma 4.3 and Lemma 4.4, respectively. Since the condition $\tilde{\delta}(x,\mu) < 1$ of Lemma 4.3 is not directly verifiable, we give an upper bound on $\tilde{\delta}(x,\mu)$ by $\delta(x,\mu)$ in Lemma 4.6. The lower bound on the reduction of objective value per Newton step is established in Lemma 4.2.

Algorithm 1 Prototype long step algorithm for two-stage stochastic convex problem

Initialize. Given $\epsilon, x^0, \mu^0, \gamma$ such that $\delta(x^0, \mu^0) \leq \kappa, \gamma \in (0, 1)$. Set $x = x^0$, $\mu = \mu^0$ and $k = 0$.

while $\mu^k > \epsilon$

$\mu^{k+1} = \gamma\mu^k$

while $\delta(x, \mu^{k+1}) > \kappa$

$x := x^k$

solve subproblems $\eta^i(x, \mu)$, $i = 1, \ldots, K$

compute the Newton direction Δx using (36)

update $x := x + \frac{1}{1+\delta(x,\mu)}\Delta x$ if $\delta(x, \mu^{k+1}) > 2\kappa$; otherwise $x := x + \Delta x$.

$x^{k+1} := x$

$k := k + 1$

Theorem 4.2 *Let μ^0 be the initial barrier parameter, and ϵ be the target precision. The long step algorithm needs at most $O(\tilde{\vartheta} ln\mu^0/\epsilon)$ damped Newton iterations to generate a point (x^k, μ) satisfying $\delta(x^k, \mu) \leq \kappa$ from a starting point x^0 satisfying $\delta(x^0, \mu^0) \leq \kappa = (2 - \sqrt{3})/2$ while reducing μ^0 to ϵ at a linear rate $0 < \gamma < 1$.*

Proof: In Algorithm 1 we would like (x^k, μ^k) returning to the central path at every major iteration, i.e. x^k satisfying $\delta(x^k, \mu^k) \leq \kappa, \forall k = 0, \ldots, K$. We update μ by a constant factor $\gamma \in (0, 1), \mu^{k+1} = \gamma\mu^k$. If $\delta(x^k, \mu^{k+1})$ is not less than κ, we start the kth inner loop to generate iterates $\hat{x}^0, \ldots, \hat{x}^M$, where $\hat{x}^0 \equiv x^k$ and $\hat{x}^M$ is the first iterate satisfying $\delta(x^{k+1}, \mu^{k+1}) \leq 2\kappa$. Then by Lemma 4.2, since $\delta(\hat{x}^M, \mu^{k+1}) \leq 2\kappa$, one full Newton step will restore closeness to central path, i.e, $\delta(\hat{x}^M + \Delta x, \mu^{k+1}) \leq \kappa$. Hence, we only need to bound M. Since $\forall m = 0, \ldots, M - 1, \delta(\hat{x}^m, \mu^{k+1}) \geq 2\kappa$, by Lemma 4.2 each damped Newton step reduces objective value by at least $\mu^{k+1}(\delta(\hat{x}^j, \mu^{k+1}) - ln(1 + \delta(\hat{x}^j, \mu^{k+1})) \geq \iota, j = 1, \ldots, M - 1$, where $\iota = \mu^{k+1}(2\kappa - ln(1 + 2\kappa)) > 0$. On the other hand, by Lemma 4.4 we know that $\phi(x^k, \mu^{k+1}) = f(x^k, \mu^{k+1}) - f(x^*(\mu^{k+1}), \mu^{k+1})$ is at most $O(\tilde{\vartheta})$. Hence, the number of inner loops is at most $O(\tilde{\vartheta})$. The conclusion follows since it is trivial to see that we need only $ln(\epsilon/\mu_0)/ln\gamma$ outer loops to reach the target precision ϵ. $\square$

Lemma 4.3 *Let $\tilde{\delta} := \tilde{\delta}(x, \mu) < 1$. Then,*

$$\begin{aligned} \phi(x, \mu) &\leq \mu\left(\frac{\tilde{\delta}}{1 - \tilde{\delta}} + ln(1 - \tilde{\delta})\right) \\ |\phi(x, \mu)'| &\leq -ln(1 - \tilde{\delta})\tilde{\vartheta}. \end{aligned}$$

Proof: By the Fundamental Theorem of Calculus

$$\phi(x, \mu) = \nabla_x f(x(\mu), \mu)^T \tilde{\triangle} x + \int_0^1 \int_0^\alpha \tilde{\Delta} x^T \nabla_x^2 f(x(\mu) + s\tilde{\Delta} x, \mu)^T \tilde{\Delta} x d_s d_\alpha. \quad (39)$$

