Convergence Analysis of a Weighted Barrier Decomposition Algorithm for Two Stage Stochastic Programming *

Sanjay Mehrotra M. Gokhan Ozevin

Technical Report, 2007-07

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

Abstract

Mehrotra and Ozevin [7] computationally found that a weighted primal barrier decomposition algorithm significantly outperforms the barrier decomposition proposed and analyzed in [11; 6; 8]. This paper provides a theoretical foundation for the weighted barrier decomposition algorithm (WBDA) in [7]. Although the worst case analysis of the WBDA achieves a first-stage iteration complexity bound that is worse than the bound shown for the decomposition algorithms of [11] and [6; 8], under a probabilistic assumption we show that the worst case iteration complexity of WBDA is independent of the number of scenarios in the problem. The probabilistic assumption uses a novel concept of self-concordant random variables.

Key Words: Two stage Stochastic Programming, linear-quadratic programming, Bender's decomposition, large scale optimization, nondifferentiable convex optimization

*This research was partially supported by NSF grants DMI-0200151, DMI-0522765, and ONR grant N0014-01-1-0048/P00002

1. Introduction

Zhao [11] developed a log-barrier decomposition algorithm (BDA) for two-stage stochastic linear programs (TSSLP). Mehrotra and Ozevin [6; 8] extended Zhao's analysis to more general two-stage stochastic quadratic and semi-definite programs. The essential feature of these algorithms is that they perform primal Newton iterations on a barrier function for the first problems. The gradient and Hessian required to compute the Newton direction of the primal barrier algorithm are computed by solving second-stage centering problems which decompose in the number of scenarios.

For two stage stochastic programs with discrete support, starting from a suitably centered first-stage solution (sufficiently accurate solution for the barrier problem with parameter $\mu = \mu_0$) [11; 8] showed that for a short-step primal interior algorithm the number of first-stage Newton iterations required to obtain a suitably centered solution (for $\mu = \mu_k$) of a two-stage stochastic linear or semi-definite program is $O(\sqrt{n + K\hat{n}} \ln(\mu_k/\mu_0))$. Here n is the dimension of the first stage linear and/or semi-definite cone, $\hat{n}$ is the dimension of the second stage linear and/or semi-definite cone, and K is the number of scenarios in the stochastic program. The value μ_k is chosen to ensure that the computed solution has ϵ-accuracy in the objective function for any $\epsilon > 0$. The long-step analysis in [11; 8] proves a worst case bound of $O((n + K\hat{n}) \ln(\mu_k/\mu_0))$ primal Newton iterations. The long-step primal interior algorithms are known to take fewer iterations in practice even though they have an inferior worst case bound.

In their computational experiments Mehrotra and Ozevin [7] found that a log-barrier decomposition algorithm which weighs the second-stage barrier parameters with corresponding probabilities achieves significantly superior performance when compared to the standard decomposition approach analyzed in [11; 6; 8]. The purpose of this paper is to provide a theoretical foundation for the algorithm proposed in [7]. In particular, we show that the complexity of the long-step WBDA is $O(K(n + \hat{n}) \ln(\mu_k/\mu_0))$ first-stage Newton iterations. Under a probabilistic assumption we show that the complexity of the long-step WBDA is $O(\tilde{K}(n+\hat{n}) \ln(\mu_k/\mu_0))$ first-stage Newton iterations for problems with discrete or continuous support, where $\tilde{K}$ is a self-concordance parameter of a random matrix appearing in the computations. This bound is independent of the number of second-stage scenarios when $\tilde{K}$ does not depend on K. As in [11; 6; 8] these bounds are shown under the assumption that the second-stage centering problems are solved exactly. The analysis of BDA in [11; 6; 8] and the one presented here is based on showing that the log-barrier functions associated with two-stage stochastic programs are strongly self-concordant and form a self-concordant family (see Appendix A). For simplicity we perform this analysis for TSSLP.

This paper is organized as follows. Section 2 describes the weighted barrier recourse formulation of TSSLP, establishes its basic properties, and compares it with the standard barrier recourse formulation. Additional notation, and assumptions are also introduced in this section. Section 3 gives expressions for the gradient and Hessian of the barrier recourse with respect to the first stage variables, and with respect to the barrier parameter. Some basic bounds on the derivative of the barrier recourse function and second stage solutions are also proved in this Section. Section 4 shows that the second stage solutions are real analytic functions of the barrier parameter and the first stage solutions. Section 5 shows that the expected barrier recourse function is also a real analytic function of the first stage solutions. Section 6 establishes bounds on the value of the self-concordance parameters under a probabilistic assumption. Section 7 establishes worst case bounds on the value of self-concordance parameters for problems with finite support. Section 8 provides a convergence analysis of WBDA. Appendix A gives the definition of a self-concordance function and the self-concordant barrier. Appendix B gives the definition of a real analytic function and a sufficient condition for a function to be a real analytic function. Appendix B also gives sufficient conditions for commuting the integration with derivatives. Appendix C proves a generalized Holder, and a projection inequality needed in the analysis of Section 6.

2. Barrier Recourse Formulations of TSSLP

2.1 Assumptions and Notations for TSSLP

The two-stage stochastic linear program (TSSLP) with recourse is:

$$\min_{x\in\mathcal{P}} \tilde{\eta}(x), \quad \tilde{\eta}(x) := c^T x + E[\tilde{\rho}(x)], \quad \mathcal{P} := \{x \mid Ax = b, x \in \mathbb{R}^n_+\}, \text{ where} \tag{2.1}$$

$$E[\tilde{\rho}(x)] := \textstyle\int_{\Xi} \tilde{\rho}^\xi(x) dF(\xi), \tag{2.2}$$

$$\tilde{\rho}^\xi(x) := min_{y^\xi \in \mathcal{P}^\xi(x)} \, {p^\xi}^T y^\xi, \quad \mathcal{P}^\xi(x) := \{y^\xi \mid W^\xi y^\xi = h^\xi - T^\xi x, y^\xi \in \mathbb{R}^{n^\xi}_+\}, \xi \in \Xi. \tag{2.3}$$

Here Ξ is the support of random parameters $\tilde{\xi}$ with probability distribution $F(\xi)$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. For a realization ξ of $\tilde{\xi}$, $T^\xi \in \mathbb{R}^{m^\xi \times n^\xi}$, and $W^\xi \in \mathbb{R}^{m^\xi \times n^\xi}$.

We define feasibility sets:

$\mathcal{P}^1 := \{x \mid x \in \mathcal{P}, E[\tilde{\rho}(x)] < \infty\}, \mathcal{P}_o := int(\mathcal{P}^1)$,
for $x \in \mathcal{P}_o$, and any realization $\xi \in \Xi, \mathcal{P}^\xi_o(x) := \{y^\xi \mid W^\xi y^\xi = h^\xi - T^\xi x, y^\xi \geq \frac{1}{\iota} e > 0\}$,
and $\mathcal{D}^\xi_o(x) := \{(z^\xi, s^\xi) \mid {W^\xi}^T z^\xi + s^\xi = p^\xi, s^\xi \geq \frac{1}{\iota} e > 0\}$.

We make the following assumptions:

A1 The matrices A and W^ξ, $\forall \xi \in \Xi$ have a full row rank.

A2 The set $\mathcal{P}_o$ is non-empty and bounded set.

A3 For some $\iota > 0$, sets $\mathcal{P}^\xi_o(x)$ and $\mathcal{D}^\xi_o(x)$ have a feasible solution and they are bounded for all $x \in \mathcal{P}_o$. In particular, let $\|y^\xi\| \leq \tilde{\iota} < \infty$.

Assumption A1 is for convenience. We can satisfy this assumption by deleting linearly dependent rows in A and W^ξ. Assumptions A2, A3 require that the first and second-stage problems (2.1–2.3) and their dual have a strictly feasible solution. These assumptions also imply that problems (2.29–2.32) below have a unique optimum solution. The assumption that the feasible region has a non-empty interior and it is bounded is standard in interior point methods. Note that Assumption A3 also rules out situations where the second stage problem may be infeasible for a scenario. We may ensure Assumptions A2 and A3 by introducing artificial variables and bounds on variables when $\tilde{\xi}$ has finite support.

In the case where $\tilde{\xi}$ has a finite support $\Xi = \{\xi^1, \ldots, \xi^K\}$ with probabilities $\{\pi^1, \ldots, \pi^K\}$, the entities T^ξ, W^ξ, p^ξ, h^ξ, y^ξ, s^ξ, z^ξ, $\tilde{\rho}^\xi(\cdot)$, $\rho^\xi(\cdot)$, $\mathcal{P}^\xi(x)$, $\mathcal{P}^\xi_o(x)$, and $\mathcal{D}^\xi_o(x)$ are denoted by T^i, W^i, p^i, h^i, y^i, s^i, z^i, $\tilde{\rho}^i(\cdot)$, $\rho^i(\cdot)$, $\mathcal{P}^i(x)$, $\mathcal{P}^i_o(x)$, and $\mathcal{D}^i_o(x)$, $i = 1, \ldots, K$. In this case (2.1–2.3) is:

$$min_{x\in\mathcal{P}} c^T x + E[\bar{\rho}(x)], \text{ where } E[\bar{\rho}(x)] := \textstyle\sum_{i=1}^K \pi^i \rho^i(x), \tag{2.4}$$

$$\bar{\rho}^i(x) := \min_{y^i\in\mathcal{P}^i(x)} {p^i}^T y^i, \text{ for } i = 1, \ldots, K. \tag{2.5}$$

If all scenarios are equally likely, then $\pi^i = 1/K$, e.g., when scenarios are generated from the sample average approximation method [3; 10]. We will use this value of π^i when stating our iteration complexity results in the analysis. In addition, for simplicity, $\hat{n} = n^1 = \ldots = n^K = n^\xi$, and $\hat{m} = m^1 = \ldots = m^K = m^\xi$.

2.2 The Weighted Log-Barrier Problem

The dual of (2.3) is given by

$$max_{(z^\xi,s^\xi)\in\mathcal{D}^\xi(x)} (h^\xi - T^\xi x)^T z^\xi, \;\; \mathcal{D}^\xi(x) := \{(z^\xi, s^\xi) \mid {W^\xi}^T z^\xi + s^\xi = p^\xi, s^\xi \geq 0\}. \tag{2.6}$$

The weighted barrier problems for (2.1–2.3, 2.6) are defined as:

$$\min_{x\in\mathcal{P}_o} \eta(\mu, x), \;\; \eta(\mu, x) := c^T x - \mu \textstyle\sum_{i=1}^n \ln x_i + E[\rho(\mu, x)], \text{ where} \tag{2.7}$$

$$E[\rho(\mu, x)] := \textstyle\int_\Xi \rho^\xi(x, \mu) dF(\xi), \tag{2.8}$$

$$\rho^\xi(\mu, x) := min_{y^\xi\in\mathcal{P}^\xi(x)} {p^\xi}^T y^\xi - \mu \textstyle\sum_{j=1}^{n^\xi} \ln y^\xi_j, \text{ for } \xi \in \Xi, \tag{2.9}$$

$$\max_{(z^\xi,s^\xi)\in\mathcal{D}^\xi(x)} -(h^\xi - T^\xi x)^T z^\xi + \mu \textstyle\sum_{j=1}^{n^\xi} \ln s^\xi_j. \tag{2.10}$$

We denote the optimum solution of the first-stage problem (2.7) by $x(\mu)$ and the optimum solution of the second-stage problems (2.9–2.10) by $(y^\xi(\mu, x), s^\xi(\mu, x), z^\xi(\mu, x))$. For a given μ the first order KKT-conditions for the first-stage problem (2.7) are:

$$\nabla\eta(\mu, x) - A^T\lambda = 0, Ax = b. \tag{2.11}$$

Since problems (2.9) and (2.10) are respectively convex and concave, y^ξ and (z^ξ, s^ξ) are optimum solutions of (2.9) and (2.10), respectively, if and only if they satisfy the KKT-conditions:

$$\begin{aligned} &Y^\xi s^\xi = \mu e, \\ &W^\xi y^\xi = h^\xi - T^\xi x, \\ &{W^\xi}^T z^\xi + s^\xi = p^\xi, \\ &y^\xi > 0, \;\; s^\xi > 0. \end{aligned} \tag{2.12}$$

The following theorem shows that as $\mu \to 0$, the objective value of TSSLP evaluated at the optimum solution of the first stage barrier problem (2.7) converges to the optimum objective value of TSSLP. Hence, we can solve a sequence of (2.7) with decreasing value of μ to solve TSSLP.

Theorem 2.1 *Consider TSSLP (2.1–2.3) and the associated weighted barrier problem (2.7–2.10), and let $\mathcal{X}^*$ be the set of optimum solutions of (2.1–2.3). Suppose TSSLP (2.1–2.3) satisfies Assumptions A1–A3. Then,* *(i)* *The second-stage primal-dual central path* $\{y^\xi(\mu,x), z^\xi(\mu,x), s^\xi(\mu,x), \mu > 0\}$ *is well defined for any* $x \in \mathcal{P}_o$.

(ii) $\lim_{\mu\to 0} E[\rho(\mu,x)] = E[\tilde{\rho}(x)]$.

(iii) *The problem (2.7) has a unique minimizer* $x(\mu)$ *for all* $\mu > 0$.

(iv) $\lim_{\mu\to 0} \tilde{\eta}(\rho(x(\mu))) = \tilde{\eta}(x^*)$ *for some* $x^* \in \mathcal{X}^*$.

We need the following lemma in the proof of Theorem 2.1. It provides bounds on the magnitude of second stage solutions and related quantities generated during the course of the barrier decomposition algorithm.

Lemma 2.1 *Let* $(y^\xi(\mu,x), z^\xi(\mu,x), s^\xi(\mu,x))$ *be the solution of (2.12). Let* $Q^\xi(\mu,x) := Y^\xi(\mu,x)^{1/2} S^\xi(\mu,x)^{-1/2} = \frac{1}{\sqrt{\mu}} Y^\xi(\mu,x)$ *and* $R^\xi(\mu,x) := W^\xi Q^\xi(\mu,x)$. *Then, we have*

$$\left\|Q^\xi(\mu,x)\right\|_\infty, \left\|Q^\xi(\mu,x)^{-1}\right\|_\infty \le \sqrt{\hat{\iota}/\mu}, \text{ and } \frac{\hat{\iota}}{\mu} I \succeq (R^\xi(\mu,x) R^\xi(\mu,x)^T)^{-1},$$

where $\hat{\iota} > 0$ *is a constant. Furthermore,*

$$\left(T_j^{\xi T} (R^\xi(\mu,x) R^\xi(\mu,x)^T)^{-1} T_j^\xi\right)^{1/2} \le \frac{\sqrt{\hat{\iota}}\|T_j^\xi\|}{\sqrt{\mu}} \le \frac{\hat{\iota}}{\sqrt{\mu}},$$

where $T_j^\xi = T^\xi e_j$, *and we have assumed that* $\|T_j^\xi\| \le \sqrt{\hat{\iota}}$, $j = 1, \ldots, n$.