From KKT conditions for the first stage problem we have $\nabla_x f(x(\mu),\mu)^T = {u^*}^T A$, and $A\tilde{\Delta}x = 0$, where u^* is the Lagrangian multiplier corresponding to the first stage equality constraints $Ax = b$. This gives

$$\nabla_x f(x(\mu),\mu)^T \tilde{\Delta}x = 0. \tag{40}$$

Furthermore, by Lemma 4.1(i) and that $x(\mu) = x - \tilde{\Delta}x$ following the definition of $\tilde{\Delta}x$ in (38), we have

$$\tilde{\Delta}x^T \nabla_x^2 f(x(\mu) + t\tilde{\Delta}x, \mu)\tilde{\Delta}x \le \frac{\mu\tilde{\delta}^2}{(1-\tilde{\delta}+t\tilde{\delta})^2}. \tag{41}$$

Hence, by using (40) and (41) in (39) we have

$$\begin{aligned}
\phi(x,\mu) &= \int_0^1 \int_0^\alpha \tilde{\Delta}x^T \nabla_x^2 f(x(\mu) + t\tilde{\Delta}x, \mu)\tilde{\Delta}x dt\, d\alpha \\
&\le \int_0^1 \int_0^\alpha \frac{\mu\tilde{\delta}^2}{(1-\tilde{\delta}+t\tilde{\delta})^2} dt\, d\alpha \\
&\le \left(\frac{\tilde{\delta}}{1-\tilde{\delta}} + ln(1-\tilde{\delta}) \right)\mu.
\end{aligned}$$

Now,

$$\begin{aligned}
\phi(x,\mu)' &= f(x,\mu)' - f(x(\mu),\mu)' - \nabla_x f(x(\mu),\mu)x(\mu)' \\
&= f(x,\mu)' - f(x(\mu),\mu)' - \left(A^T u^*\right)^T x(\mu)' \\
&= f(x,\mu)' - f(x(\mu),\mu)'.
\end{aligned} \tag{42}$$

The second equality uses the KKT conditions for the first stage problem. The last equality follows from using $Ax(\mu)' = 0$. Now by applying Fundamental Theorem of Calculus to $\phi(x,\mu)'$, and using Lemma 4.1 (ii) we have

$$\begin{aligned}
|\phi(x,\mu)'| &= \left| \int_0^1 \nabla_x f(x(\mu) + \alpha\tilde{\Delta}x, \mu)'^T \tilde{\Delta}x d_\alpha \right| \\
&\le \int_0^1 \sqrt{\tilde{\Delta}x^T \nabla_x^2 f(x(\mu) + \alpha\tilde{\Delta}x, \mu)\tilde{\Delta}x} \cdot \\
&\quad \sqrt{\nabla_x f(x(\mu) + \alpha\tilde{\Delta}x, \mu)'^T [\nabla^2 f(x(\mu) + \alpha\tilde{\Delta}x, \mu)]^{-1} \nabla_x f(x(\mu) + \alpha\tilde{\Delta}x, \mu)' d_\alpha} \\
&\le \int_0^1 \frac{\sqrt{\mu}\tilde{\delta}}{1-\tilde{\delta}+\alpha\tilde{\delta}} \sqrt{\frac{\tilde{\vartheta}}{\mu}} d_\alpha = -ln(1-\tilde{\delta})\sqrt{\tilde{\vartheta}}.
\end{aligned}$$

The last inequality uses (41), and Theorem 3.1, especially its first inequality of compatibility with neighbors. $\square$

Lemma 4.4 *For a given* $0 < \zeta < 1$, *if* $\tilde{\delta} := \tilde{\delta}(x,\mu) \le \zeta$, *then*

$$\phi(x,\mu^+) = f(x,\mu^+) - f(x(\mu^+),\mu^+) \le O(\tilde{\vartheta})\mu^+$$

Proof: By differentiating (42) we have

$$\begin{aligned} \phi(x,\mu)'' &= f(x,\mu)'' - f(x(\mu),\mu)'' - \nabla_x f(x(\mu),\mu)'^T x(\mu)' \\ &\leq -f(x(\mu),\mu)'' - \nabla_x f(x(\mu),\mu)'^T x(\mu)'. \end{aligned} \tag{43}$$

We bound $-f(x(\mu),\mu)''$ by applying (i) of Corollary 3.1 to obtain

$$-f(x(\mu),\mu)'' \leq \frac{\tilde{\vartheta} - \vartheta^0}{\mu}. \tag{44}$$

To bound $-\nabla_x f(x(\mu),\mu)'^T x(\mu)'$ we first use Corollary 3.1(ii), then using the fact that $P(x(\mu)) := I - R(x(\mu))^T(R(x(\mu))R(x(\mu))^T)^{-1}R(x(\mu))$ is an orthogonal projection matrix we get

$$\begin{aligned} -\nabla_x f(x(\mu),\mu)'^T x(\mu)' &= \nabla_x f(x(\mu),\mu)'^T H(x(\mu))^{-\frac{1}{2}} P(x(\mu)) H(x(\mu))^{-\frac{1}{2}} \nabla_x f(x(\mu),\mu)'/\mu \\ &\leq \nabla_x f(x(\mu),\mu)'^T H(x(\mu))^{-1} \nabla_x f(x(\mu),\mu)'/\mu \\ &\leq \frac{\tilde{\vartheta}}{\mu}, \end{aligned} \tag{45}$$

where the last inequality follows from Theorem 3.1 (6) and Lemma 3.1. Hence, we have

$$\phi(x,\mu)'' \leq \frac{2\tilde{\vartheta} - \vartheta^0}{\mu}. \tag{46}$$

Using Lemma 4.3 and (46) we have

$$\begin{aligned} \phi(x,\mu^+) &= \phi(x,\mu) + (\mu^+ - \mu)\phi(x,\mu)' + \int_{\mu^+}^{\mu}\int_{\alpha}^{\mu} \phi(t,x)'' dt\, d\alpha \\ &\leq \left(\frac{\tilde{\delta}}{1-\tilde{\delta}} + ln(1-\tilde{\delta})\right)\mu - (\mu - \mu^+)\sqrt{\tilde{\vartheta}} ln(1-\tilde{\delta}) + (2\tilde{\vartheta} - \vartheta^0)\int_{\mu^+}^{\mu}\int_{\alpha}^{\mu} \frac{1}{t} dt\, d\alpha \\ &\leq \left(\frac{\tilde{\delta}}{1-\tilde{\delta}} + ln(1-\tilde{\delta})\right)\mu - (\mu - \mu^+)\sqrt{\tilde{\vartheta}} ln(1-\tilde{\delta}) - (2\tilde{\vartheta} - \vartheta^0)(\mu - \mu^+) ln\gamma \end{aligned}$$

The conclusion follows since $\tilde{\delta} \leq \zeta$, $\mu^+ = \gamma\mu$, and ζ, γ are constants. □

Since $\tilde{\delta}$ in Lemma 4.3 is not directly computable, in the following Lemma we bound $\tilde{\delta}$ using $\delta(x,\mu)$. This is possible as a consequence of the following result.

Lemma 4.5 *[13, Theorem 2.2.3] Let $\psi(x)$ be a 1-self-concordant function, x^* be its minimizer, $x \in domain(\psi)$, and $\|x - x^*\| \leq 1$. Then, $x^+ = x - [\nabla^2\psi(x)]^{-1}\nabla\psi(x)$ satisfies $\|x^+ - x^*\|_x \leq \frac{\|x-x^*\|_x^2}{1-\|x-x^*\|_x}$, where $\|y\|_x^2 = y^T\nabla^2\psi(x)y$.*

Lemma 4.6 *If $\delta(x,\mu) \leq \frac{1}{6}$, then $\tilde{\delta}(x,\mu) \leq 2\delta(x,\mu)$.*

Proof: Let $x^+ = x + \Delta x$ be generated from x by taking a full Newton step. Since $\frac{1}{\mu}f(x,\mu)$ is a 1-self-concordant function and $x(\mu)$ is the minimizer of $f(x,\mu)$, applying Lemma 4.5 we have

$$\|x - x(\mu)\|_x \le \|x - x^+\|_x + \|x^+ - x(\mu)\|_x \le \|x - x^+\|_x + \frac{\|x - x(\mu)\|_x^2}{1 - \|x - x(\mu)\|_x}.$$

Since $\|x - x(\mu)\|_x = \tilde{\delta}$ and $\|x - x^+\|_x = \delta$ as defined in (37) and (38), we have for $\delta \in [0, \frac{1}{6}]$, $\delta \ge \tilde{\delta} - \frac{\tilde{\delta}^2}{1-\tilde{\delta}} \ge \frac{1}{2}\tilde{\delta}$.□

5 Multistage Stochastic Convex Problem

5.1 Multistage Barrier Recourse Problem

Let $\xi_t, t = 1,\ldots,T$ be a discrete stochastic process, where each ξ_t represents data of an optimization problem. Since the data of the first stage problem is known, ξ_1 has only one realization. Along any path $\xi_1,\ldots,\xi_{t-1}$, ξ_t has only finitely many realizations as well. Hence there are only finitely many $t-$stage scenarios considering all possible history. So we can order the set $\{(\xi_1,\ldots,\xi_t)\}$, or equivalently establish a one-to-one mapping between $\{(\xi_1,\ldots,\xi_t)\}$ and $\{1,\ldots,K(t)\}$ for any $t = 1,\ldots,T$, where $K(t)$ is the cardinality of the set $\{(\xi_1,\ldots,\xi_t)\}$. For example, a binary tree has $K(1) = 1, K(2) = 2, K(3) = 4, \ldots, K(t) = 2^t$. Hence a tuple $(t,k), k \le K(t)$ uniquely identifies a node in a scenario tree for the given mapping. For the purpose of uniquely identifying any node in a scenario tree, any such one-to-one mapping will suffice.