Proof: We let $T = T^\xi$ and $T_j = T_j^\xi$. Since, $Q^\xi(\mu,x) = \left(Y^\xi(\mu,x) S^\xi(\mu,x)\right)^{-1/2} Y^\xi(\mu,x)$, we have $\|Q^\xi(\mu,x)\|_\infty \le \frac{1}{\mu}\|y^\xi(\mu,x)\|_\infty \le \frac{\tilde{\iota}}{\mu}$. Also, from (2.12) we have $p - \mu Y^\xi(\mu,x)^{-1} e - W^T z^\xi(\mu,x) = 0$. Let $\hat{y} \in \mathcal{P}_o^\xi(x)$. Taking the inner product with $(\hat{y} - y^\xi(\mu,x))$ gives $(\hat{y} - y^\xi(\mu,x))^T Y^\xi(\mu,x)^{-1} e = \frac{(\hat{y}-y^\xi(\mu,x))^T p}{\mu}$. Hence,

$$\sum_{i=1}^{\hat{n}} \frac{\hat{y}_i}{y_i^\xi(\mu,x)} = \frac{(\hat{y} - y^\xi(\mu,x))^T p}{\mu} + \hat{n} \le \frac{\|p\|_1 \|\hat{y} - y^\xi(\mu,x)\|_\infty}{\mu} + \hat{n} \le \frac{2\tilde{\iota}\|p\|_1}{\mu} + \hat{n}.$$

Consequently,

$$\frac{1}{\mu}\|s^\xi(\mu,x)\|_1 = \|Y^\xi(\mu,x)^{-1} e\|_1 \le \iota\left(\frac{2\tilde{\iota}\|p\|_1}{\mu} + \hat{n}\right). \tag{2.13}$$

This gives

$$\left\|Q^\xi(\mu,x)^{-1}\right\|_\infty = \left\|Y^\xi(\mu,x)^{-1}(Y^x i(\mu,x) s^\xi(\mu,x))\right\|_\infty = \sqrt{\mu}\left\|Y^\xi(\mu,x)^{-1}\right\|_\infty \le \frac{\iota}{\sqrt{\mu}}\left(2\tilde{\iota}\|p\|_1 + \hat{n}\mu\right).$$

Since

$$u^T (R^\xi(\mu,x) R^\xi(\mu,x)^T) u = u^T (W Q^\xi(\mu,x)^2 W^T) u \ge (\min_i \{Q_{ii}^\xi(\mu,x)\})^2 u^T W W^T u$$

for all u, i.e., $R^{\xi}(\mu,x)R^{\xi}(\mu,x)^T \succeq (\min_i\{Q^{\xi}_{ii}(\mu,x)\})^2 WW^T$, we have

$$\frac{\hat{\iota}}{\mu}I \succeq \frac{1}{(\min_i\{Q^{\xi}_{ii}(\mu,x)\})^2}(WW^T)^{-1} \succeq (R^{\xi}(\mu,x)R^{\xi}(\mu,x)^T)^{-1}, \tag{2.14}$$

where $\hat{\iota} = \max\{\|T_j\|^2, \iota\,(2\tilde{\iota}\|p\|_1 + \hat{n}\mu)\}$ □

Proof of Theorem 2.1. Assumption A3 and the strict convexity (concavity) of the objective functions in (2.9–2.10) imply that for every $\mu > 0$ and $x \in \mathcal{P}_o$ the KKT-conditions (2.12) have a unique solution. Now, by using (2.12) for any $\mu > 0$, $\rho^{\xi}(\mu,x) = p^{\xi^T} y^{\xi}(\mu,x) - \mu\sum_{i=1}^{\hat{n}} y_i^{\xi}(\mu,x) = p^T y^{\xi}(\mu,x) + \mu\sum_{i=1}^{\hat{n}} \ln s_i^{\xi}(\mu,x) - \hat{n}\mu\ln\mu < \infty$ because of (2.13) in Lemma 2.1. Hence, $E[\rho(\mu,x)] < \infty$ for every $\infty > \mu > 0$. Let $y^{\xi^*} = argmin_{y^{\xi}\in\mathcal{P}^{\xi}(x)}\ p^{\xi^T} y^{\xi}$. Clearly, $0 \le p^{\xi^T}(y^{\xi}(\mu,x)) - p^{\xi^T} y^{\xi^*}$. Also, from (2.12) we have,

$$p^{\xi^T} y^{\xi}(\mu,x) - p^{\xi^T} y^{\xi^*} = s^{\xi^T}(\mu,x)(y^{\xi}(\mu,x) - y^{\xi^*}) \le \hat{n}\mu.$$

Hence,

$$0 \le E[\rho(\mu,x)] - E[\tilde{\rho}(x)] \le \hat{n}\mu, \tag{2.15}$$

and (ii) follows. The existence and uniqueness of $x(\mu)$ now follows because the set $\mathcal{P}_o$ is bounded with a non-empty interior and $\eta(x,\mu)$ is a strictly convex function. We now prove (iv). Let $\hat{x}(\mu)$ be a minimizer of (2.16):

$$\min_{x\in\mathcal{P}_o} \hat{\eta}(\mu,x), \qquad \hat{\eta}(\mu,x) := c^T x + E[\tilde{\rho}(x)] - \textstyle\sum_{i=1}^{n} \ln(x_i). \tag{2.16}$$

Observe that

$$\begin{aligned}
&\tilde{\eta}(x^*) - \tilde{\eta}(x(\mu)) \\
&= \tilde{\eta}(x^*) - \tilde{\eta}(\hat{x}(\mu)) + \tilde{\eta}(\hat{x}(\mu)) - \mu\sum_{i=1}^{n}\ln\hat{x}_i(\mu) + \mu\sum_{i=1}^{n}\ln\hat{x}_i(\mu) \\
&\quad -\tilde{\eta}(x(\mu)) - E[\rho(\mu,x(\mu))] + E[\rho(\mu,x(\mu))] - \mu\sum_{i=1}^{n}\ln x_i(\mu) + \mu\sum_{i=1}^{n}\ln x_i(\mu) \\
&= \tilde{\eta}(x^*) - \tilde{\eta}(\hat{x}(\mu)) + \hat{\eta}(\mu,\hat{x}(\mu)) \\
&\quad +\mu\sum_{i=1}^{n}\ln(\hat{x}_i(\mu)/x_i(\mu)) - \eta(\mu,x(\mu)) - E[\tilde{\rho}(x(\mu))] + E[\rho(\mu,x(\mu))].
\end{aligned} \tag{2.17}$$

We now bound terms in the right hand side of (2.17). Let $\partial\hat{\eta}(\mu,x)$ represent a subgradient (gradient if differentiable) of $\hat{\eta}(\mu,x)$ with respect to x, and $\hat{X}(\mu) = diag(\hat{x}_1(\mu),\ldots,\hat{x}_n(\mu))$. The KKT-conditions at $\hat{x}(\mu)$ ensure that

$$\partial\hat{\eta}(\mu,\hat{x}(\mu)) + A^T\pi = c + \partial E[\tilde{\rho}(\hat{x}(\mu))] - \hat{X}^{-1}(\mu) + A^T\pi = 0. \tag{2.18}$$

Hence,

$$(c + \partial E[\tilde{\rho}(\hat{x}(\mu))])^T(\hat{x}(\mu) - x^*) = \mu\hat{X}(\mu)^{-1}(\hat{x}(\mu) - x^*). \tag{2.19}$$

Since, $\partial\tilde{\eta}\left(\hat{x}(\mu)\right) = (c + \partial E[\tilde{\rho}(\hat{x}(\mu))])$ and because $\tilde{\eta}(\cdot)$ is a convex function, we have

$$0 \geq \tilde{\eta}(x^*) - \tilde{\eta}(\hat{x}(\mu)) \geq \left(\partial\tilde{\eta}\left(\hat{x}(\mu)\right)\right)^T (x^* - \hat{x}(\mu)) = \mu\hat{X}(\mu)^{-1}(x^* - \hat{x}(\mu)) \geq -n\mu. \tag{2.20}$$

Also,

$$\begin{aligned}
\hat{\eta}(\mu, \hat{x}(\mu)) - \eta(\mu, x(\mu)) &= \hat{\eta}(\mu, \hat{x}(\mu)) - \eta(\mu, x(\mu)) + E[\rho(\mu, \hat{x}(\mu))] - E[\rho(\mu, \hat{x}(\mu))] \\
&= \eta(\mu, \hat{x}(\mu)) - \eta(\mu, x(\mu)) + E[\tilde{\rho}(\hat{x}(\mu))] - E[\rho(\mu, \hat{x}(\mu))] \\
&\geq E[\tilde{\rho}(\hat{x}(\mu))] - E[\rho(\mu, \hat{x}(\mu))] \\
&\geq -\hat{n}\mu \quad (\text{using (2.15)}).
\end{aligned} \tag{2.21}$$

Furthermore,

$$\begin{aligned}
\hat{\eta}(\mu, \hat{x}(\mu)) - \eta(\mu, x(\mu)) &= \hat{\eta}(\mu, \hat{x}(\mu)) - \eta(\mu, x(\mu)) + E[\tilde{\rho}(x(\mu))] - E[\tilde{\rho}(x(\mu))] \\
&= \hat{\eta}(\mu, \hat{x}(\mu)) - \hat{\eta}(\mu, x(\mu)) + E[\tilde{\rho}(x(\mu))] - E[\rho(\mu, x(\mu))] \\
&\leq E[\tilde{\rho}(x(\mu))] - E[\rho(\mu, x(\mu))] \\
&\leq 0 \quad (\text{using (2.15)}).
\end{aligned} \tag{2.22}$$

From (2.18) we also have $\left(c + \partial E[\tilde{\rho}(\hat{x}(\mu))]\right)^T (\hat{x}(\mu) - x(\mu)) = \mu\hat{X}(\mu)^{-1}(\hat{x}(\mu) - x(\mu))$. Hence, we have a constant θ_1 satisfying

$$\begin{aligned}
e^T\hat{X}(\mu)^{-1}x(\mu) &= n + \frac{1}{\mu}\left(c - \partial E[\tilde{\rho}(\hat{x}(\mu))]\right)^T (\hat{x}(\mu) - x(\mu)) \\
&\leq n + \frac{1}{\mu}\left\|c - \partial E[\tilde{\rho}(\hat{x}(\mu))]\right\| \left\|\hat{x}(\mu) - x(\mu)\right\| \\
&\leq n + \frac{\theta_1}{\mu}.
\end{aligned} \tag{2.23}$$

The last inequality holds because $\mathcal{P}_o$ is bounded, hence $\|\hat{x}(\mu) - x(\mu)\|$ is bounded, and $\partial E[\tilde{\rho}(\hat{x}(\mu))]$ is bounded because the set of subgradients of a convex function is bounded [2, Proposition 1.1.2]. Similarly, using (2.11) and a bound on $\nabla E[\rho(\mu, x(\mu))]$ from (5.2) we have

$$e^TX(\mu)^{-1}\hat{x}(\mu) = n + \frac{1}{\mu}\left(c - \nabla E[\rho(\mu, x(\mu))]\right)^T (x(\mu) - \hat{x}(\mu)) \leq n + \frac{\theta_2}{\mu}. \tag{2.24}$$

The bounds in (2.23–2.24) imply $(n + \frac{\theta}{\mu})^{-1} \leq \hat{x}_i(\mu)/x_i(\mu) \leq n + \frac{\theta}{\mu}$, where $\theta := \max\{\theta_1, \theta_2\}$. Hence,

$$-n\ln\left(n + \frac{\theta}{\mu}\right) \leq \sum_{i=1}^{n} \ln(\hat{x}_i(\mu)/x_i(\mu))) \leq n\ln\left(n + \frac{\theta}{\mu}\right), \tag{2.25}$$

which implies that

$$\lim_{\mu\to 0}\left(\mu\sum_{i=1}^{n}(\ln\hat{x}_i(\mu)/x_i(\mu))\right) = 0. \tag{2.26}$$

The result in (iv) now follows by using (2.15,2.20–2.22,2.26) in (2.17). $\square$

2.3 Comparison of the Weighted Barrier with the Standard Barrier Formulation

In this section we discuss the subtle difference between the weighted barrier problems (2.7–2.10) and those considered in [11; 6; 8]. Zhao [11], and Mehrotra and Ozevin [6; 8] consider the following log-barrier decomposition problem in the discrete case:

$$\min_{x\in\mathcal{P}_o} c^T x - \mu \textstyle\sum_{i=1}^{n} \ln x_i + \sum_{i=1}^{K} \hat{\rho}^i(\mu, x), \text{ where} \tag{2.27}$$

$$\hat{\rho}^i(\mu, x) := \min_{y^i\in\mathcal{P}^i} \ \pi^i {p^i}^T y^i - \mu \textstyle\sum_{j=1}^{n^i} \ln y_j, \quad \text{for } i = 1, \ldots, K, \tag{2.28}$$

and show that the central path associated with (2.27–2.28) is same as the central path associated with the extensive barrier formulation of (2.4–2.5). Furthermore, the analysis in [11; 6; 8] shows that the number of first-stage iterations required in the primal barrier algorithm is of the same order as the number of iterations required to solve the extensive barrier formulation using standard primal (or primal-dual) interior point methods.

In comparison the weighted barrier (2.7–2.10) introduced in Mehrotra and Ozevin [7] considers

$$min_{x\in\mathcal{P}_o} \eta(\mu, x), \eta(\mu, x) := c^T x - \mu \textstyle\sum_{i=1}^{n} \ln x_i + E[\rho(\mu, x)], \tag{2.29}$$

$$E[\rho(\mu, x)] := \textstyle\sum_{i=1}^{K} \pi^i \rho^i(\mu, x), \tag{2.30}$$

$$\rho^i(\mu, x) := min_{y^i\in\mathcal{P}^i_o} \ {p^i}^T y^i - \mu \textstyle\sum_{j=1}^{n^i} \ln y^i_j, \text{ for } i = 1, \ldots, K. \tag{2.31}$$

The difference between (2.27–2.28) and (2.29–2.31) is that the latter scales the log-barrier for the second-stage problem with scenario probabilities. Consequently, we can view the weighted barrier as an expected value of the second-stage barrier subproblems. We denote the optimum solution of the first-stage problem (2.29) by $x(\mu)$ and the solution of the second-stage problems (2.31) by $(y^i(\mu, x), s^i(\mu, x))$. Note that for simplicity we have used the same notation to define $E[\rho(\mu, x)]$ and $\eta(\mu, x)$ for both the discrete and the continuous support case.

The extensive deterministic equivalent formulation of (2.29-2.31) is given by:

$$\begin{aligned} min \quad & c^T x - \mu \sum_{i=1}^{n} \ln x_i + \sum_{i=1}^{K} \pi^i \left(p_i^T y_i - \mu \sum_{j=1}^{n^i} \ln y_j \right) \\ s.t. \quad & Ax = b, x \geq 0, \\ & W_i y_i = h_i - T_i x, \ y_i \geq 0, \quad i = 1, \ldots, K. \end{aligned} \tag{2.32}$$

The following proposition states that the central path associated with (2.29-2.31) and (2.32) are same. Its proof follows from considering KKT-conditions for (2.29–2.31) and (2.32).

Proposition 2.1 *For a given $\mu > 0$, if $(x(\mu)^*, y^1(\mu)^*, \ldots y^K(\mu)^*)$ is the optimum solution of (2.32), then $x(\mu)^*$ is the optimum solution of (2.29), and $(y^1(\mu)^*, \ldots y^K(\mu)^*)$ are the optimal solutions of subproblems (2.31). Conversely, if for a given μ, $x(\mu)^*$ is the optimum solution of (2.29) and $(y^1(\mu)^*, \ldots, y^K(\mu)^*)$ are the optimum solutions of (2.31) with $x = x(\mu)^*$, then $(x(\mu)^*, y^1(\mu)^*, \ldots, y^K(\mu)^*)$ is the optimum solution of (2.32).* □

2.4 A Weighted Barrier Decomposition Algorithm

The weighted barrier decomposition algorithm presented in this section is a standard primal interior point method, which reduces μ by a constant factor at each iteration and seeks to approximate the minimizer $x(\mu)$ by taking one or more Newton steps. The novelty is in computing the Newton direction from the solutions of the decomposed second-stage problems using (6.3), (6.6), or (7.9), (7.11) in the finite support case. Starting from an appropriately centered solution, the algorithm approximately traces the central path. The procedure terminates with μ sufficiently small to generate a strictly feasible ϵ-optimum solution of (2.1), or (2.4) in the finite support case.

At an iterate x^k the Newton direction d_x is computed from

$$\nabla^2\eta(\mu^k, x^k)d_x - A^T d_\lambda = -\nabla\eta(\mu^k, x^k) + A^T\lambda, Ad_x = 0. \tag{2.33}$$

Let $\beta > 0, \gamma \in (0, 1)$ and $\theta > 0$ be suitable scalars. We make their values more precise in Theorems 8.1 and 8.2. The desired precision ϵ, a $\mu^0 > 0$ and a suitably centered initial point $x^0 \in \mathcal{P}_o$ are assumed given as input.

Weighted Barrier Decomposition Algorithm.

Initialization. $x = x^0$; $\mu = \mu^0$, $k = 0$.

Step 1.

1.1. Compute $\nabla\eta[\rho^\xi(x^k, \mu^k)], \nabla^2\eta[\rho^\xi(x^k, \mu^k)]$ from (6.3), (6.6), or (7.9), (7.11) in the finite support case.

1.2. Compute the Newton direction (d_x, d_λ) from (2.33).

1.3. Let $\delta(\mu^k, x^k) = \sqrt{\frac{1}{\mu^k} d_x^T[\nabla^2\eta(\mu^k, x^k)]d_x}$. If $\delta \leq \beta$ go to Step2.

1.4. Set $x^k = x^k + \theta d_x$ and $\lambda^k = \lambda^k + \theta d_\lambda$, and go to Step 1.1.

Step 2. If $\mu \leq \epsilon$ stop, otherwise set $\mu^{k+1} = \gamma\mu^k$, $x^{k+1} = x^k$, $\lambda^{k+1} = \lambda^k$, $k = k + 1$, and go to Step 1.1.

In the case of finite support, a practical approach to initialize the algorithm was studied in Mehrotra and Ozevin [7], while a theoretical approach was presented in Zhao [11]. In the above algorithm we assume that we can find exact solutions of the optimality conditions (2.12) and compute $\nabla E[\rho(\mu, x)]$ and $\nabla^2 E[\rho(\mu, x)]$ exactly. These assumptions considerably simplifies the complexity analysis. A practical implementation of this algorithm will use approximations of $\nabla E[\rho(\mu, x)], \nabla^2 E[\rho(\mu, x)]$ and an approximate solution of (2.12) (see [7] for further discussions).