We use $a(t,k)$ to denote the ancestor of (t,k) and $\mathcal{D}_t^k$ to denote the set of s such that $(t+1,s)$ are direct descendants of (t,k). Since any t-stage scenario is a decedent of one of the $(t-1)$-stage scenarios, and $\mathcal{D}_t^k \cap \mathcal{D}_t^j = \emptyset, \forall k \ne j$, we have $\sum_{k=1}^{K(t)} |\mathcal{D}_t^k| = K(t+1)$. For convenience, a superscript tk means (t,k), i.e., x^{tk} means $x^{(t,k)}$. We use these two notations interchangeably. We define a T-stage problem as follows.

$$\bar{\eta}^{tk}(x^{a(t,k)}) := \mathit{min}\, c^{tk} \cdot x^{tk} + \sum_{s \in \mathcal{D}_t^k} \bar{\eta}^{(t+1,s)}(x^{tk}) \;\; s.t.\, x^{tk} \in G^{tk} \cap L^{tk}(x^{a(t,k)}) \qquad (47)$$
$$\forall t = 1,\ldots,T, k = 1,\ldots,K(t),$$

where G^{tk} are nonempty convex domains, $L^{tk} \equiv \{x^{tk} \mid Q^{tk}x^{tk} = q^{tk} + T^{tk}x^{a(t,k)}\}$ are affine spaces, $x^{a(1,1)} := 0$, and $\bar{\eta}^{(T+1,k)} := 0$. Without loss of generality, we assume $Q^{tk}, \forall t = 1,\ldots,T, k = 1,\ldots,K(t)$, have full row rank. The definition clearly shows that the kth t-stage subproblem, $\bar{\eta}^{tk}(x^{a(t,k)})$, is well defined only if $x^{a(t,k)}$ is given, and that its objective function contains the next stage subproblems $\bar{\eta}^{(t+1,s)}(x^{tk})$, $\forall s \in \mathcal{D}_t^k$.

We define the multistage barrier problem as follows.

$$\begin{aligned} f^{tk}(x^{tk}) &= c^{tk^T} x^{tk} + \mu B^{tk}(x^{tk}) + \sum_{s \in \mathcal{D}_t^k} \eta^{(t+1,s)}(x^{tk}), \\ \eta^{tk}(x^{a(t,k)}) &= \mathit{min}\, f^{tk}(x^{tk}) \, s.t.\, x^{tk} \in L^{tk}(x^{a(t,k)}), \\ &\forall t = 1,\ldots,T, k = 1,\ldots,K(t), \\ &x^{a(1,1)} := 0, \bar{\eta}^{(T+1,k)} := 0, \end{aligned} \qquad (48)$$

where each $B^{tk}(\cdot)$, $t = 1, \ldots, T, k = 1, \ldots, K(t)$ is a self-concordant barrier function of corresponding domain $intG^{tk}$ with complexity value ϑ^{tk}. We also define that

$$\tilde{\vartheta}^{tk} := \vartheta^{tk} + \sum_{s \in \mathcal{D}_t^k} \tilde{\vartheta}^{(t+1,s)}, \tag{49}$$

i.e., $\tilde{\vartheta}^{tk}$ is the sum of complexity values of the sub-tree rooted at scenario (t, k).

At least three alternative approaches are possible for solving a multistage stochastic convex program. The first possibility is to formulate its deterministic equivalent. Subsequently we can exploit the structure of the problem to perform linear algebra in parallel Hegland *et al.* [6]. The second approach is to formulate the problem using nonanticipativity constraints, and subsequently relax these constraints by Lagrangian dual, the progressive hedging method Rockafellar and Wets [14], and methods proposed in Ruszczynskii [16] and Mulvey and Ruszczynskii [10]. More recently, Zhao [19] has given an interior decomposition method based on the Lagrangian dual approach. The third approach is to use the Bender decomposition formulation (47). The L-shaped method of Van Slyke and Wets [17] was extended in Birge [12] for the multistage problems using this formulation. The prototype algorithm of this section also uses the Bender decomposition formulation. However, instead of using cutting planes, as is the case with L-shaped methods, it proposes an interior barrier decomposition scheme that is an extension of the method for the two-stage problem discussed in the previous section. Our proposal is based on the property that the multistage barrier problems forms a self-concordant family tree.

5.2 Self-Concordant Tree Property in Multistage Problem

For a two-stage problem, Theorems 2.2 shows that the barrier recourse functions $\{\eta^k(x, \mu), \mu > \}$ and $\{r(x) := c^T x + \mu b(x) + \sum \eta^k(x, \mu), \mu > 0\}$ are self-concordant families. The following two theorems establish recursively that all recourse functions and objective functions in a multistage barrier problem (48) are also self-concordant families, and show that the complexity values accumulate additively.