The iterate x^k is close to the central path at iteration k of the algorithm, i.e. $\delta(\mu^k, x^k) \leq \beta$. In the short-step algorithm after reducing the parameter from μ^k to $\mu^{k+1} = \gamma\mu^k$, we have $\delta(\mu^{k+1}, x^k) \leq 2\beta$. In this variant just one Newton step with step size $\theta = 1$ (i.e., one loop of Step 1) suffice to give a new point x^{k+1} with $\delta(\mu^{k+1}, x^{k+1}) \leq \beta$. In the long-step algorithm μ is decreased by an arbitrarily constant factor ($\lambda \in (0, 1)$), and several damped Newton steps are taken to restore proximity $\delta(\mu^{k+1}, \cdot)$ to the central path. Although the worst case complexity of the long-step algorithm is worse than the complexity of the short-step algorithm, it performs much better in practice. The theoretical complexity results for these algorithms are given in Section 8.

3. Derivatives of the Barrier Recourse Function

In this section we give explicit expressions for derivatives of $y^\xi(\mu,x)$, $s^\xi(\mu,x)$ and $z^\xi(\mu,x)$ with respect to x and μ. We further give expressions for derivative of $\nabla_x \rho^\xi(\mu,x)$ with respect to μ. These expressions are used during the analysis in the following sections. We let $y^* := y^\xi(\mu,x)$, $s^* := s^\xi(\mu,x)$, and $z^* := z^\xi(\mu,x)$. Also, $Y^* = diag(y_1^*,\ldots,y_{\hat{n}}^*)$, $S^* = diag(s_1^*,\ldots,s_{\hat{n}}^*)$, $p := p^\xi$, $W := W^\xi$, and $T := T^\xi$. From (2.10–2.12) we have

$$\rho^\xi(\mu,x) = (h^\xi - T^\xi x)^T z^\xi(\mu,x) + \mu \sum_{j=1}^{n^\xi} \ln s_j^\xi(\mu,x) + n^\xi \mu(1-\ln\mu). \tag{3.1}$$

From (2.12) we get

$$z^* = (WW^T)^{-1} W(p - \mu Y^{*-1} e). \tag{3.2}$$

Also,

$$z^* = \left(WY^*S^{*-1}W^T\right) WY^*S^{*-1}p = (R^*R^{*T})^{-1}R^*Q^*(p-\mu Y^{*-1}e), \tag{3.3}$$

where

$$Q^* := Q^\xi(\mu,x) := Y^{*1/2}S^{*-1/2} = \frac{1}{\sqrt{\mu}}Y^* \quad \text{and} \quad R^* := R^*(\mu,x) := WQ^*. \tag{3.4}$$

3.1 Derivatives with respect to x

Differentiating (2.12) with respect to x gives

$$\begin{aligned} &Y^*\nabla_x s^* + S^*\nabla_x y^* = 0, \\ &W\nabla_x y^* = -T, \\ &W^T\nabla_x z^* + \nabla_x s^* = 0, \end{aligned} \tag{3.5}$$

where $\nabla_x s^* = \left[\frac{\partial s^*}{\partial x_1},\ldots,\frac{\partial s^*}{\partial x_n}\right]$, $\nabla_x y^* = \left[\frac{\partial s^*}{\partial x_1},\ldots,\frac{\partial s^*}{\partial x_n}\right]$, and $\nabla_x z^* = \left[\frac{\partial z^*}{\partial x_1},\ldots,\frac{\partial z^*}{\partial x_n}\right]$ are the Jacobian matrices, i.e., they are matrices whose columns are vectors of partial derivatives of s^*, y^* and z^* with respect to $x_1,\ldots,x_n$. The solution of (3.5) gives

$$\begin{aligned} \nabla_x z^* &= -\left(R^*R^{*T}\right)^{-1} T, \\ \nabla_x y^* &= -Q^*R^{*T}\left(R^*R^{*T}\right)^{-1} T, \\ \nabla_x s^* &= Q^{*-1}R^{*T}\left(R^*R^{*T}\right)^{-1} T. \end{aligned} \tag{3.6}$$

Differentiating (3.1), using KKT conditions (2.12), and Jacobian expressions from (3.6) gives

$$\nabla_x \rho^\xi(\mu,x) = -T^T z^* = -T^T(R^*R^{*T})^{-1}R^*Q^*(p-\mu Y^{*-1}e). \tag{3.7}$$

Also using (3.6) and differentiating (3.7) gives

$$\nabla_x^2 \rho^\xi(\mu,x) = -T^T\nabla_x z^* = T^T\left(R^*R^{*T}\right)^{-1} T. \tag{3.8}$$

3.2 Derivatives with respect to μ

We use the the symbols $(\cdot)'$ and $(\cdot)''$ to represent first and second derivatives of an entity with respect to μ. Differentiating (2.12) with respect to μ gives

$$\begin{aligned} &Y^* s^{*\prime} + S^* y^{*\prime} = e, \\ &W y^{*\prime} = 0, \\ &W^T z^{*\prime} + s^{*\prime} = 0. \end{aligned} \tag{3.9}$$

Solving (3.9) we obtain

$$\begin{aligned} z^{*\prime} &= -\frac{1}{\mu^{1/2}}(R^* R^{*T})^{-1} R^* e, \\ y^{*\prime} &= \frac{1}{\mu^{1/2}} Q^* (I - R^{*T}(R^* R^{*T})^{-1} R^*) e, \\ s^{*\prime} &= \frac{1}{\mu^{1/2}} Q^{*-1} R^{*T} (R^* R^{*T})^{-1} R^* e. \end{aligned} \tag{3.10}$$

Differentiating (3.7) with respect to μ and by using (3.10) we get

$$(\nabla_x \rho^\xi(\mu, x))' = -T^T z^{*\prime} = \frac{1}{\mu^{1/2}} T^T (R^* R^{*T})^{-1} R^* e. \tag{3.11}$$

Also differentiating (3.8) with respect to μ and using (3.11) we get

$$\begin{aligned} \left[\nabla_x^2 \rho^\xi(\mu, x)\right]' &= T^T \left[\left(R^* R^{*T}\right)^{-1}\right]' T \\ &= T^T \left(R^* R^{*T}\right)^{-1} \left[R^* R^{*T}\right]' \left(R^* R^{*T}\right)^{-1} T \\ &= T^T \left(R^* R^{*T}\right)^{-1} W \left[\{Q^*\}^2\right]' W^T \left(R^* R^{*T}\right)^{-1} T \\ &= \frac{1}{\mu} T^T \left(R^* R^{*T}\right)^{-1} W \left[Y^{*2}\right]' W^T \left(R^* R^{*T}\right)^{-1} T, \end{aligned} \tag{3.12}$$

where using (3.10) implies

$$\left[Y^{*2} e\right]' = 2 Y^* y^{*\prime} = 2 Q^{*2} (I - R^{*T}(R^* R^{*T})^{-1} R^*) e. \tag{3.13}$$

Now differentiating (3.1) and using (2.12) we obtain

$$\begin{aligned} \rho^\xi(\mu, x)' &= d^T y^{*\prime} - \sum_{i=1}^{\hat{n}} \ln y_i^* - \mu \sum_{i=1}^{\hat{n}} Y^{*-1} {y^{*\prime}}_i = (p - s^*)^T y^{*\prime} - \sum_{i=1}^{\hat{n}} \ln y_i^* \\ &= z^{*T} W y^{*\prime} - \sum_{i=1}^{n} \ln y_i^* = -\sum_{i=1}^{n} \ln y_i^* \qquad \text{(noting that } W y^{*\prime} = 0). \end{aligned}$$

Hence,

$$
\begin{aligned}
0 \le -\rho^{\xi''}(\mu,x) &= e^T Y^{*-1} y^{*\prime} && (3.14)\\
&= e^T Y^{*-1} Q^* (I - R^{*T}(R^* R^{*T})^{-1} R^*) Q^* Y^{*-1} e \\
&\qquad\qquad \text{(using (3.10) and } \mathrm{Y}^{*-1}\mathrm{Q}^*\mathrm{e} = \frac{1}{\sqrt{\mu}}\mathrm{e}) \\
&\le \|Y^{*-1} Q^* e\|_2^2 = \frac{\hat{n}}{\mu}. && (3.15)
\end{aligned}
$$

4. Analytic Properties of the Second-Stage Solutions

We will follow the notation from Section 3. The goal of this section is to show that all higher partial derivatives of y^*, s^*, and z^* are bounded. Consequently, we also show that y^*, s^*, and z^* are analytic functions of x and μ. To simplify notation throughout this section we define $x_0 := \mu$, $x := (x_0, x_1, \ldots, x_n)$ while treating the barrier parameter μ as a variable. We use μ to represent a particular choice of x_0. To simplify expressions we also assume that $\hat{n}\mu = O(1)$, $\|d^\xi\|, \|T^\xi\| = O(1), \iota = O(1)$, and $\tilde{\iota} = O(1)$.

4.1 Infinite Differentiability of the Second-Stage Solutions

Equations (3.5, 3.9) inductively define all higher derivatives of s^* and y^* with respect to x. In particular, let $x_i \in \{x_0, \ldots, x_n\}$, $\frac{\partial^k s^*}{\partial x_0^{k^0} \ldots \partial x_n^{k^n}}$, $\frac{\partial^k y^*}{\partial x_0^{k^0} \ldots \partial x_n^{k^n}}$ and $\frac{\partial^k z^*}{\partial x_0^{k^0} \ldots \partial x^{k^n}}$ represent the vector of k^ith partial derivatives of the vector functions $s^\xi(x), y^\xi(x)$, and $z^\xi(x)$, with respect to x_i, $k = \sum_{i=0}^n k^i$. Here k^i represents the number of times a partial derivative with respect to variable i is repeated. We follow the convention that the 0th order partial derivative of a vector function is the vector itself.

Proposition 4.1 *Let $y^* := y^\xi(\mu, x)$ $s^* := s^\xi(\mu, x)$, and $z^* := z^\xi(\mu, x)$ be the solution of (2.12). Redefine $x_0 := \mu$, and let $k^i, l^i \ge 0$ be integers and $k = \sum_{i=0}^n k^i, l = \sum_{i=0}^n l^i$. For $k \ge 2$, the partial derivatives of y^* and s^* with respect to $x_0, \ldots, x_n$ satisfy the recursion*

$$\sum_{l^0=0}^{k^0} \cdots \sum_{l^n=0}^{k^n} \begin{pmatrix} k^1 \\ l^1 \end{pmatrix} \cdots \begin{pmatrix} k^n \\ l^n \end{pmatrix} \frac{\partial^l Y^*}{\partial x_n^{l^n} \cdots \partial x_1^{l^1}} \frac{\partial^{k-l} s^*}{\partial x_n^{k^n - l^n} \cdots \partial x_1^{k^1 - l^1}} = 0. \tag{4.1}$$

In the special case where the partial derivatives are taken with respect to the same variable, the recursion is given by

$$\sum_{l=0}^{k} \begin{pmatrix} k \\ l \end{pmatrix} \frac{\partial^l Y^*}{\partial x_i^l} \frac{\partial^{k-l} s^*}{\partial x_i^{k-l}} = 0, i = 0, \ldots, n. \tag{4.2}$$

In particular, the partial derivatives with respect to $x_0, \ldots, x_n$ are the solution of the system of equations:

$$Y^* \left(\frac{\partial^k s^*}{\partial x_0^{k^0} \ldots \partial x_n^{k^n}} \right) + S^* \left(\frac{\partial^k y^*}{\partial x_0^{k^0} \ldots \partial x_n^{k^n}} \right) = v^{(k^0, \ldots, k^n)}, \text{ where} \tag{4.3}$$

$$v^{(k^0,\ldots,k^n)} := \sum_{\substack{l^0=0 \\ \sum_{j=0}^n l^j \neq 0, \sum_{j=0}^n k^j \neq k}}^{k^0} \cdots \sum_{l^n=0}^{k^n} \begin{pmatrix} k^0 \\ l^0 \end{pmatrix} \cdots \begin{pmatrix} k^n \\ l^n \end{pmatrix} \frac{\partial^l Y^*}{\partial x_n^{l^n} \cdots \partial x_0^{l^0}} \frac{\partial^{k-l} s^*}{\partial x_n^{k^n - l^n} \cdots \partial x_0^{k^0 - l^0}},$$

$$W\left(\frac{\partial^k y^*}{\partial x_0^{k^0} \cdots \partial x_n^{k^n}}\right) = 0,$$
$$W^T\left(\frac{\partial^k z^*}{\partial x_0^{k^0} \cdots \partial x_n^{k^n}}\right) + \left(\frac{\partial^k s^*}{\partial x_0^{k^0} \cdots \partial x_n^{k^n}}\right) = 0.$$

Proof: The proof of (4.2) follows by induction and taking the partial derivatives of the first equation in (3.5,3.9) repeatedly. The recursion in (4.1) follows by using (4.2) while taking partial derivatives with respect to variables $x_0, \ldots, x_n$. (4.3) follows from using the implicit function theorem. □

4.2 Bounds on Partial Derivatives of the Second-Stage Solutions

The discussion in Section 4.1 shows that s^*, y^*, and z^* are infinitely differentiable functions of x. We now show that all partial derivatives computed in (4.3) are bounded. We need the following well known result on Catalan numbers in the subsequent analysis.

Proposition 4.2 *Let* $p(1) = 1$ *and* $p(k) = \sum_{l=1}^{k-1}(p(l)(p(k-l))$, $k \geq 2$. *Then,* $p(k) = \frac{1}{k}\begin{pmatrix} 2(k-1) \\ k-1 \end{pmatrix} \leq 4^{k-1}/k$.

Proof: $p(k)$ is the $k-1$st Catalan number C_{k-1}. For Catalan numbers, $C_k = \frac{2k!}{k!\ k+1!}$. □

Lemma 4.1 *For* $i = 0, \ldots, n$, $k \geq 2$, *let* $\Delta_i^k y^\xi = \frac{1}{k!}\frac{\partial^k y^\xi}{\partial x_i^k}$ and $\Delta_i^k s^\xi = \frac{1}{k!}\frac{\partial^k s^\xi}{\partial x_i^k}$. *Then,*

$$\|Q^\xi \Delta_i^k s^\xi\|, \|Q^{\xi^{-1}} \Delta_i^k y^\xi\| \leq \frac{\theta^k}{\mu^{(k-1)/2}} p(k),$$

where $p(k)$ *is the* $k-1$*st Catalan number, and* $\theta = \max\{\hat{\imath}, \hat{n}\}/\mu^{1/2}$.

Proof: For $k = 1$ from (3.6), (3.10) and using Lemma 2.1 we have

$$\left\|Q^{*-1}\frac{\partial y^*}{\partial x_j}\right\|_2 = \|Q^{*-1}\Delta_j^1 y^*\| = \left(T_j^T(R^* R^{*T})^{-1}T_j\right)^{1/2} = \theta_j \leq \theta, j \geq 1,$$
$$\left\|Q^*\frac{\partial s^*}{\partial x_j}\right\|_2 = \|Q^*\Delta_j^1 s^*\| = \left(T_j^T(R^* R^{*T})^{-1}T_j\right)^{1/2} = \theta_j \leq \theta, j \geq 1,$$
$$\left\|Q^{*-1}\frac{\partial y^*}{\partial x_0}\right\|_2 = \|Q^{*-1}\Delta_0^1 y^*\| \leq \frac{\hat{n}}{\mu^{1/2}} \leq \theta,$$
$$\left\|Q^*\frac{\partial s^*}{\partial x_0}\right\|_2 = \|Q^{*-1}\Delta_0^1 s^*\| \leq \frac{\hat{n}}{\mu^{1/2}} \leq \theta.$$

Now assume that the result is true for $l = 1, \ldots, k-1$. From (4.2) we have

$$\sum_{l=0}^k \left(\Delta_i^k y_j^*\right)\left(\Delta_i^{k-l} s_j^*\right) = 0, i = 1, \ldots, n, j = 1, \ldots, \hat{n},$$

or equivalently, for $k \geq 2$

$$Y^*\Delta_i^k s^* + S^*\Delta_i^k y^* = -\sum_{l=1}^{k-1}\left(\Delta_i^k Y^*\right)\left(\Delta_i^{k-l}s^*\right) = 0, i = 1,\dots,n, j = 1,\dots,\hat{n}. \tag{4.4}$$

Multiplying both sides in (4.4) by $(Y^*S^*)^{-1/2} = \frac{1}{\sqrt{\mu}}I$ gives

$$Q^*\Delta_i^k s^* + Q^{*-1}\Delta_i^k y^* = -\frac{1}{\sqrt{\mu}}\sum_{l=1}^{k-1}\left(\Delta_i^k Y^*\right)\left(\Delta_i^{k-l}s^*\right) = 0, i =,\dots,n. \tag{4.5}$$

Since from the last two equations in (4.3), for $k \geq 2$, $(Q^*\Delta_i^k s^*)^T(Q^{*-1}\Delta_i^k y^*) = 0$, from (4.5) we have

$$\begin{aligned}
\left\|Q^*\Delta_i^k s^*\right\|, \left\|Q^{*-1}\Delta_i^k y^*\right\| &\leq \frac{1}{\sqrt{\mu}}\left\|\sum_{l=1}^{k-1}\left(Q^{*-1}\Delta_i^k y^*\right)\left(Q^*\Delta_i^{k-l}s^*\right)\right\| \\
&\leq \frac{1}{\sqrt{\mu}}\sum_{l=1}^{k-1}\left\|Q^{*-1}\Delta_i^k Y^*\right\|\left\|Q^*\Delta_i^{k-l}s^*\right\| \\
&\leq \frac{1}{\sqrt{\mu}}\sum_{l=1}^{k-1}\left(p(l)\frac{\theta^l}{\mu^{(l-1)/2}}\right)\left(p(k-l)\frac{\theta^{(k-l)}}{\mu^{(k-l-1)/2}}\right) \\
&= \frac{\theta^k}{\mu^{(k-1)/2}}\sum_{l=1}^{k-1}p(l)p(k-l) = \frac{\theta^k}{\mu^{(k-1)/2}}p(k).
\end{aligned}$$

Here the last equality used Proposition 4.2. □

The following technical combinatorial equality is needed to establish a bound on the partial derivatives of y^* and s^* with respect to several variables.

Proposition 4.3 *Let $k \geq l \geq 1$ be given integers. Then,*

$$\binom{k}{l} = \sum_{\substack{\sum_{i=1}^n l^k = l \\ \sum_{i=1}^n k^i = k \\ l^i \in \{0,\dots,k^i\}}} \binom{k^1}{l^1}\cdots\binom{k^n}{l^n}.$$

Proof: As in (4.2) taking l derivatives of a function $a(t)b(t) = 0$, $a(t), b(t) : \mathbb{R} \to \mathbb{R}$ with respect to t we get

$$\sum_{l=0}^{k}\binom{k}{l}\frac{d^l a(t)}{dt^l}\frac{d^{k-l}b(t)}{dt^{k-l}} = 0. \tag{4.6}$$

If we stop condensing the terms after $l^1,\dots,l^n$ steps, (following the step used to obtain (4.1)) while taking the derivatives of $a(t)b(t) = 0$ we get

$$
\begin{aligned}
0 &= \sum_{l^1=0}^{k^1} \cdots \sum_{l^n=0}^{k^n} \begin{pmatrix} k^1 \\ l^1 \end{pmatrix} \cdots \begin{pmatrix} k^n \\ l^n \end{pmatrix} \frac{d^{\sum_{i=1}^n l^i} a(t)}{dx^{\sum_{i=1}^n l^i}} \frac{d^{\sum_{i=1}^n k^i - l^i} b(t)}{dt^{\sum_{i=1}^n k^i - l^i}} \\
&= \sum_{l=0}^{k} \sum_{\substack{\sum_{i=1}^n l^k = l \\ \sum_{i=1}^n k^i = k \\ l^i \in \{0,\ldots,k^i\}}} \begin{pmatrix} k^1 \\ l^1 \end{pmatrix} \cdots \begin{pmatrix} k^n \\ l^n \end{pmatrix} \frac{d^l a(t)}{dt^l} \frac{d^{t-l} b(t)}{dt^{k-l}}. \qquad (4.7)
\end{aligned}
$$

The result follows by comparing the coefficients of each term in (4.6) with (4.7). $\square$

The following Lemma bounds the partial derivatives of y^* and s^* with respect to several variables. Lemma 4.1 is a special case of Lemma 4.2.

Lemma 4.2 *Let* $k = \sum_{i=0}^n k^i$ *and define* $\Delta^{(k^0,\ldots,k^n)} y^* = \frac{1}{k!} \frac{\partial^k y^*}{\partial x_0^{k^0} \cdots \partial x_n^{k^n}}$, $\Delta^{(k^0,\ldots,k^n)} s^* = \frac{1}{k!} \frac{\partial^k s^*}{\partial x_0^{k^0} \cdots \partial x_n^{k^n}}$. *Then,* $\left\| Q^{*-1} \Delta^{(k^0,\ldots,k^n)} y^* \right\|_2 \leq \frac{\theta^k p(k)}{\mu^{k-1/2}}$, *and* $\left\| Q^{*-1} \Delta^{(k^0,\ldots,k^n)} s^* \right\|_2 \leq \frac{\theta^k p(k)}{\mu^{k-1/2}}$, *where* $p(k)$ *is the* $k-1st$ *Catalan number.*

Proof: The result is proved for $k = 1$ in Lemma 4.1. Substituting the above definitions in (4.1) gives

$$
\sum_{l^0=0}^{k^0} \cdots \sum_{l^n=0}^{k^n} l!(k-l)! \begin{pmatrix} k^0 \\ l^0 \end{pmatrix} \cdots \begin{pmatrix} k^n \\ l^n \end{pmatrix} \left(\Delta^{(l^0,\ldots,l^n)} Y^* \right) \Delta^{(k^0-l^0,\ldots,k^n-l^n)} s^* = 0.
$$

Hence,

$$
\begin{aligned}
& Y^* \Delta^{(k^0,\ldots,k^n)} s^* + S^* \Delta^{(k^0,\ldots,k^n)} y^* \\
&= - \sum_{\substack{l^0=0 \\ \sum_{j=0}^n l^j \neq 0, \sum_{j=0}^n k^j \neq k}}^{k^0} \cdots \sum_{l^n=0}^{k^n} l!(k-l)! \begin{pmatrix} k^0 \\ l^0 \end{pmatrix} \cdots \begin{pmatrix} k^n \\ l^n \end{pmatrix} \left(\Delta^{(l^0,\ldots,l^n)} Y^* \right) \Delta^{(k^0-l^0,\ldots,k^n-l^n)} s^*,
\end{aligned}
$$

which gives

$$
\begin{aligned}
& Q^* \Delta^{(k^0,\ldots,k^n)} s^* + Q^{*-1} \Delta^{(k^0,\ldots,k^n)} y^* \\
&= -\frac{1}{\sqrt{\mu}} \sum_{\substack{l^0=0 \\ \sum_{j=0}^n l^j \neq 0, \sum_{j=0}^n k^j \neq k}}^{k^0} \cdots \sum_{l^n=0}^{k^n} l!(k-l)! \begin{pmatrix} k^0 \\ l^0 \end{pmatrix} \cdots \begin{pmatrix} k^n \\ l^n \end{pmatrix} \left(\Delta^{(l^0,\ldots,l^n)} Y^* \right) \Delta^{(k^0-l^0,\ldots,k^n-l^n)} s^*.
\end{aligned}
$$

By induction assume that $\left\| Q^{*-1} \Delta^{(l^0,\ldots,l^n)} y^* \right\| \leq \frac{\theta^l p(l)}{\mu^{(l-1)/2}}$ and $\left\| Q^* \Delta^{(l^0,\ldots,l^n)} s^* \right\| \leq \frac{\theta^l p(l)}{\mu^{(l-1)/2}}$, for all $\{l^0,\ldots,l^n\}$ such that $\sum_{j=0}^n l^j \leq k-1$. Now for any $k^0 \ldots k^n$ such that $\sum_{j=0}^n k^j = k$, since $(Q^* \Delta^{(k^0,\ldots,k^n)} s^*)^T (Q^{*-1} \Delta^{(k^0,\ldots,k^n)} y^*) = 0$, we have

$$
\left\| Q^{*-1} \Delta^{(k^0,\ldots,k^n)} y^* \right\|_2, \left\| Q^* \Delta^{(k^0,\ldots,k^n)} s^* \right\|_2
$$

$$
\begin{aligned}
&= \frac{1}{k!\sqrt{\mu}} \left\| v^{k^0,\ldots,k^n} \right\|_2 \\
&\le \frac{1}{k!\sqrt{\mu}} \left\| \sum_{\substack{l^0=0 \\ l=\sum_{j=0}^n l^j \neq 0, \\ k=\sum_{j=0}^n k^j \neq k}}^{k^0} \cdots \sum_{l^n=0}^{k^n} l!(k-l)! \binom{k^0}{l^0} \cdots \binom{k^n}{l^n} Q^{*-1} \Delta^{(l^0,\ldots,l^n)} Y^* Q^* \Delta^{(k^0-l^0,\ldots,k^n-l^n)} s^* \right\| \\
&\le \frac{1}{k!\sqrt{\mu}} \sum_{\substack{l^0=0 \\ l=\sum_{j=0}^n l^j \neq 0, \\ k=\sum_{j=0}^n k^j \neq k}}^{k^0} \cdots \sum_{l^n=0}^{k^n} l!(k-l)! \binom{k^0}{l^0} \cdots \binom{k^n}{l^n} \left\| Q^{*-1} \Delta^{(l^0,\ldots,l^n)} y^* \right\| \left\| Q^* \Delta^{(k^0-l^0,\ldots,k^n-l^n)} s^* \right\| \\
&\le \frac{1}{k!\sqrt{\mu}} \sum_{\substack{l^0=0 \\ \sum_{j=0}^n l^j \neq 0, \sum_{j=0}^n k^j \neq k}}^{k^0} \cdots \sum_{l^n=0}^{k^n} \left(\frac{l!(k-l)! \binom{k^0}{l^0} \cdots \binom{k^n}{l^n} p(l)\theta^l p(k-l)\theta^{k-l}}{\mu^{l-1/2}\mu^{k-l-1/2}} \right) \\
&= \frac{\theta^k}{k!\mu^{k-1/2}} \sum_{\substack{l^0=0 \\ \sum_{j=0}^n l^j \neq 0, \sum_{j=0}^n k^j \neq k}}^{k^0} \cdots \sum_{l^n=0}^{k^n} \left(l!(k-l)! p(l) p(k-l) \binom{k^0}{l^0} \cdots \binom{k^n}{l^n} \right) \\
&= \frac{\theta^k}{k!\mu^{k-1/2}} \sum_{l=0}^{k} \sum_{\substack{\sum_{i=0}^n l^k = l \\ \sum_{i=0}^n k^i = k \\ l^i \in \{0,\ldots,k^i\}}} \left(l!(k-l)! p(l) p(k-l) \binom{k^0}{l^0} \cdots \binom{k^n}{l^n} \right) \\
&= \frac{\theta^k}{k!\mu^{k-1/2}} \sum_{l=0}^{k} l!(k-l)! p(l) p(k-l) \sum_{\substack{\sum_{i=0}^n l^k = l \\ \sum_{i=0}^n k^i = k \\ l^i \in \{0,\ldots,k^i\}}} \binom{k^0}{l^0} \cdots \binom{k^n}{l^n}
\end{aligned}
$$

$$
= \frac{\theta^k}{k!\mu^{k-1/2}} \sum_{l=0}^{k} l!(k-l)! p(l) p(k-l) \binom{k}{l} = \frac{\theta^k}{\mu^{k-1/2}} \sum_{l=0}^{k} p(l) p(k-l) = \frac{\theta^k p(k)}{\mu^{k-1/2}} \quad \square
$$

Lemma 4.3 *Let $k = \sum_{j=0}^n k^j$. Then, for some constant $\hat{\theta} \ge 0$ such that $\hat{\theta}^k \ge \iota^{k+1} p(k)$,*

$$
\left\| \frac{\partial^k y^*}{\partial x_n^{k^n} \cdots \partial x_0^{k^0}} \right\|_2, \left\| \frac{\partial^k s^*}{\partial x_n^{k^n} \cdots \partial x_0^{k^0}} \right\|_2 \le \frac{\hat{\theta}^k k!}{\mu^k}, \tag{4.8}
$$

$$
\left\| \frac{\partial^k z^*}{\partial x_n^{k^n} \cdots \partial x_0^{k^0}} \right\| \le \frac{\hat{\theta}^k k!}{\mu^k}. \tag{4.9}
$$

Proof: The inequalities in (4.8) follow from combining Lemma 4.2, Lemma 2.1, (2.14), and noting that $p(k) = \frac{1}{k}\begin{pmatrix} 2(k-1) \\ k-1 \end{pmatrix} \leq 4^{k-1}/k$. From (3.6) we have $\frac{\partial z^*}{\partial x_i} = (R^\xi R^{\xi^T})^{-1}T_i$, and from (3.10) we have $\frac{dz^*}{dx_0} = (R^* R^{*T})^{-1}R^* e$. Now from (4.3) we also see that for $k \geq 2$,

$$\frac{\partial^k z^*}{\partial x_0^{k^0} \dots \partial x_n^{k^n}} = (R^* R^{*T})^{-1} W S^{*-1} v^{k^0,\dots,k^n} = \frac{1}{\sqrt{\mu}}(R^* R^{*T})^{-1} R^* v^{k^0,\dots k^n}.$$

Hence,

$$\left\| \frac{\partial^k z^*}{\partial x_0^{k^0} \dots \partial x_n^{k^n}} \right\|^2 = \frac{1}{\mu}\left(\hat{v}^{k^0,\dots,k^n}\right)^T (R^* R^{*T})^{-1} \hat{v}^{k^0,\dots,k^n}, \tag{4.10}$$

where $\hat{v}^{k^0,\dots,k^n} := (R^* R^{*T})^{-1/2} R^* v^{k^0,\dots k^n}$ if $k \geq 2$, and $\hat{v}^{k^0,\dots,k^n} = (R^* R^{*T})^{-1/2} T_i$, if $k = 1$, and $k^i = 1$, $i = 1, \dots, n$, $\hat{v}^{k^0,\dots,k^n} = (R^* R^{*T})^{-1/2} R^* e$, if $k = 1$, and $k^0 = 1$. Now by observing that $(R^*)^T (R^* R^{*T})^{-1} R^*$ is an orthogonal project matrix and using (2.14) we have (4.9) for $k = 1$. Also using (2.14) and that $(R^*)^T (R^* R^{*T})^{-1} R^*$ is an orthogonal projection matrix from (4.10) we have

$$\left\| \frac{\partial^k z^*}{\partial x_0^{k^0}, \dots, \partial x_n^{k^n}} \right\| \leq \frac{\hat{\iota}}{\mu} \left\| v^{k^1,\dots,k^n} \right\| \leq \frac{\hat{\iota} p(k) \theta^k}{\mu^{k/2}} = \frac{\hat{\theta}^k k!}{\mu^k},$$

where the last inequality uses the bound on $\left\| \frac{1}{\sqrt{\mu}} v^{k^1,\dots,k^n} \right\|$ proved during the proof of Lemma 4.2. This concludes the proof of (4.9). □

4.3 Real Analytic Function Properties of the Second-Stage Solutions

Theorem 4.1 *Let y^*, s^*, z^* be solutions of (2.12). For any $x \in \mathcal{P}_o$ and $\mu > 0$, the functions y^*, s^*, and z^*, are real analytic functions of $x = (x_0, \dots, x_n)$.*

Proof: It is sufficient to show that the Taylor expansion of these functions agree with the function at all points in a neighborhood of a given point x. Let $\hat{x} = x + th$, $\sum_{i=0}^n |h_i| = 1$, $t > 0$. From the Taylor expansion of y_j^*:

$$\begin{aligned} y_j^*(\hat{x}) &= y_j^*(x) + \sum_{l=0}^{\infty} \frac{t^l}{l!} \left(\sum_{i=0}^n h_i \frac{\partial}{\partial x_i} \right)^l y_j^*(x) \\ &= y_j^*(x) + \sum_{l=0}^{k-1} \frac{t^l}{l!} \left(\sum_{i=0}^n h_i \frac{\partial}{\partial x_i} \right)^l y_j^*(x) + \frac{t^k}{k!} \left(\sum_{i=0}^n h_i \frac{\partial}{\partial x_i} \right)^k y_j^*(\bar{x}), \end{aligned}$$

for $\bar{x} = x + \bar{t}h, \bar{t} \in [0, t]$, and $k \geq 1$. Now,

$$\frac{t^k}{k!} \left(\sum_{i=0}^n h_i \frac{\partial}{\partial x_i} \right)^k y_j^*(\bar{x}) \leq \max_{\substack{k^i \geq 0 \\ \sum_{i=0}^n k^i = k}} \left| \frac{\partial^k y_j^*(\bar{x})}{\partial^{k^n} x_n \cdots \partial^{k^0} x_0} \right| \frac{t^k}{k!} \left(\sum_{i=0}^n |h_i| \right)^k \leq \frac{t^k \hat{\theta}^k}{\mu^k}, \tag{4.11}$$

where the last inequality follows from using Lemma 2.1, and observing that the definition of $\hat{\theta}$ in the bound of this Lemma is independent of x, since in Assumption A3 the constants ι, $\tilde{\iota}$ are independent of x. The bound in (4.11) goes to zero for t sufficiently small ($t < \mu^k/\hat{\theta}$)as $k \to \infty$. Hence, y^ξ is a real analytic function. The proof that s^ξ, and z^ξ are real analytic functions is similar. □

5. Analytic Properties of the Expected Barrier Recourse Function

Let $x := (x_1, \dots, x_n)$. In this section we show that the higher partial derivatives of $\nabla_x E\left[\rho^\xi(\mu,x)\right]$ in (2.8) are bounded. This property is used to commute the integral with differential when finding the derivatives of $E\left[\rho^\xi(\mu,x)\right]$ with respect to x for the stochastic programming problems with continuous support. We also show that $E[\rho^\xi(\mu,x)]$ is a real analytic function of x.