Theorem 5.1 *For a multistage barrier problem in (48), fix a pair* (t, k) *and assume that* $\{\eta^{(t+1,s)}(x^{tk}, \mu), \mu > 0\}, \forall s \in \mathcal{D}_t^k$ *are self-concordant families with parameter functions* $\alpha^{(t+1,s)}(\mu) = \mu$, $\gamma^{(t+1,s)}(\mu) = \nu^{(t+1,s)}(\mu) = 1$, $\xi^{(t+1,s)}(\mu) = \sqrt{\tilde{\vartheta}^{(t+1,s)}}/\mu$, $\sigma^{(t+1,s)}(\mu) = 1/2\mu$, *then* $\{f^{tk}(x^{tk}, \mu), \mu > 0\}$ *is a self-concordant family with following parameter functions:*

$$\alpha^{tk}(\mu) = \mu, \gamma^{tk}(\mu) = \nu^{tk}(\mu) = 1, \xi^{tk}(\mu) = \frac{\sqrt{\tilde{\vartheta}^{tk}}}{\mu}, \sigma^{tk}(\mu) = \frac{1}{2\mu}.$$

Proof: The proof of convexity and differentiability of $f^{tk}(x^{tk})$ is straight forward. It is also clear that $f^{tk}(x^{tk}, \mu)$ is a self-concordant function following a proof similar to that of Lemma 3.1. We need to prove condition (SCF3) in Definition 2.2, i.e., that $f^{tk}(x^{tk}, \mu)$ are compatible with neighbors. Let $g^{tk} = \nabla B^{tk}(x^{tk}), H^{tk} = \nabla^2 B^{tk}(x^{tk})$. To simplify notation we let

$$g^s := \nabla \eta^{(t+1,s)}(x^{tk}, \mu), H^s := \nabla^2 \eta^{(t+1,s)}(x^{tk}, \mu), R^s := Q^{(t+1,s)} {H^s}^{-\frac{1}{2}}, s \in \mathcal{D}_t^k.$$

Hence,

$$\{\nabla f^{tk}(x^{tk})\}'_\mu = g^{tk} + \sum_{s \in \mathcal{D}_t^k} \{g^s\}', \text{ and } \nabla^2 f^{tk}(x^{tk}) = \mu g^{tk} + \sum_{s \in \mathcal{D}_t^k} H^s.$$

We note that $g^s, H^s, \{g^s\}'_\mu$ are computed recursively by accumulating from stage $t+1$ up to the final stage T. We now show that $\{f^{tk}(x^{tk},\mu), \mu > 0\}$ is a self-concordant family. Let $\hat{I} = [I, \dots, I]$, $\bar{h}^T = h^T \hat{I}$, $\bar{g}^T(x,\mu) = \left[g^{tk^T}, \dots, g^{s^T}, s \in \mathcal{D}_t^k\right]$, $D^{tk} = \sqrt{\mu} H^{tk^{\frac{1}{2}}}$, $\tilde{D}^{tk} = \frac{1}{\sqrt{\mu}} H^{tk^{-\frac{1}{2}}}$, $D^s = \sqrt{\mu} T^{s^T}\left(R^s R^{s^T}\right)^{-\frac{1}{2}}$, $\tilde{D}^s(x,\mu) = \frac{1}{\sqrt{\mu}}(R^s R^{sT})^{-\frac{1}{2}} R^s H^{s^{-\frac{1}{2}}}$, $s \in \mathcal{D}_t^k$, and

$$D = blkdiag(D^{tk}, \dots, D^s, s \in \mathcal{D}_t^k), \tilde{D}(x,\mu) = blkdiag(\tilde{D}^{tk}, \dots, \tilde{D}^s, s \in \mathcal{D}_t^k). \tag{50}$$

Here *blkdiag* represents a block diagonal diagonal matrix.

Hence,

$$\begin{aligned}
\left| h^T \left\{\nabla f^{tk}(x^{tk})\right\}'_\mu \right| &= \left|\bar{h}^T D \tilde{D} \bar{g}\right| \le \sqrt{\bar{h}^T D D^T \bar{h}} \sqrt{\bar{g}^T \tilde{D}^T \tilde{D} \bar{g}} \\
&= \sqrt{h^T \left(\mu H^{tk} + \sum_{s \in \mathcal{D}_t^k} H^s\right) h} \cdot \sqrt{\frac{1}{\mu} g^{tk^T} H^{tk^{-1}} g^{tk} + \sum_{s \in \mathcal{D}_t^k} g^{s'^T} H^{s^{-1}} g^{s'}} \\
&\le \sqrt{h^T \nabla^2 f^{tk}(x) h} \sqrt{\frac{1}{\mu}\left(\vartheta^{tk} + \sum_{s \in \mathcal{D}_t^k} \tilde{\vartheta}^s\right)} \qquad (51) \\
&= \sqrt{\frac{\tilde{\vartheta}^{tk}}{\mu}} \sqrt{h^T \nabla^2 f^{tk}(x) h}. \qquad (52)
\end{aligned}$$

We draw reader's attention to the differences between (51) and (33). The inequality (51) is due to the assumption that all barrier recourse functions $\{\eta^{(t+1,s)}(x^{tk},\mu), \mu > 0\}$ are self-concordant families with $\nu^{tk}(\mu) = 1$, hence (51) follows from (SCF3), which shows that $g^{s'^T} H^{s^{-1}} g^{s'}$ is bounded from above by $\xi(\mu)^2 \alpha(\mu) = \frac{1}{\mu}\tilde{\vartheta}^s$. While (33) holds because the kth second stage barrier function has complexity value, which means $\frac{1}{\mu} g^{k^T} H^{k^{-1}} g^k$ is bounded from above by $\frac{1}{\mu}\vartheta^k$. We emphasize that $g^{sT} H^{s-1} g^s$ is not necessarily bounded from above, and hence we don't require $\eta^{(t+1,s)}(x^{tk})$ to be a barrier.