Proposition 5.1 *Let $x_0 := \mu$, and $\nabla_x \rho^\xi(\mu,x)$ be given as in (3.7). Then,*

$$\frac{\partial^k(\nabla_x \rho^\xi(\mu,x))}{\partial x_0^{k^0} \dots \partial x_n^{k^n}} = -{T^\xi}^T \frac{\partial^k z^\xi(\mu,x)}{\partial x_0^{k^0} \dots \partial x_n^{k^n}}. \tag{5.1}$$

Also, there exists a constant $\theta' \geq 0$ such that

$$\|\nabla_x \rho^\xi(\mu,x)\| \leq \theta'. \tag{5.2}$$

Furthermore, for $k \geq 1$

$$\left\| \frac{\partial^k \nabla_x \rho^\xi(\mu,x)}{\partial x_0^{k^0} \cdots \partial x_n^{k^n}} \right\| \leq \theta' \left(\frac{\hat{\theta}^k k!}{\mu^k} \right), \tag{5.3}$$

where $\hat{\theta}$ is defined in Lemma 4.3.

Proof: The relationship (5.1) follows from (3.7). From (3.7) and (3.2) we have

$$\begin{aligned}
(\nabla \rho^\xi)^T \nabla \rho^\xi &= {z^\xi}^T T^\xi {T^\xi}^T z^\xi \leq O({z^\xi}^T z^\xi) \\
&= O\left((p^\xi - \mu {Y^\xi}^{-1} e)^T {W^\xi}^T (W^\xi {W^\xi}^T)^{-2} W^\xi (p^\xi - \mu Y^\xi\right) \\
&\leq O(\|p^\xi - \mu {Y^\xi}^{-1} e\|^2) = O(\|p^\xi - s^\xi\|^2) := \theta'
\end{aligned}$$

Hence, $\|\nabla \rho^\xi\| \leq O(\|p^\xi\| + \mu \|{Y^\xi}^{-1} e\|)$. Now (5.2) follows from using (2.13). Also (5.3) follows from (5.1) and using Lemma 4.3. □

Theorem 5.1 *Let $E[\rho^\xi(\mu,x)] = \int_\Xi \rho^\xi(\mu,x) dF(\xi)$ be defined as in (2.8). Then,*

$$\nabla_x E[\rho^\xi(\mu,x)] = \int_\Xi \nabla_x \rho^\xi(\mu,x) dF(\xi), \tag{5.4}$$

$$\frac{d}{d\mu} \nabla_x E[\rho^\xi(\mu,x)] = \int_\Xi \frac{d}{d\mu} (\nabla_x \rho^\xi(\mu,x)) dF(\xi), \text{ and} \tag{5.5}$$

$$\nabla_x^2 E[\rho^\xi(\mu,x)] = \int_\Xi \nabla_x^2 \rho^\xi(\mu,x) dF(\xi). \tag{5.6}$$

Furthermore, for all $x \in \mathcal{P}_o$ and $\mu > 0$, $E[\rho^\xi(\mu,x)]$, $E[\nabla \rho^\xi(\mu,x)]$, and $E[\nabla^2 \rho^\xi(\mu,x)]$ are real analytic functions of x. In addition, $E[\nabla \rho^\xi(\mu,x)]$ and $E[\nabla^2 \rho^\xi(\mu,x)]$ are real analytic functions of μ.

Proof. Since $\rho^\xi(\mu,x)$ is a bounded function for all $\xi\in\Xi$, it is Lebesgue-integrable (also Riemann-integrable). Equalities (5.4–5.6) follow because from Proposition 5.1 the partial derivatives of $\rho^\xi(x,\mu)$ are bounded and Theorem B.2 applies. We now show that $E[\rho^\xi(x,\mu)]$ is a real analytic function of x. From Proposition 5.1 and Theorem B.2 we have

$$\frac{\partial^k E[\rho^\xi(\mu,x)]}{\partial^{k^n}x_n\cdots\partial^{k^1}x_1}=\int_\Xi \frac{\partial^k\rho^\xi(\mu,x)}{\partial^{k^n}x_n\cdots\partial^{k^1}x_1}dF(\xi). \tag{5.7}$$

From (5.7) and using (5.2 –5.3) we have

$$\left|\frac{\partial^k E[\rho^\xi(\mu,x)]}{\partial^{k^n}x_n\cdots\partial^{k^1}x_1}\right|\le\int_\Xi\left|\frac{\partial^k\rho^\xi(\mu,x)}{\partial^{k^n}x_n\cdots\partial^{k^1}x_1}\right|dF(\xi)\le\theta'\left(\frac{\hat\theta^{k-1}(k-1)!}{\mu^{k-1}}\right).$$

Now let $\hat x=x+th$, $\sum_{i=1}^n|h_i|=1$, $t>0$. The Taylor expansion of $E[\rho^\xi(\mu,x)]$ is given as:

$$\begin{aligned}E[\rho^\xi(\mu,\hat x)] &= E[\rho^\xi(\mu,x)]+\sum_{l=1}^{\infty}\frac{t^l}{l!}\left(\sum_{i=1}^n h_i\frac{\partial}{\partial x_i}\right)^l E[\rho^\xi(\mu,x)]\\ &= E[\rho^\xi(\mu,x)]+\sum_{l=1}^{k-1}\frac{t^l}{l!}\left(\sum_{i=1}^n h_i\frac{\partial}{\partial x_i}\right)^l E[\rho^\xi(\mu,x)]+\frac{t^k}{k!}\left(\sum_{i=1}^n h_i\frac{\partial}{\partial x_i}\right)^k E[\rho^\xi(\mu,\bar x)],\end{aligned}$$

where $\bar x=x+\bar th, \bar t\in[0,t]$, and $k\ge1$. Now,

$$\begin{aligned}\frac{t^k}{k!}\left(\sum_{i=1}^n h_i\frac{\partial}{\partial x_i}\right)^k E[\rho^\xi(\mu,\bar x)] &\le \max_{\substack{k^i\ge0\\ \sum_{i=1}^k k^i=k}}\left|\frac{\partial^k E[\rho^\xi(\mu,x)]}{\partial^{k^n}x_n\cdots\partial^{k^1}x_1}\right|\frac{t^k}{k!}\left(\sum_{i=1}^n|h_i|\right)^k &(5.8)\\ &\le \theta'\frac{t^{k-1}\hat\theta^{k-1}}{\mu^{k-1}}, &(5.9)\end{aligned}$$

where the last inequality follows from using (5.8), and observing that the definition of $\hat\theta$ is independent of x. Since for all $\mu>0$ the bound in (5.9) goes to zero for t sufficiently small as $k\to\infty$, we conclude that $E[\rho^\xi(\mu,\hat x)]$ is a real analytic function. $E[\nabla_x\rho^\xi(\mu,x)]$ and $E[\nabla_x^2\rho^\xi(\mu,x)]$ are proved to be a real analytic function of x and μ by following the above proof for each element. □

6. Self-Concordance Under a Probabilistic Assumption

In this section we show that the self-concordance parameters of $\eta(\mu,x)$ do not depend on the number of scenarios under a probabilistic assumption. For this purpose we use a concept of a self-concordant random variable and self-concordant random matrix [5]. Mehrotra [5] shows that many known random variables (uniform, gamma, Chi-square, Normal, bounded, etc.) are self-concordant with a constant random variable dependent self-concordance parameter.

6.1 The Gradient and Hessian of Expected Barrier Recourse Function

As a consequence of Theorem 5.1, (3.7), and (3.8) and we have

$$\begin{aligned}
\nabla_x \eta(\mu, x) &= c - \mu X^{-1} e + \nabla_x E[\rho^\xi(\mu, x)] && (6.1)\\
&= c - \mu X^{-1} e + \int_\Xi \nabla_x \rho^\xi(\mu, x) dF(\xi) && (6.2)\\
&= c - \mu X^{-1} e - \int_\Xi T^{\xi T} z^\xi(\mu, x) dF(\xi), && (6.3)\\
\nabla_x^2 \eta(\mu, x) &= \mu X^{-2} + \nabla_x^2 E[\rho^\xi(\mu, x)] && (6.4)\\
&= \mu X^{-2} + \int_\Xi \nabla_x^2 \rho^\xi(\mu, x) dF(\xi) && (6.5)\\
&= \mu X^{-2} + \int_\Xi T^{\xi T} \left(R^\xi R^{\xi T}\right)^{-1} T^\xi dF(\xi). && (6.6)
\end{aligned}$$

6.2 The Self-Concordance Property of the Barrier Recourse Family

Lemma 6.1 *For any $\mu > 0$, $\rho^\xi(\mu, x)$ is strongly μ-self-condordant on $\mathcal{P}_o$. In particular, for $x \in \mathcal{P}_o$ and $\mu > 0$,*

$$|\nabla_x^3 \rho^\xi(\mu, x)[h, h, h]| \le 2\mu^{-1/2}(h^T \nabla_x^2 \rho^\xi(\mu, x) h))^{3/2}. \quad (6.7)$$

Furthermore, $\rho^\xi(\mu, x)$ form a self-concordant family with parameters $\alpha(\mu) = \mu$, $\gamma(\mu) = \nu(\mu) = 1$, $\xi(\mu) = \sqrt{\hat{n}}/\mu$, $\sigma(\mu) = \frac{\sqrt{\hat{n}}}{\mu}$. In particular,

$$\left|\{h^T \nabla_x \rho^\xi(\mu, x)\}'_\mu\right| \le \sqrt{\frac{\hat{n}}{\mu}} \sqrt{h^T \nabla_x^2 \rho^\xi(\mu, x) h} \quad (6.8)$$

$$\left|\{h^T \nabla_x^2 \rho^\xi(\mu, x) h\}'_\mu\right| \le \frac{2\sqrt{\hat{n}}}{\mu} h^T \nabla_x^2 \rho^\xi(\mu, x) h. \quad (6.9)$$

Proof: Let $\rho(\mu, x) := \rho^\xi(\mu, x)$, $\phi(t) := h^T \nabla_x^2 \rho^\xi(\mu, x + th) h$, $R(t) := R^\xi(\mu, x + th)$, $Y(t) := Y^\xi(\mu, x + th)$, $S(t) := S^\xi(\mu, x + th)$, $Q(t) = Q^\xi(\mu, x + th)$, $R := R(0)$, $Q := Q(0)$, $T := T^\xi$, and $W := W^\xi$. Since $\nabla_x^2 \rho(\mu, x)$ in (3.8) is a positive definite matrix, $\rho(\mu, x)$ is a convex function. Also, $\frac{d\phi(0)}{dt} = \nabla_x^3 \rho(\mu, x)[h, h, h]$. Now,

$$\begin{aligned}
\left|\frac{d\phi(t)}{dt}\right| &= \left|h^T \frac{d}{dt}\left[\nabla_x^2 \rho(\mu, x + th)\right] h\right|\\
&= \left|h^T \frac{d}{dt}\left[T^T (R(t)R(t)^T)^{-1} T\right] h\right| \text{(using (3.8))}\\
&= \left|h^T T^T (R(t)R(t)^T)^{-1} \left[\frac{d}{dt}(R(t)R(t)^T)\right] (R(t)R(t)^T)^{-1} T h\right|\\
&= \left|h^T T^T (R(t)R(t)^T)^{-1} W \left[\frac{d}{dt}(Q(t)^2)\right] W^T (R(t)R(t)^T)^{-1} T h\right| && (6.10)
\end{aligned}$$

From (3.6)

$$\begin{aligned}
\left.\frac{d}{dt} Q^2(t) e\right|_{t=0} &= \frac{1}{\mu} \frac{d}{dt} Y^2(t) e|_{t=0} = \frac{2}{\mu} Y(0) \nabla_x y(0) h\\
&= -\frac{2}{\sqrt{\mu}} Q^3 W^T \left(RR^T\right)^{-1} T h && (6.11)
\end{aligned}$$

Letting $\bar{h} = W^T \left(RR^T\right)^{-1} Th$, and evaluating (6.10) at $t = 0$, and using (6.11) we have

$$\begin{aligned}
\left|\frac{d\phi(0)}{dt}\right| &= \frac{2}{\sqrt{\mu}} \sum_{i=1}^{\hat{n}} (Q_{ii}\bar{h}_i)^3 \\
&\leq \frac{2}{\sqrt{\mu}} \left(\sum_{i=1}^{\hat{n}} (Q_{ii}\bar{h}_i)^2\right)^{3/2} \quad (\text{using} \|\cdot\|_3 \leq \|\cdot\|_2) \\
&= \frac{2}{\sqrt{\mu}} \left(\sum_{i=1}^{\hat{n}} (QW^T \left(RR^T\right)^{-1} Th)^2\right)^{3/2} \\
&= \frac{2}{\sqrt{\mu}} \left(\sum_{i=1}^{\hat{n}} (h^T T^T \left(RR^T\right)^{-1} Th)^2\right)^{3/2} \\
&= \frac{2}{\sqrt{\mu}} (h^T \nabla_x^2 \rho(\mu, x) h))^{3/2} \quad (\text{using (3.8)}).
\end{aligned}$$

This proves (6.7). Now to show (6.9) observe that from (3.11) we have

$$\begin{aligned}
\left|\frac{d}{d\mu}\{h^T \nabla_x \rho(\mu, x)\}\right| &= \frac{1}{\mu^{1/2}} h^T T^T (RR^T)^{-1} Re \\
&\leq \frac{1}{\mu^{1/2}} \left(h^T T^T (RR^T)^{-1} h\right)^{1/2} (e^T R^T (RR^T)^{-1} Re)^{1/2} \\
&\leq \sqrt{\frac{\hat{n}}{\mu}} \sqrt{h^T \nabla_x^2 \rho(\mu, x) h},
\end{aligned}$$

where the last inequality follows because $R^T(RR^T)^{-1}R$ is an orthogonal project matrix. Now let $\bar{h} = W^T \left(RR^T\right)^{-1} Th$, and $\tilde{e} = (I - R^T(RR^T)^{-1}R)e$. Then, from (6.9) and using (3.12–3.13)

$$\left|\{h^T \nabla_x^2 \rho(\mu, x) h\}'_\mu\right| = \frac{2}{\mu} h^T T^T \left(RR^T\right)^{-1} W \left[\frac{d}{d\mu} Y^2\right] W^T \left(RR^T\right)^{-1} T^h \tag{6.12}$$

$$= \frac{2}{\mu} \sum_{i=1}^{\hat{n}} \bar{h}_i^2 Q_{ii}^2 \tilde{e}_i$$

$$\leq \frac{2}{\mu} \|\tilde{e}\| \, \|Q\bar{h}\|^2 \leq \frac{2\sqrt{\hat{n}}}{\mu} h^T \nabla_x^2 \rho(\mu, x) h. \quad \square \tag{6.13}$$

6.3 The Self-Concordance Property of the Expected Barrier Family

We now study the self-concordance properties of $\eta(\mu, x)$.