The proof of the second inequality of (7) for $f^{tk}(x^{tk})$ follows steps similar to that for (35). $\square$

Theorem 5.2 *For a multistage barrier problem in (48), fix a pair (t,k) and assume that $\{f^{tk}(x^{tk},\mu), \mu > 0\}$ is a self-concordant family, then $\{\eta^{tk}(x^{a(t,k)},\mu), \mu > 0\}$ is also a self-concordant family with exactly the same parameter functions*

$$\alpha^{tk}(\mu) = \mu, \gamma^{tk}(\mu) = \nu^{tk}(\mu) = 1, \xi^{tk}(\mu) = \frac{\sqrt{\tilde{\vartheta}^{tk}}}{\mu}, \sigma^{tk}(\mu) = \frac{1}{2\mu}.$$

Proof: By following steps similar to those in the proof for the two-stage problem we can show that $\eta^{tk}(x^{a(t,k)},\mu)$ is convex, twice continuously differentiable, and self-concordant. We only need to show that $\eta^{tk}(\mu, x^{a(t,k)})$ is compatible with neighbors and compute its parameter functions. Using the approach of Lemma 2.2 and Lemma 2.3 we get

$$\nabla^2\eta^{tk}(\mu, x^{a(t,k)}) = \mu\bar{T}^T(\bar{R}\bar{R}^T)^{-1}\bar{T} \tag{53}$$

$$\{\nabla\eta^{tk}(\mu, x^{a(t,k)})\}' = \bar{T}^T(\bar{R}\bar{R}^T)^{-1}\bar{R}\bar{H}^{-\frac{1}{2}}\{\bar{g}\}', \tag{54}$$

$$\text{where } \bar{T} = T^{tk}, \bar{Q} = Q^{tk}, \bar{g} = \nabla f^{tk}(\mu, x^{tk^*}), \bar{H} = \nabla^2 f^{tk}(\mu, x^{tk^*}), \bar{R} = Q^{tk}\bar{H}^{-\frac{1}{2}}. \tag{55}$$

Hence, we have

$$\begin{aligned} \left| h^T \left\{ \nabla\eta^{tk}(x^{a(t,k)}, \mu) \right\}_\mu' \right| &= \left| h^T\bar{T}^T(\bar{R}\bar{R}^T)^{-1}\bar{R}\bar{H}^{-\frac{1}{2}}\{\bar{g}\}' \right| \\ &\le \sqrt{h^T\bar{T}^T(\bar{R}\bar{R}^T)^{-1}\bar{T}h}\sqrt{(\{\bar{g}\}')^T\,\bar{H}^{-\frac{1}{2}}\bar{R}^T(\bar{R}\bar{R}^T)^{-1}\bar{R}\bar{H}^{-\frac{1}{2}}\{\bar{g}\}'} \\ &\le \sqrt{\frac{1}{\mu}h^T\nabla^2\eta^{tk}(x^{a(t,k)}, \mu)h}\sqrt{(\{\bar{g}\}')^T\,\bar{H}^{-1}\{\bar{g}\}'} \\ &\le \sqrt{\frac{\tilde{\vartheta}^{tk}}{\mu}}\sqrt{h^T\nabla^2\eta^{tk}(x^{a(t,k)}, \mu)h}. \end{aligned} \tag{56}$$

The last inequality of (56) holds because $\{f^{tk}(x^{tk}, \mu), \mu > 0\}$ is a self-concordant family by assumption. Now, we prove the second inequality (7) in condition (SCF3) of Definition 2.2.

$$\begin{aligned} & \left| \left\{ h^T\nabla^2\eta^{tk}(x^{a(t,k)}, \mu)h \right\}_\mu' \right| \\ = & \left| h^T \left\{ \mu\bar{T}^T\left(\bar{R}\bar{R}^T\right)^{-1}\bar{T} \right\}_\mu' h \right| \\ = & \left| h^T\bar{T}^T\left(\bar{R}\bar{R}^T\right)^{-1}\bar{Q}\bar{H}^{-1}\left\{\bar{H}\right\}_\mu'\bar{H}^{-1}\bar{Q}^T\left(\bar{R}\bar{R}\right)^{-1}\bar{T}h \right| \\ = & \left| \bar{h}^T\left\{\nabla^2 f^{tk}(x^{tk^*}, \mu)\right\}_\mu'\bar{h} \right| \le \frac{1}{\mu}\bar{h}^T\nabla^2 f^{tk}(x^{tk^*}, \mu)\bar{h} = \frac{1}{\mu}h^T\nabla^2\eta^{tk}(x^{a(t,k)}, \mu)h, \end{aligned} \tag{57}$$

where $\bar{h} = \bar{H}^{-1}\bar{Q}^T\left(\bar{R}\bar{R}\right)^{-1}\bar{T}h$. The inequality holds because $\{f^{tk}(x^{tk}, \mu), \mu > 0\}$ is self-concordant family by assumption. The last equality follows from a straightforward computation. □

Finally by induction we show the structure of the barrier Bender decomposition formulation: all members are self-concordant families with respect to a single control parameter μ and the complexity values accumulate across the tree additively.

Theorem 5.3 *For any subproblem (t, k) in a multistage problem, both $\eta^{tk}(x^{a(t,k)}, \mu)$ and its objective function $\{f^{tk}(x^{tk}, \mu), \mu > 0\}$ are self-concordant families with exactly the same parameter functions*

$$\alpha^{tk}(\mu) = \mu, \gamma^{tk}(\mu) = \nu(\mu) = 1, \xi^{tk}(\mu) = \frac{\sqrt{\tilde{\vartheta}^{tk}}}{\mu}, \sigma^{tk}(\mu) = \frac{1}{2\mu}.$$

Especially the first stage objective function is a self-concordant family with $\xi^{(1,1)}(\mu) = \frac{\sqrt{\tilde{\vartheta}^{(1,1)}}}{\mu}$, which is the total complexity values of the scenario tree.

Proof: For any fixed $x^{(T-2,\cdot)}$, $\eta^{(T-1,\cdot)}\left(x^{(T-2,\cdot)},\mu\right)$ is a two-stage problem, hence $\{f^{(T-1,\cdot)}(x^{(T-1,\cdot)},\mu)\}$ is a self-concordant family as shown in our analysis of a two-stage problem. Assume that $\{f^{tk}(x^{tk},\mu),\mu>0\}$ is a self-concordant family for any $2\leq t\leq T-2$, then Theorem 5.2 shows that $\{\eta^{tk}(x^{a(t,k)},\mu),\mu>0\}$ is a self-concordant family. Given this conclusion and by using Theorem 5.1 we further conclude that $\{f^{(t-1,\cdot)}(x^{(t-1)(\cdot)},\mu)\}$ is a self-concordant family. Hence, $\{f^{tk}(x^{tk},\mu),\mu>0\}$, and $\{\eta^{tk}(x^{a(t,k)},\mu),\mu>0\}$, $\forall t=1,\ldots,T,k=1,\ldots,K(t)$ are self-concordant families. □

5.3 A Prototype Barrier Decomposition Algorithm for multistage Stochastic Programs

From Theorem 5.3 the complexity of the first stage composite barrier recourse function is $\tilde{\vartheta}^{(1,1)}$. This suggests that if the Hessian and gradient of the recourse function can be computed exactly (or with sufficient accuracy) then we can apply a short or long step algorithm to the first stage problem. Computing the gradient and Hessian of the recourse function requires recursive solutions of second and subsequent stage centering problems, which also decompose. This results in a prototype decomposition algorithm for multistage stochastic programs. This algorithm is outlined as Algorithm 2. It is straightforward to follow the proofs of the two-stage problem, and draw conclusions about the number of first stage Newton steps of a short step or a long step algorithm. We state the result without proof here.

Theorem 5.4 *Let μ_0 be the initial barrier parameter, and ϵ be its target value. If $\delta(\mu^0,x^0)\leq\kappa=(2-\sqrt{3})/2$, and if the short step algorithm reduces μ at a constant rate $\gamma=1-\sigma/\sqrt{\tilde{\vartheta}^{(1,1)}}$, where $\sigma\leq 0.1$ then the short step algorithm terminates with a solution x satisfying $\delta(\epsilon,x)\leq\kappa$ in $O\left(\sqrt{\tilde{\vartheta}^{(1,1)}}ln\frac{\mu^0}{\epsilon}\right)$ number of first stage Newton iterations.*

Theorem 5.5 *Let μ_0 be the initial barrier parameter, and ϵ be its target value. If $\delta(\mu^0,x^0)\leq\kappa=(2-\sqrt{3})/2$, and if long step algorithm reduces μ at rate $\gamma<1$, then this algorithm terminates with a solution x satisfying $\delta(\epsilon,x)\leq\kappa$ in $O\left(\tilde{\vartheta}^{(1,1)}ln\frac{\mu^0}{\epsilon}\right)$ number of first stage Newton iterations.*

6 Conclusions And Future Work

In this paper we have shown that we can regularize two and mulit-stage stochastic programming problems by using self-concordant barriers on the feasible regions. We showed that the barrier decomposition problems resulting from this regularization form a self-concordant family tree whose complexity value accumulates additively. These properties allows us to apply classical path following interior point algorithms and allow us to bound the number of first stage Newton iteration required to solve these problems. The algorithms for two and multistage problems proposed in this paper are prototype algorithms since they require exact solutions for second and subsequent stage centering problems. In practice it is only possible to find an approximate solution of these problems. The computational results for the two-stage conic programs in Mehrotra and Ozevin [7] suggest that in the two-stage setting it suffices to find inaccurate