Definition 6.1 *Let $\varsigma : \Omega \to \mathbb{R}$ be a random variable with probability measure F^ς. We call ς an α-self-concordant random variable if*

$$\left|E[\varsigma^3]\right| \leq 2\alpha^{-1/2} (E[\varsigma^2])^{3/2}. \tag{6.14}$$

Let $P : \Omega \to \mathbb{R}^{m\times n}$ be a random matrix with probability measure F^{ς}. The matrix P is called an α-self-concordant random matrix if

$$\left|E[\|Ph\|^3]\right| \leq 2\alpha^{-1/2}(E[\|Ph\|^2])^{3/2}, \text{ for all } h \in \mathbb{R}^n. \tag{6.15}$$

Lemma 6.2 *Let $R^{\xi} := R^{\xi}(\mu, x)$, $\tilde{K} = \max\{1, \hat{K}\}$ and assume that $P^{\xi} := (R^{\xi}R^{\xi^T})^{-1}R^{\xi}$ is a $\frac{4}{\hat{K}}$-self-concordant random matrix. Then, the function $\eta(\mu, x)$ defined in (2.7) is strongly $\mu/\tilde{K}$-self-concordant on $\mathcal{P}_o$. In particular,*

$$\left|\nabla_x^3 \eta(\mu, x)[h, h, h]\right| \leq \frac{2\tilde{K}^{1/2}}{\sqrt{\mu}} \left(h^T \nabla_x^2 \eta(\mu, x) h\right)^{3/2}.$$

Proof. Since $P^{\xi}P^{\xi^T} = (R^{\xi}R^{\xi^T})^{-1}$, the self-concordance assumption on P^{ξ} implies that for all $h \in \mathbb{R}^n$

$$\begin{aligned}
\left|E[(h^T\nabla_x^2\rho^{\xi}(\mu,x)h)^{3/2}]\right| &= \left|E[\|P^{\xi}Th\|^3]\right| \\
&\leq \hat{K}^{1/2}\left(E[\|P^{\xi}Th\|^2]\right)^{3/2} \\
&= \hat{K}^{1/2}(E[(h^TT^{\xi^T}P^{\xi^T}P^{\xi}T^{\xi}h)])^{3/2} \\
&= \hat{K}^{1/2}(E[h^T\nabla_x^2\rho^{\xi}(\mu,x)h])^{3/2}. \qquad (6.16)
\end{aligned}$$

Now,

$$\begin{aligned}
\left|\nabla_x^3\eta(\mu,x)[h,h,h]\right| &\leq \left|\sum_{i=1}^n \mu(h_i/x_i)^3\right| + \left|\nabla_x^3 E[\rho^{\xi}(\mu,x)][h,h,h]\right| \\
&\leq \sum_{i=1}^n \mu(|h_i/x_i|)^3 + \left|\nabla_x^3 E[\rho^{\xi}(\mu,x)][h,h,h]\right| \\
&= \sum_{i=1}^n \mu(|h_i/x_i|)^3 + \left|E[\nabla_x^3\rho^{\xi}(\mu,x)[h,h,h]]\right| \quad \text{(using Theorem 5.1)} \\
&\leq \sum_{i=1}^n \mu(|h_i/x_i|)^3 + \frac{2}{\sqrt{\mu}}E[(h^T\nabla_x^2\rho^{\xi}(\mu,x)h)^{3/2}] \quad \text{(using Lemma 6.1)} \\
&\leq \sum_{i=1}^n \mu(|h_i/x_i|)^3 + \frac{2\hat{K}^{1/2}}{\sqrt{\mu}}(E[h^T\nabla_x^2\rho^{\xi}(\mu,x)h])^{3/2} \quad \text{(using (6.16))} \\
&\leq \frac{2\max\{1,\hat{K}^{1/2}\}}{\sqrt{\mu}}\left(\sum_{i=1}^n \mu^{3/2}(|h_i/x_i|)^3 + (E[h^T\nabla_x^2\rho^{\xi}(\mu,x)h])^{3/2}\right) \\
&\leq \frac{2\tilde{K}^{1/2}}{\sqrt{\mu}}\left(\sum_{i=1}^n \mu(h_i/x_i)^2 + (E[h^T\nabla_x^2\rho^{\xi}(\mu,x)h])\right)^{3/2} \quad \text{(using } \|\cdot\|_3 \leq \|\cdot\|_2\text{)} \\
&= \frac{2\tilde{K}^{1/2}}{\sqrt{\mu}}\left(\sum_{i=1}^n \mu(h_i/x_i)^2 + h^T\nabla_x^2 E[\rho^{\xi}(\mu,x)]h\right)^{3/2} \quad \text{(using Theorem 5.1)} \\
&= \frac{2\tilde{K}^{1/2}}{\sqrt{\mu}}\left(h^T\nabla_x^2\eta(\mu,x)h\right)^{3/2} \text{(using 6.4)} \quad \square
\end{aligned}$$

6.4 Parameters of the Self-Concordance Family

Theorem 6.1 *The family of functions $\eta(x,\mu)$ is a strongly self-concordant family with parameters $\alpha(\mu) = \frac{\mu}{\tilde{K}}$, $\gamma(\mu) = \nu(\mu) = 1$, $\xi(\mu) = \frac{\sqrt{\tilde{K}(n+\hat{n})}}{\mu}$ and $\sigma(\mu) = \frac{\sqrt{\hat{n}}}{\mu}$. In particular, for any $\mu > 0$, $x \in \mathcal{P}_o$ and $h \in \mathbb{R}^n$ we have*

$$\left|\frac{d}{d\mu}\{\nabla\eta(\mu,x)^T h\}\right| \leq \left[\frac{n+\hat{n}}{\mu}\nabla^2\eta(\mu,x)^T[h,h]\right]^{1/2}, \tag{6.17}$$

$$\{\nabla\eta(\mu,x)^T\}'_\mu[\nabla^2_x\eta(\mu,x)]^{-1}\{\nabla\eta(\mu,x)\}'_\mu \leq \mu^{-1}(n+\hat{n}), \tag{6.18}$$

$$\left|\frac{d}{d\mu}\{h^T\nabla^2\eta(\mu,x)h\}\right| \leq \frac{2\sqrt{\hat{n}}}{\mu}\nabla^2\eta(\mu,x)[h,h]. \tag{6.19}$$

Proof. Condition (SF1) in Definition A.2 is easy to verify. Lemma 6.2 shows (SF2) for our choice of $\alpha(\mu)$. We now show (SF3). By differentiating (6.3) with respect to μ, using (3.11) and Theorem 5.1 we get

$$\{\nabla\eta(x,\mu)\}'_\mu \quad = \quad -X^{-1}e + \frac{1}{\mu^{1/2}}\int_\Xi \left(T^{\xi^T}(R^\xi R^{\xi^T})^{-1}R^\xi e\right)dF(\xi), \tag{6.20}$$

$$\begin{aligned}
h^T\{\nabla\eta(x,\mu)\}' &= -h^T X^{-1}e + \frac{1}{\mu^{1/2}}\int_\Xi \left(h^T T^\xi (R^\xi R^{\xi^T})^{-1}R^\xi e\right)dF(\xi)\\
&= \int_\Xi \left(-h^T X^{-1}e + \frac{1}{\mu^{1/2}}h^T T^\xi (R^\xi R^{\xi^T})^{-1}R^\xi e\right)dF(\xi)\\
&\leq \mu^{-1/2}\left(\int_\Xi (\mu h^T X^{-2}h + h^T T^\xi (R^\xi R^{\xi^T})^{-1}T^\xi h)dF(\xi)\right)^{1/2}\cdot\\
&\qquad \left(\int_\Xi (n + \|(R^\xi R^{\xi^T})^{-1}R^\xi e\|^2)dF(\xi)\right)^{1/2} \text{(using Proposition C.1)}\\
&\leq \frac{(n+\hat{n})^{1/2}}{\mu^{1/2}}\left(h^T\nabla^2_x\eta(x)h\right)^{1/2},
\end{aligned}$$

where the last inequality uses the fact that $R^{\xi^T}(R^\xi R^{\xi^T})^{-1}R^\xi$ is an orthogonal project matrix, $\int_\Xi dF(\xi) = 1$, and (6.6). Also, using (3.11), and Proposition C.2 in the first inequality below, we have

$$\{\nabla\eta(\mu,x)^T\}'[\nabla^2_x\eta(\mu,x)]^{-1}\{\nabla\eta(\mu,x)\}'$$

$$= \frac{1}{\mu}\left(\int_\Xi \left[-X^{-1} \vdots \frac{1}{\mu^{1/2}} T^\xi (R^\xi R^{\xi^T})^{-1/2}\right] \begin{pmatrix} e \\ (R^\xi R^{\xi^T})^{-1/2} R^\xi e \end{pmatrix} dF(\xi)\right)^T$$

$$\left[\int_\Xi \left(\left[-X^{-1} \vdots \frac{1}{\mu^{1/2}} T^\xi (R^\xi R^{\xi^T})^{-1/2}\right]\left[-X^{-1} \vdots \frac{1}{\mu^{1/2}} T^\xi (R^\xi R^{\xi^T})^{-1/2}\right]^T dF(\xi)\right)\right]^{-1}$$

$$\left(\int_\Xi \left[-X^{-1} \vdots \frac{1}{\mu^{1/2}} T^\xi (R^\xi R^{\xi^T})^{-1/2}\right] \begin{pmatrix} e \\ (R^\xi R^{\xi^T})^{-1/2} R^\xi e \end{pmatrix} dF(\xi)\right)$$

$$\leq \frac{1}{\mu} \int_\Xi \begin{pmatrix} e \\ (R^\xi R^{\xi^T})^{-1/2} R^\xi e \end{pmatrix}^T \begin{pmatrix} e \\ (R^\xi R^{\xi^T})^{-1/2} R^\xi e \end{pmatrix} dF(\xi)$$

$$= \frac{1}{\mu}\left(n + \int_\Xi \left(e^T R^{\xi^T} (R^\xi R^{\xi^T})^{-1} R^\xi e\right) dF(\xi)\right) \leq \frac{1}{\mu}(n + \hat{n}).$$

From (6.6) and differentiating with respect to μ and using Theorem 5.1, for any $h \in \mathbb{R}^n$ we have

$$\begin{aligned}
\left|\frac{d}{d\mu}\{h^T \nabla^2 \eta(\mu, x) h\}\right| &= \left|h^T X^{-2} h + \int_\Xi \frac{d}{d\mu}(h^T \nabla_x^2 \rho^\xi(x,\mu) h) dF(\xi)\right| \\
&\leq h^T X^{-2} h + \int_\Xi \left|\frac{d}{d\mu}(h^T \nabla_x^2 \rho^\xi(x,\mu) h)\right| dF(\xi) \\
&\leq h^T X^{-2} h + \frac{2\sqrt{\hat{n}}}{\mu} \int_\Xi (h^T \nabla_x^2 \rho^\xi(x,\mu) h) dF(\xi) \text{ (using Lemma 6.1)} \\
&\leq \frac{2\sqrt{\hat{n}}}{\mu} h^T \nabla^2 \eta(x,\mu) h \quad \square
\end{aligned}$$

7. Properties of the WB Recourse Function under Finite Support

In this section we establish the self-concordance properties of the barrier recourse function in the worst case under the finite support assumption. The analysis of this section parallels the analysis in Section 6 for the discrete case. The barrier decomposition algorithm based on the bounds given in this section achieves an iteration complexity that is slightly worse than the iteration complexity proved in Zhao [11]. In particular, the long step iteration complexity in Zhao [11] is $O(n + K\hat{n}\ln(1/\epsilon))$ as compare to $O(K(n+\hat{n})\ln(1/\epsilon))$ proved here for the algorithm in Section 2.4. This is because the self-concordance parameter α for $\eta(x,\mu)$ in Lemma 7.2 is worse than the self-concordance parameter for the log-barrier function (2.27).

The dual of (2.31) is:

$$max_{(z^i,s^i)\in\mathcal{D}^i(x)} (h^i - T^i x)^T z^i, \quad \mathcal{D}^i(x) := \{(z^i, s^i) \mid W^{i^T} z^i + s^i = d^i, s^i \geq 0\}. \tag{7.1}$$

The barrier problem associated with (7.1) is defined as:

$$max_{(z^i,s^i)\in\mathcal{D}^i(x)} - (h^i - T^i x)^T z^i + \mu \textstyle\sum_{j=1}^{n^i} \ln s^i_j, \tag{7.2}$$

$$\text{and } \rho^i(\mu, x) = (h^i - T^i x)^T z^i(\mu, x) + \mu \textstyle\sum_{j=1}^{n^i} \ln s^\xi_j(\mu, x) + n^i \mu (1 - \ln \mu). \tag{7.3}$$

Since the objective in problems (2.31) and (7.2) are respectively convex and concave, y^i and (z^i, s^i) are optimal solutions of (2.31) and (7.2), respectively, if and only if they satisfy the KKT conditions:

$$Y^i s^i = \mu e, W^i y^i = h^i - T^i x, W^{i^T} z^i + s^i = d^i, y^i > 0,\ \ s^i > 0. \tag{7.4}$$

From (7.4) we get

$$z^i = (R^i R^{i^T})^{-1} R^i Q^i (d^i - Y^{i^{-1}} e),\ \text{where}\ Q^i := Y^{i^{1/2}} S^{i^{-1/2}} \quad \text{and} \quad R^i := W^i Q^i. \tag{7.5}$$

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

Let $(y^i, z^i, s^i) := (y^i(\mu, x), z^i(\mu, x), s^i(\mu, x))$ be the optimum solution of (2.31,7.1). Differentiating (7.4) with respect to x and μ we get

$$\begin{array}{ll} Y^i \nabla_x s^i + S^i \nabla_x y^i = 0, & Y^i s^{i\prime} + S^i y^{i\prime} = e, \\ W^i \nabla_x y^i = -T^i, & W^i y^{i\prime} = -T^i, \\ W^{i^T} \nabla_x z^i + \nabla_x s^i = 0, & W^{i^T} z^{i\prime} + s^{i\prime}, = 0 \end{array} \tag{7.6}$$

where $\nabla_x s^i, \nabla_x y^i$, and $\nabla_x z^i$ are Jacobian matrices. Solving the system (7.6) we obtain

$$\begin{array}{ll} \nabla_x z^i = -\left(R^i R^{i^T}\right)^{-1} T^i, & z^{i\prime} = -\frac{1}{\mu^{1/2}} \left(R^i R^{i^T}\right)^{-1} R^i e, \\ \nabla_x y^i = -Q^{i^2} W^{i^T} \left(R^i R^{i^T}\right)^{-1} T^i, & y^{i\prime} = \frac{1}{\mu^{1/2}} Q^i \left(I - R^i R^{i^T}\right)^{-1} R^i e \\ \nabla_x s^i = W^{i^T} \left(R^i R^{i^T}\right)^{-1} T^i, & s^{i\prime} = \frac{1}{\mu^{1/2}} Q^{i^{-1}} R^{i^T} \left(R^i R^{i^T}\right)^{-1} R^i e \end{array} \tag{7.7}$$

Differentiating (7.3) and using the optimality conditions (7.4) and (7.7) we get

$$\nabla_x \rho^i(\mu, x) = -T^{i^T} z^i(\mu, x), \quad \{\nabla_x \rho^i(\mu, x)\}' = -T^{i^T} z^{i\prime} = \frac{1}{\mu^{1/2}} T^{i^T} \left(R^i R^{i^T}\right)^{-1} R^i e \tag{7.8}$$

Hence,

$$\nabla_x \eta(\mu, x) = c - \mu X^{-1} e - \sum_{i=1}^{K} \pi^i \nabla_x \rho^i(x) = c - \mu X^{-1} e - \sum_{i=1}^{K} \pi^i T^{i^T} z^i(\mu, x), \tag{7.9}$$

$$\nabla_x^2 \eta(\mu, x) = \mu X^{-2} + \sum_{i=1}^{K} \pi^i \nabla_x^2 \rho^i(x) = \mu X^{-2} - \sum_{i=1}^{K} \pi^i T^{i^T} \nabla_x z^i(\mu, x). \tag{7.10}$$

Then substituting for $\nabla_x z^i$ in (7.10) we get

$$\nabla_x^2 \eta(\mu, x) = \mu X^{-2} + \sum_{i=1}^{K} \pi^i T^{i^T} \left(R^i R^{i^T}\right)^{-1} T^i. \tag{7.11}$$

7.2 Self-Concordance of the Recourse Function

The following lemma is a consequence of Nesterov and Nemirovskii [9, Proposition 5.1.5] as noted by Zhao [11, Lemma 1] in the context of linear two-stage stochastic programs. See Mehrotra and Ozevin [6] for a direct proof.

Lemma 7.1 *For any $\mu > 0$, $\rho^i(\mu, x)$ is strongly μ-self-condordant on $\mathcal{P}^{1,i}, i = 1, \ldots, K$. In particular,*

$$|\nabla_x^3 \rho^i(\mu, x)[h, h, h]| \leq 2\mu^{-1/2}(\nabla_x^2 \rho^i(x, \mu))[h, h])^{3/2}. \tag{7.12}$$

Furthermore, $\rho^i(\mu, x)$ form a self-concordant family with parameters $\alpha(\mu) = 1$, $\gamma(\mu) = \nu(\mu) = \mu$, $\xi(\mu) = \sqrt{n^i}/\mu$, $\sigma(\mu) = \sqrt{n^i/\mu}$. In particular,

$$|\{h^T \nabla_x \rho^i(\mu, x)\}'| \leq \frac{\sqrt{n^i}}{\mu} h^T \nabla_x^2 \rho^i(\mu, x) h, \tag{7.13}$$

$$|\{h^T \nabla_x^2 \rho^i(\mu, x) h\}'| \leq \frac{2\sqrt{n^i}}{\mu} h^T \nabla_x^2 \rho^i(\mu, x) h. \ \Box \tag{7.14}$$

We now discuss the self-concordance properties of $\eta(\mu, x)$.

Lemma 7.2 *The function $\eta(\mu, x)$ defined in (2.29) is strongly $\frac{\mu}{K}$-self-concordant on $\mathcal{P}_o$. In particular,*

$$\left|\nabla_x^3 \eta(\mu, x)[h, h, h]\right| \leq \frac{2K^{1/2}}{\mu^{1/2}} \left(h^T \nabla_x^2 \eta(\mu, x) h\right)^{3/2}.$$

Proof. From (7.11) we have

$$\begin{aligned}
& \left(h^T \nabla_x^2 \eta(\mu, x) h\right)^{1/2} \\
= & \left(\mu h^T X^{-2} h + \sum_{i=1}^{K} \pi^i h^T \nabla_x^2 \rho^i(\mu, x) h\right)^{1/2} \\
\geq & \left(\left(\mu \sum_{i=1}^{n} (h_i/x_i)^2\right)^{3/2} + \sum_{i=1}^{K} \pi^{i3/2} (h^T \nabla_x^2 \rho^i(\mu, x) h)^{3/2}\right)^{1/3} \quad (\text{using } \|\cdot\|_3 \leq \|\cdot\|_2), \\
\geq & \left(\mu^{1/2} \sum_{i=1}^{n} \mu (h_i/x_i)^3 + \frac{\mu^{1/2}}{2} \sum_{i=1}^{K} \pi^{i3/2} \nabla_x^3 \rho^i(\mu, x)[h, h, h]\right)^{1/3} \quad (\text{using 7.12}) \\
\geq & \left(\min_{i=1,\ldots,K} \{\pi^i\}^{1/2} \frac{\mu^{1/2}}{2} \left(\sum_{i=1}^{n} (h_i/x_i)^3 + \frac{\mu^{1/2}}{2} \sum_{i=1}^{K} \pi^i \nabla_x^3 \rho^i(\mu, x)[h, h, h]\right)\right)^{1/3} \\
= & \left(\frac{\mu^{1/2}}{2K^{1/2}} \nabla_x^3 \eta(\mu, x)[h, h, h]\right)^{1/3}
\end{aligned}$$

This proves that $\eta(\mu, x))$ is $\frac{\mu}{K}$-self-concordant. $\Box$

7.3 Parameters of the Self-Concordance Family $\{\eta(\mu, x), \mu > 0\}$

In this section we prove the values of the self-concordance family parameters for the discrete case.