Algorithm 2 Prototype long step algorithm for multistage convex problem

solve LSMNC$(1, 1, \epsilon)$;

LSMNC$(t, k, \epsilon)\{$

find μ_0, x_0, γ such that $\delta(\mu_0, x_0) \leq \kappa, \gamma = 1 - \sigma/\sqrt{\tilde{\vartheta}^{tk}}, \sigma \leq 0.1$

set $x = x_0, \mu = \mu_0$

while $(\mu > \epsilon)$

update $\mu = \gamma\mu$

while $(\delta(x, \mu) > \kappa)$

if $t = T - 1$, solve $\eta^{Ts}, \forall s \in \mathcal{D}_t^k$; otherwise, solve $LSMNC(t + 1, s, \mu), \forall s \in \mathcal{D}_t^k$

compute Newton direction Δx and update $x = x + \frac{1}{1+\delta(x,\mu)}\Delta x$ if $\delta(x, \mu) > 2\kappa$; otherwise $x = x + \Delta x$.

$\}$

solutions of the second stage problems. This is not surprising because, although the analysis of interior point algorithm presented here requires exact Newton direction computations, it is well understood that in practical application of this method the Newton direction computation need not be exact. It remains to be seen if this empirically observed property continues to hold for the multistage case. It is a topic of future research to see how the analysis of this paper should be extended to the situation where exact solutions of second and subsequent stage centering problems are not available. It was also observed in [7] that warm-start and adaptive addition of scenarios was possible for the two-stage conic programming problems. It is a topic of future research to find if such properties generalize to the multistage case.

References

[1] K.A. Ariyawansa and Y. Zhu. A class of polynomial volumetric barrier decomposition algorithms for stochastic semidefinite programming. Technical report, Washington State University, 2006.

[2] John R. Birge and Francois Louveaux. *Introduction To Stochastic Programming*. Springer, New York, 1997.

[3] Leonid Faybusovich and Michael Gekhtman. Calculation of universal barrier functions for cones generated by chebyshev systems over finite sets. *SIAM J. on Optimization*, 14(4):965–979, 2004.

[4] Leonid Faybusovich, Thanasak Moutonglang, and Takashi Tsuchiya. Numerical experiments with universal barrier functions for cones of chebyshev systems. *Computational Optimization And Applications (To Appear)*, 2007.

[5] Anthony V. Fiacco. *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming.* Academic Press, New York, 1983.

[6] M. Hegland, M.R. Osborne, and J. Sun. Parallel interior point schemes for solving multistage convex programming. *Annals of Operations Research*, 108(1-4), 2001.

[7] Sanjay Mehrotra and M. Gokhan Ozevin. On the implementation of interior point decomposition algorithms for two-stage stochastic conic programs. Technical report, IEMS Department, Northwestern University, 2005.

[8] Sanjay Mehrotra and M. Gokhan Ozevin. Decomposition-based interior point methods for two-stage stochastic convex quadratic programs with recourse. *Operations Research*, 2007.

[9] Sanjay Mehrotra and M. Gokhan Ozevin. Decomposition-based interior point methods for two-stage stochastic semidefinite programming. *SIAM Journal on Optimization*, 18(1):206–222, 2007.

[10] John M. Mulvey and Andrzej Ruszczyński. A new scenario decomposition method for large-scale stochastic optimization. *Operations Research*, 43(3):477–490, May 1995.

[11] Yurii Nesterov and Arkadii Nemirovskii. *Interior Point Polynomial Algorithms In Convex Programming.* SIAM, Philadelphia, 1994.

[12] John R.Birge. An L-shaped method computer code for multistage stochastic linear programs. *Numerical Methods in Stochastic Programming, R. Wets and Y. Ermoliev, eds*, pages 255–266, 1988.

[13] James Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization.* SIAM, 2001.

[14] R. T. Rockafellar and R.J.B Wets. Scenarios and policy aggregation inoptimization under uncertainty. *Mathematics of Operations Research*, 16(1):1–29, 1991.

[15] Walter Rudin. *Principals of Mathematical Analysis.* McGraw-Hill, 1976.

[16] Andrzej Ruszczyński. On convergence of an augmented lagrangian decomposition method for sparse convex optimization. *Mathematics of Operations Research*, 20(3):634–656, 1995.

[17] R. Van Slyke and R. Wets. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal of Applied Mathematics*, 17 (4):638–663, 1967.

[18] Gongyun Zhao. A log-barrier method with Benders decomposition for solving two-stage stochastic linear programs. *Mathematical Programming*, 90(3):507, 2001.

[19] Gongyun Zhao. A lagrangian dual method with self-concordant barriers for multi-stage stochastic convex programming. *Mathematical Programming*, 102(1):1–24, January 2005.