Theorem 7.1 *The family of functions $\eta(\mu, x)$ is a strongly self-concordant family with parameters $\alpha(\mu) = \frac{\mu}{K}$, $\gamma(\mu) = \nu(\mu) = 1$, $\xi(\mu) = \frac{\sqrt{K(n+\hat{n})}}{\mu}$ and $\sigma(\mu) = \frac{\sqrt{n}}{\mu}$. In particular, for any $\mu > 0$,*

$x \in \mathcal{P}_o^1$ and $h \in \mathbb{R}^n$ we have

$$|\{\nabla\eta(\mu,x)^T h\}'| \leq \left[\frac{n+\hat{n}}{\mu}\nabla^2\eta(\mu,x)^T[h,h]\right]^{1/2}, \tag{7.15}$$

$$\{\nabla\eta(\mu,x)^T\}'[\nabla_x^2\eta(\mu,x)]^{-1}\{\nabla\eta(\mu,x)\}' \leq \mu^{-1}(n+\hat{n}), \tag{7.16}$$

$$|\{h^T\nabla^2\eta(\mu,x)h\}'| \leq \frac{2\sqrt{\hat{n}}}{\mu}\nabla^2\eta(\mu,x)[h,h]. \tag{7.17}$$

Proof. Condition SF1 in Definition A.2 is easy to verify. Lemma 7.2 shows Condition SF2. We now show SF3. Let

$$D := \left[\sqrt{\mu}X^{-1} \vdots \sqrt{\pi^1}T^{1^T}(R^1R^{1^T})^{-1/2} \vdots \cdots \vdots \sqrt{\pi^K}T^{K^T}(R^KR^{K^T})^{-1/2}\right], \text{ and}$$

$$\bar{D} := \left[-\frac{1}{\sqrt{\mu}}I \vdots \sqrt{\pi^1}T^{1^T}(R^1R^{1^T})^{-1/2} \vdots \cdots \vdots \sqrt{\pi^K}T^{K^T}(R^KR^{K^T})^{-1/2}\right]$$

Then, $DD^T = \nabla_x^2\eta(x,\mu)$. Differentiating (7.9) with respect to μ and applying (7.7) gives

$$\{\nabla\eta(x,\mu)\}' \quad = \quad -X^{-1}e + \frac{1}{\mu^{1/2}}\sum_{i=1}^K \pi^i T^{i^T}(R^iR^{i^T})^{-1}R^ie = D\bar{g}, \tag{7.18}$$

where $\bar{g} := \frac{1}{\mu^{1/2}}\left(-e, {\pi^1}^{1/2}(R^1R^{1^T})^{-1/2}R^1e, \ldots, {\pi^K}^{1/2}(R^KR^{K^T})^{-1/2}R^Ke\right) = \bar{D}^Te$,

$$h^T\{\nabla\eta(x,\mu)\}' = h^TD\bar{D}^Te \leq \|h^TD\|_2\|\bar{D}^Te\| = \|\bar{D}^Te\|\sqrt{h^T\nabla_x^2\eta(x)h}, \text{ and}$$

$$\|\bar{D}e\|^2 = \bar{g}^T\bar{g} = \mu^{-1}\left(n + \sum_{i=1}^K \pi^ie^TR^{i^T}(R^iR^{i^T})^{-1}R^ie\right) \leq \mu^{-1}(n+\hat{n}), \tag{7.19}$$

where the last inequality uses the fact that $R^{i^T}(R^iR^{i^T})^{-1}R^i$ is an orthogonal project matrix, $\hat{n} = n^i, i = 1, \ldots, K$ and $\sum_{i=1}^K \pi^i = 1$. Also, using (7.19)

$$\{\nabla\eta(\mu,x)^T\}'[\nabla_x^2\eta(\mu,x)]^{-1}\{\nabla\eta(\mu,x)\}' = \bar{g}^TD^T(DD^T)^{-1}D\bar{g} \leq \bar{g}^T\bar{g} \leq \mu^{-1}(n+\hat{n}).$$

Differentiating (7.11) with respect to μ and using Lemma 7.1, for any $h \in \mathbb{R}^n$ we have

$$\begin{aligned} |\{h^T\nabla^2\eta(\mu,x)h\}'| &= |h^TX^{-2}h + \sum_{i=1}^K(h^T\nabla_x^2\rho^i(x,\mu)h)'| \leq |h^TX^{-2}h| + \sum_{i=1}^K|(h^T\nabla_x^2\rho^i(x,\mu)h)'| \\ &\leq h^TX^{-2}h + \frac{2\sqrt{\hat{n}}}{\mu}\sum_{i=1}^K(h^T\nabla_x^2\rho^i(x,\mu)h) = \frac{2\sqrt{\hat{n}}}{\mu}h^T\nabla^2\eta(x,\mu)h \quad \square \end{aligned}$$

8. Convergence Analysis for the Weighted Barrier Algorithms

Theorems 8.1 and 8.2 below give complexity results for the short-step and long-step variants of Algorithm 1.

Theorem 8.1 *Consider the discrete and continuous versions of stochastic linear programs (2.1–2.3) and (2.29–2.31), respectively. 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.* $\delta(\mu_0, x^0) \le \beta$*, then the short-step algorithm reduces the barrier parameter* μ *at a linear rate and terminates within* $O(\sqrt{(n+m)K}\ln\mu_0/\epsilon)$ *iterations when solving (2.29–2.31). If the self-concordance assumption in Lemma 6.2 is satisfied, then the short-step algorithm terminates in* $O(\sqrt{(n+m)\tilde{K}}\ln\mu_0/\epsilon)$ *iterations (2.1–2.3) and (2.29–2.31).*

Theorem 8.2 *Consider the discrete and continuous versions of stochastic linear programs (2.1–2.3) and (2.29–2.31), respectively. 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.* $\delta(\mu_0, x^0) \le \beta$*, then the long-step algorithm reduces the barrier parameter* μ *at a linear rate and terminates within* $O((n+m)K\ln\mu_0/\epsilon)$ *iterations when solving (2.29–2.31). If the the self-concordance assumption in Lemma 6.2 is satisfied, then the long-step algorithm terminates in* $O(\sqrt{(n+m)\tilde{K}}\ln\mu_0/\epsilon)$ *iterations when solving (2.1–2.3) and (2.29–2.31).*

We now prove Theorems 8.1 and 8.2. The proof given here follows the steps of proofs in [11; 6], and it is given here for completeness. In this section we give a proof of Theorems 8.1–8.2 for the worst case analysis of the discrete case. The proof under the self-concordance assumption is identical, where K is replaced by $\tilde{K}$ in various inequality in the proofs of this section. The following proposition follows directly from the definition of self-concordance.

Proposition 8.1 *[9, Theorem 2.1.1] For any* $\mu > 0$*,* $x \in \mathcal{P}_o$ *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*

$$(1-\tau\delta)^2 h^T\nabla^2\eta(\mu,x)h \le h^T\nabla^2\eta(\mu,x+\tau\Delta x)h \le (1-\tau\delta)^{-2}h^T\nabla^2\eta(\mu,x)h. \tag{8.1}$$

In order to estimate the number of Newton steps needed for recentering the centrality measure $\delta(\mu,x)$, and the first stage objective $\eta(\mu,x)$ are used to measure progress for short and long step algorithms respectively. The following lemma describes the behavior of the Newton direction.

Lemma 8.1 *[9, Theorem 2.2.3] For any* $\mu > 0$ *and* $x \in \mathcal{P}_o$*, let* Δx *be the Newton direction calculated by (2.33) and let* $\delta := \delta(\mu,x) := \sqrt{\frac{K}{\mu}d_x^T\nabla^2\eta(\mu,x)d_x}$*. Then, the following relations hold:*

(i) If $\delta < 2-\sqrt{3}$*, then* $\delta(\mu, x+d_x) \le \left(\frac{\delta}{1-\delta}\right)^2 \le \frac{\delta}{2}$*.*
(ii) If $\delta \ge 2-\sqrt{3}$*, then* $\eta(\mu,x) - \eta(\mu, x+\overline{\theta}d_x) \ge \frac{\mu}{K}[\delta - \ln(1+\delta)]$*, where* $\overline{\theta} = (1+\delta)^{-1}$*.*

8.1 Complexity of the Short-Step Algorithm

We will show that in the short-step version of the algorithm a single Newton step is sufficient for recentering after updating the barrier parameter μ. To this end we will make use of [9, Theorem 3.1.1], which is restated for the present context in the next proposition.

Proposition 8.2 *Let* $\varphi_\kappa(\eta;\mu,\mu^+) := \left(\frac{1+\sqrt{m}}{2} + \frac{\sqrt{(n+\hat{n})K}}{\kappa}\right)\ln\gamma^{-1}$. *Suppose* $\delta(\mu,x) \le \kappa$, *and* $\mu^+ := \gamma\mu$ *satisfies* $\varphi_\kappa(\eta;\mu,\mu^+) \le 1 - \frac{\delta(\mu,x)}{\kappa}$. *Then,* $\delta(\mu^+,x) \le \kappa$.

Lemma 8.2 *Let* $\beta = (2-\sqrt{3})/2$, *and* $\mu^+ = \gamma\mu$ *where* $\gamma = 1 - \sigma/\sqrt{(n+\hat{n})K}$ *and* $\sigma \le 0.1$. *If* $\delta(\mu,x) \le \beta$ *then* $\delta(\mu^+,x) \le 2\beta$.

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

$$\varphi_\kappa(\eta;\mu,\mu^+) = \left(\frac{1+\sqrt{\hat{n}}}{2} + \frac{\sqrt{(n+\hat{n})K}}{\kappa}\right)\ln(1-\sigma/\sqrt{(n+\hat{n})K})^{-1} \le \frac{1}{2} \le\ 1 - \frac{\delta(\mu,x)}{\kappa}.$$

Using Proposition 8.2 gives $\delta(\mu^+,x) \le \kappa = 2\beta$. □

From Lemma 8.1 and Lemma 8.2 it is clear that we can reduce μ by the factor $\gamma = 1-\sigma/\sqrt{(n+\hat{n})K}$, $\sigma < 0.1$ at each iteration and a single Newton step is sufficient for recentering.
Theorem 8.1 follows.

8.2 Complexity of the Long-Step Algorithm

For the analysis of the long-step algorithm we use $\eta(\mu,x)$ 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 8.1. Our analysis follows the template in Zhao [11].

Assume that we have a point x^{k-1} sufficiently close to $x(\mu^{k-1})$. The barrier parameter from μ^{k-1} is reduced to $\mu^k = \gamma\mu^{k-1}$, where $\gamma \in (0,1)$. The long step algorithm generates a finite sequence of points $\tilde{x}^1,\ldots,\tilde{x}^N \in \mathcal{P}_o$ and finally sets $x^k = \tilde{x}^N$ when $\tilde{x}^N$ is sufficiently close to $x(\mu^k)$. We need to upper bound N, the number of damped Newton iterations needed for recentering. From Lemma 8.1(ii) at any $\tilde{x}^i \in \mathcal{P}_o$ a damped Newton step with step size $\overline{\theta} = (1+\delta)^{-1}$ decreases $\eta(\mu^k,\tilde{x}^i)$ at least by a certain amount which depends on the current value of δ and μ, when δ is 'large', i.e., when $\tilde{x}^i$ is not close to $x(\mu^k)$. Consequently, it is sufficient to prove a bound on

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

to bound N. This bound is established in Lemma 8.4 below. The next proposition and lemma give upper bounds on $\phi(\mu^{k-1})$ and $\phi'(\mu^{k-1})$, respectively.

Proposition 8.3 *[11, Lemma 7] For any* $\mu > 0$ *and* $x \in \mathcal{P}_o$, *let* $\tilde{d}_x := x - x(\mu)$ *and define* $\tilde{\delta} := \tilde{\delta}(\mu,x) := \sqrt{\frac{K}{\mu}\tilde{d}_x^T\nabla^2\eta(\mu,x)\tilde{d}_x}$. *If* $\tilde{\delta} < 1$, *then*

$$\phi(\mu,x) \le \mu\left[\frac{\tilde{\delta}}{1-\tilde{\delta}} + \ln(1-\tilde{\delta})\right]. \tag{8.2}$$

Lemma 8.3 *Let $\tilde{d}_x$ and $\tilde{\delta}$ be as defined in Proposition 8.3. For any $\mu > 0$ and $x \in \mathcal{P}_o$, if $\tilde{\delta} < 1$, then $|\phi'(\mu, x)| \le -\sqrt{n+\hat{n}}\ln(1-\tilde{\delta})$.*

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

$$\phi'(\mu, x) = \eta'(\mu, x) - \eta'(\mu, x(\mu)) - \nabla\eta(\mu, x(\mu))^T x'(\mu). \tag{8.3}$$

The optimality conditions (2.11) imply that $\nabla\eta(\mu, x(\mu))^T x'(\mu) = 0$.

Therefore,

$$\phi'(\mu, x) = \eta'(\mu, x) - \eta'(\mu, x(\mu)). \tag{8.4}$$

From (6.18 or 7.16) we have

$$\{\nabla\eta(\mu, x)^T\}'[\nabla^2\eta(\mu, x)]^{-1}\{\nabla\eta(\mu, x)\}' \le \mu^{-1}(n+\hat{n}). \tag{8.5}$$

Applying the mean-value theorem in (8.4) we get

$$\begin{aligned}
|\phi'(\mu, x)| &= \left|\int_0^1 \{\nabla\eta(\mu, x(\mu) + \tau\tilde{\Delta}x)^T\}'\tilde{\Delta}x \; d\tau\right| \\
&\le \int_0^1 \left[\tilde{\Delta}x^T\nabla^2\eta(\mu, x(\mu) + \tau\tilde{\Delta}x)\tilde{\Delta}x\right]^{1/2} \\
&\quad \left[\{\nabla\eta(\mu, x(\mu) + \tau\tilde{\Delta}x)^T\}'[\nabla^2\eta(\mu, x(\mu) + \tau\tilde{\Delta}x)]^{-1}\{\nabla\eta(\mu, x(\mu) + \tau\tilde{\Delta}x)^T\}'\right]^{1/2} d\tau \\
&\le \int_0^1 \frac{\sqrt{\tilde{d}_x^T\nabla^2\eta(\mu, x)\tilde{d}_x}}{1-\tilde{\delta}+\tau\tilde{\delta}}\sqrt{\frac{n+\hat{n}}{\mu}} \; d\tau \quad \text{(using 8.5)} \\
&\le \int_0^1 \frac{\sqrt{\mu}\tilde{\delta}}{1-\tilde{\delta}+\tau\tilde{\delta}}\sqrt{\frac{n+\hat{n}}{\mu}} \; d\tau \quad \text{(using Proposition8.1)} \\
&= -\sqrt{n+\hat{n}}\ln(1-\tilde{\delta}) \quad \Box
\end{aligned}$$

Lemma 8.4 *Let $\mu > 0$ and $x \in \mathcal{P}_o$ be such that $\tilde{\delta} < 1$, where $\tilde{\delta}$ is defined in Proposition 8.3. Let $\mu_+ = \gamma\mu$ with $\gamma \in (0, 1)$. Then, $\eta(\mu_+, x) - \eta(\mu_+, x(\mu_+)) \le O(n+\hat{n})\mu$.*

Proof. By differentiating (8.3) and using $\nabla\eta(\mu, x(\mu))^T x'(\mu) = 0$ we have

$$\phi''(\mu, x) = \eta''(\mu, x) - \eta''(\mu, x(\mu)) - \{\nabla\eta(\mu, x)^T\}'x'(\mu). \tag{8.6}$$

By taking the derivative of the optimality conditions (2.11) we have

$$\{\nabla\eta(\mu, x(\mu))\}' + Hx'(\mu) - A^T\lambda'(\mu) = 0, \quad Ax'(\mu) = 0, \tag{8.7}$$

where $H = \nabla^2\eta(\mu, x(\mu))$. Solving (8.7) gives

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

Now we have

$$\begin{aligned}
-\{\nabla\eta(\mu, x)^T\}'x'(\mu) &= \{\nabla\eta(\mu, x(\mu))^T\}'[H^{-1} - H^{-1}A^T(AH^{-1}A^T)^{-1}AH^{-1}]\{\nabla\eta(\mu, x(\mu))\}' \\
&\le \{\nabla\eta(\mu, x(\mu))^T\}'H^{-1}\{\nabla\eta(\mu, x(\mu))\}' \le \mu^{-1}(n+\hat{n}).
\end{aligned} \tag{8.8}$$

The last inequality above follows using (6.18 or 7.16). We now bound the first two terms in the right-hand-side of (8.6).

From (3.14–3.15), $0 \leq -\rho^{\xi\prime\prime}(\mu, x) \leq \hat{n}/\mu$. Hence, using Theorem B.2

$$\eta''(\mu, x) = E[\rho(\mu, x)]'' = \int_{\Xi} \rho^{\xi\prime\prime}(x, \mu) dF(\xi).$$

Hence, for any $x \in \mathcal{P}_o$

$$0 \leq -\eta''(\mu, x) \leq \frac{\hat{n}}{\mu}. \tag{8.9}$$

Now using the bounds given in (8.8), and (8.9) from (8.6) we get

$$\phi''(\mu, x) \leq \frac{n + 2\hat{n}}{\mu}. \tag{8.10}$$

Using Proposition 8.3, Lemma 8.3 and (8.10) we have

$$\begin{aligned}
&\phi(\mu_+, x) \\
&= \phi(\mu, x) + \phi'(\mu, x)(\mu_+ - \mu) + \int_{\mu}^{\mu_+} \int_{\mu}^{\tau} \phi''(\mu)\ d\mu\ d\tau \\
&\leq \mu \left[\frac{\tilde{\delta}}{1 - \tilde{\delta}} + \ln(1 - \tilde{\delta}) \right] - \sqrt{n + \hat{n}}(\mu - \mu_+) \ln(1 - \tilde{\delta}) + (n + 2\hat{n}) \int_{\mu}^{\mu_+} \int_{\mu}^{\tau} \mu^{-1}\ d\mu\ d\tau \\
&\leq \mu \left[\frac{\tilde{\delta}}{1 - \tilde{\delta}} + \ln(1 - \tilde{\delta}) \right] - \sqrt{n + \hat{n}}(\mu - \mu_+) \ln(1 - \tilde{\delta}) + (n + 2\hat{n})(\mu - \mu_+) \ln \gamma^{-1}\ .
\end{aligned} \tag{8.11}$$

Since γ and $\tilde{\delta}$ are absolute constants (8.11) implies that $\eta(\mu_+, x) - \eta(\mu_+, x(\mu_+)) \leq O(n + \hat{n})\mu$ and proves the lemma. □

Note that Proposition 8.3, Lemma 8.3 and Lemma 8.4 require $\tilde{\delta}$ be less than one. However, we cannot evaluate $\tilde{\delta}$ since we do not explicitly know $x(\mu)$. The lemma below shows that δ and $\tilde{\delta}$ are proportional to δ.

Proposition 8.4 *[11, Lemma 9] For any $\mu > 0$, $x \in \mathcal{P}_o$, let d_x be the Newton direction defined in (2.33) and $\tilde{d}_x := x - x(\mu)$. Let $\delta(\mu, x) := \sqrt{\frac{1}{\mu} d_x^T \nabla^2 \eta(\mu, x) d_x}$ $\quad \tilde{\delta}(\mu, x) := \sqrt{\frac{1}{\mu} \tilde{d}_x^T \nabla^2 \eta(\mu, x) \tilde{d}_x}$. If $\delta \leq 1/6$, then $\frac{2}{3}\delta \leq \tilde{\delta} \leq 2\delta$.* □

Lemma 8.1 implies that each line search should decrease the value of η by at least $\mu[\delta - \ln(1 + \delta)]$. Therefore, in view of Lemma 8.1 and Lemma 8.4 it is clear that after reducing μ by a factor $\gamma \in (0, 1)$, at most $O((n + \hat{n})K)$ damped Newton iterations will be needed for recentering. Also, the long-step variant algorithm updates barrier parameter μ at most $O(\ln \mu^0/\epsilon)$ times. Theorem 8.2 follows from Lemma 8.1 (ii), Lemma 8.4, and Proposition 8.4.

9. Concluding Remarks

In this paper we analyzed a prototype interior decomposition algorithm (WBDA) for two stage stochastic linear programs which uses a weighted barrier function. This algorithm was previously found to be more practical in the computational results of Mehrotra and Ozevin [7]. Our analysis here shows that the worst case first-stage iteration complexity of the weighted barrier decomposition algorithm in the finite scenario case is only slightly worse than the worst case iteration complexity of the barrier decomposition algorithm analyzed in [11; 8]. Interestingly, under a probabilistic assumption the weighted barrier decomposition algorithm has a worst case iteration complexity that depends only on a self-concordance parameter of a random matrix appearing in the algorithm, which is possibly independent of the number of scenarios (amount of discretization of a continuous random variable). The analysis of WBDA under the probabilistic self-concordance assumption is performed for problems under continuous support. This analysis also establishes several differentiability and analytic properties of the barrier recourse function. Our analysis assumes that we can compute the gradient and Hessian of the barrier recourse under continuous support exactly. In practice, this is not possible in general (except for some simple recourse problems) because numerical methods are required for integration, and because we can not compute exact solutions of second stage problems when we discretize a continuous problem. A more refined analysis that removes these assumptions remains a topic of future work. We point out that Mehrotra and Ozevin [7] numerically studied the issues of solving the second stage problems approximately, warm-starting algorithm for solving the second stage problem, practical first stage Newton step length, a practical choice of the proximity measure, selection of an initial barrier parameter and an initial solution. A proper resolution of these issues is important for developing practical implementations of WBDA.

Bibliography

[1] S. Cheng, *Differentiation under the integral sign with weak derivatives*, tech. report, http://www.gold-saucer.org/math/diff-int/diff-int.pdf, 2006.

[2] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analsyis and Minimization Algorithms I*, Springer-Verlag, 1996.

[3] A. King and R. Rockafellar, *Asymptotic theory for solutions in statistical estimation and stochastic programming*, Mathematics of Operations Research, 18 (1993), pp. 148–162.

[4] S. G. Krantz and H. R. Parks, *A Primer of Real Analytic Functions (2nd ed.)*, Springer, 2002.

[5] S. Mehrotra, *Minimizing functions of self-concordant random variables (in preparation)*.

[6] S. Mehrotra and M. G. Ozevin, *Decomposition based interior point methods for two-stage stochastic convex quadratic programs with recourse*, Operations Research (to appear).

[7] ——, *On the implementation of interior point decomposition algorithms for two-stage stochastic conic programs*, SIAM Journal on Optimization (under review).

[8] ——, *Decomposition-based interior point methods for two-stage stochastic semidefinite programming*, SIAM Journal on Optimization, 18(1) (2007), pp. 206–222.

[9] Y. Nesterov and A. Nemirovskii, *Interior Point Polynomial Algorithms In Convex Programming*, SIAM, 1994.

[10] A. Shapiro, *Asymptotic behavior of optimal solutions in stochastic programming*, Mathematics of Operations Research, 18 (1993), pp. 829–845.

[11] G. Zhao, *A log-barrier method with benders decomposition for solving two-stage stochastic linear programs*, Mathematical Programming, 90(3) (2001), pp. 507–536.

Appendix

A Self-Concordant Functions

Definition A.1 *[9, Definition 2.1.1 and 2.3.1] Let $\mathcal{C} \in \mathbb{R}^n$ be an open nonempty convex subset of $\mathcal{R}$, $\psi : \mathcal{C} \to \mathbb{R}$, and $\alpha > 0$. The function ψ is called α-self-concordant on $\mathcal{C}$ with the parameter α, if ψ is a convex function on $\mathcal{C}$, $\psi \in C^3$, and for all $x \in \mathcal{C}$ and $h \in \mathcal{E}$*

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

An α-self-concordant function ψ is called strongly α-self-concordant on $\mathcal{C}$ if $\psi(x^i)$ tends to infinity along every sequence $\{x^i \in \mathcal{C}\}$ converging to a boundary point of $\mathcal{C}$. Furthermore, an α-self-concordant function $\psi(x)$ has a complexity value ϑ, or we say that $\psi(x)$ is a ϑ-self-concordant-barrier, if it satisfies

$$\vartheta := \frac{sup_{x\in intQ}\nabla\psi(x)^T[\nabla^2\psi(x)]^{-1}\nabla\psi(x)}{\alpha} < \infty. \tag{A.1}$$

Definition A.2 *[9, Definition 3.1.1] A family of functions $\{\psi(\mu,x), \mu > 0\}$ is strongly self-concordant on nonempty convex domain $\mathcal{C} \in \mathbb{R}^n$ with continuous positive differentiable parameter functions $\alpha(\mu), \gamma(\mu), \nu(\mu), \xi(\mu)$, and $\sigma(\mu)$ if following properties hold:*

(SF1) Convexity and differentiability. $\psi(\mu,x)$ is convex in x, continuous in $(\mu,x) \in \mathcal{C}$ and is twice continuously differentiable in x.

(SF2) Self-concordance of members. For any $\mu > 0$, $\psi(\mu,x)$ is $\alpha(\mu)$-self-concordant on $\mathcal{C}$.

(SF3) Compatibility of neighbors. For every $(\mu,x) \in \mathbb{R}^+ \times \mathcal{C}$ and $h \in R^n$, $\nabla_x\psi(\mu,x)$, $\nabla_x^2\psi(x,\mu)$ are continuously differentiable in μ, and

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

$$|\{h^T\nabla_x^2\psi(\mu,x)h\}' - \{ln\,\gamma(\mu)\}'h^T\nabla_x^2\psi(\mu,x)h| \leq 2\sigma(\mu)h^T\nabla_x^2\psi(\mu,x)h. \tag{A.3}$$

B Analytic Functions

Definition B.1 *[4] A real valued function $\psi(x) : \mathcal{X} \to \mathbb{R}$ is called a real analytic function on its domain $\mathcal{X}$ if it is infinitely differentiable at all points in $\mathcal{X}$, its Taylor series expansion at any point $x \in \mathcal{X}$ is convergent, and the Taylor series expansion agrees (gives) the function value at all points $\hat{x}$ sufficiently close to x. A vector (matrix) valued function is called a real analytic vector (matrix) function if all its elements are real analytic functions.*

Proposition B.1 *[4] If the derivative of a real valued function $\psi(x)$ is bounded by θ^k, then $\psi(x)$ is a real analytic function in a neighborhood of $\bar{x}$.*

Theorem B.1 *[1] Let $\mathcal{X}$ be an open subset of $\mathbb{R}^n$, and Ξ be a measure space. Suppose that the function $\psi : \mathcal{X} \times \Xi \to \mathbb{R}$ satisfies the following conditions (1–3):*

1. *$\psi(x,\xi)$ is a Lebesgue-integrable function of ξ for each $x \in \mathcal{X}$.*

2. *For almost all* $\xi \in \Xi$, $\psi(x,\xi)$ *is continuous in* x.

3. *There is an integrable function* $\Theta : \Xi \to \mathbb{R}$ *such that* $|\psi(x,\xi)| \le \Theta(\xi)$ *for all* $x \in \mathcal{X}$.

Then, $\int_\Xi \psi(x,\xi)d\omega$ *is a continuous function of* $x \in \mathcal{X}$. $\square$

Theorem B.2 *[1] Let* $\mathcal{X}$ *be an open subset of* $\mathbb{R}^n$, *and* Ξ *be a measure space. Suppose* $\psi : \mathcal{X}\times\xi \to \mathbb{R}$ *satisfies the following conditions (1–3):*

1. $\psi(x,\xi)$ *is a Lebesgue-integrable function of* ξ *for each* $x \in \mathcal{X}$.

2. *For almost all* $\xi \in \Xi$, *the derivative* $\frac{\partial\psi(x,\xi)}{\partial x_i}$ *exists for all* $x \in \mathcal{X}$.

3. *There is an integrable function* $\Theta : \Xi \to \mathbb{R}$ *such that* $\left|\frac{\partial\psi(x,\xi)}{\partial x}\right| \le \Theta(\omega)$ *for all* $x \in \mathcal{X}$.

Then, for all $x \in \mathcal{X}$

$$\frac{\partial}{\partial x_i}\int_\Xi \psi(x,\omega)d\omega = \int_\Omega \frac{\partial\psi(x,\omega)}{\partial x_i}d\omega. \quad \square$$

C Generalized Holder and Projection Inequalities

We need the following Lemmas in the analysis of Section 6. Lemma C.1 is a generalization of Holder's inequality. Proposition C.2 is a generalization of the result that the 2-norm of a vector does not increase upon its orthogonal project onto a subspace.

Proposition C.1 *Let* $a^\xi := a^\xi(x,\mu)$, $b^\xi := b^\xi(x,\mu) : (\Xi, F(\xi), (\mathcal{X},\mu)) \to \mathbb{R}^n$ *be bounded vector functions with support* Ξ *and measure* $F(\xi)$. *Then,*

$$\int_\Xi |a^{\xi T} b^\xi| dF(\xi) \le \left(\int_\Xi \|a^\xi\|^2 dF(\xi)\right)^{1/2}\left(\int_\Xi \|b^\xi\|^2 dF(\xi)\right)^{1/2}. \tag{C.1}$$

Proof: We have,

$$\begin{aligned}
\frac{\int_\Xi |a^{\xi T} b^\xi| dF(\xi)}{\left(\int_\Xi \|a^\xi\|^2 dF(\xi)\right)^{1/2}\left(\int_\Xi \|b^\xi\|^2 dF(\xi)\right)^{1/2}} &\le \frac{\int_\Xi \left(\sum_{i=1}^n |a^\xi{}_i||b^\xi_i|\right) dF(\xi)}{\left(\int_\Xi \|a^\xi\|^2 dF(\xi)\right)^{1/2}\left(\int_\Xi \|b^\xi\|^2 dF(\xi)\right)^{1/2}} \\
&= \int_\Xi \sum_{i=1}^n \left(\frac{|a^\xi{}_i|}{\left(\int_\Xi \|a^\xi\|^2 dF(\xi)\right)^{1/2}} \frac{|b^\xi_i|}{\left(\int_\Xi \|b^\xi\|^2 dF(\xi)\right)^{1/2}}\right) dF(\xi) \\
&\le \frac{1}{2}\int_\Xi \sum_{i=1}^n \left(\frac{|a^\xi{}_i|^2}{\int_\Xi \|a^\xi\|^2 dF(\xi)} + \frac{|b^\xi_i|}{\int_\Xi \|b^\xi\|^2 dF(\xi)}\right) dF(\xi) \\
&= 1.
\end{aligned}$$

The second inequality above uses Young's inequality. $\square$

Proposition C.2 *Let* $B^\xi := B^\xi(x,\mu) : (\Xi, F(\xi), (\mathcal{X},\mu)) \to \mathbb{R}^{m\times n}$ *be a bounded real matrix function with support* Ξ *and measure* $F(\xi)$, *and assume that* $\bar{g}^\xi := \bar{g}^\xi(x,\mu) : (\Xi, F(\xi), (\mathcal{X},\mu)) \to \mathbb{R}^n$ *is a bounded real vector function with support* Ξ *and measure* $F(\xi)$. *Then,*

$$\left(\int_\Xi B^\xi \bar{g}^\xi dF(\xi)\right)^T \left[\int_\Xi B^\xi B^{\xi T} dF(\xi)\right]^{-1}\left(\int_\Xi B^\xi \bar{g}^\xi dF(\xi)\right) \le \int_\Xi \bar{g}^{\xi T}\bar{g}^\xi dF(\xi). \tag{C.2}$$

Proof: Let us consider a generalized least-squares minimization problem

$$\min_{u\in\mathbb{R}^m} E[\|B^{\xi^T}u - g^\xi\|^2] = \min_{u\in\mathbb{R}^m} \int_\Xi \left(B^{\xi^T}u - \bar{g}^\xi\right)^T \left(B^{\xi^T}u - \bar{g}^\xi\right) dF(\xi).$$

Then, our problem is to minimize

$$\int_\Xi \left(u^T B^{\xi^T} B^\xi u - 2u^T B^\xi \bar{g}^\xi + \bar{g}^{\xi T}\bar{g}^\xi\right) dF(\xi),$$

or equivalently,

$$\min_{u\in\mathbb{R}^m} u^T \left[\int_\Xi B^\xi B^{\xi^T} dF(\xi)\right] u - 2u^T \int_\Xi \left(B^\xi \bar{g}^\xi\right) dF(\xi) + \int_\Xi \bar{g}^{\xi T}\bar{g}^\xi dF(\xi). \tag{C.3}$$

From the optimality conditions, at the optimum solution u^* of (C.3) we have

$$\left[\int_\Xi B^\xi B^{\xi^T} dF(\xi)\right] u^* = \int_\Xi \left(B^\xi \bar{g}^\xi\right) dF(\xi). \tag{C.4}$$

Now substituting this value of u^* in (C.3) and observing that $\min E[\|B^{\xi^T}u - g^\xi\|^2] \geq 0$ gives us the desired result.