

A CLASS OF POLYNOMIAL VOLUMETRIC BARRIER DECOMPOSITION ALGORITHMS FOR STOCHASTIC SEMIDEFINITE PROGRAMMING

K. A. ARIYAWANSA AND YUNTAO ZHU

ABSTRACT. Ariyawansa and Zhu have recently proposed a new class of optimization problems termed stochastic semidefinite programs (SSDPs). SSDPs may be viewed as an extension of two-stage stochastic (linear) programs with recourse (SLPs). Zhao has derived a decomposition algorithm for SLPs based on a logarithmic barrier and proved its polynomial complexity. Mehrotra and Özevin have extended the work of Zhao to the case of SSDPs to derive a polynomial logarithmic barrier decomposition algorithm for SSDPs. An alternative to the logarithmic barrier is the volumetric barrier of Vaidya. There is no work based on the volumetric barrier analogous to that of Zhao for SLPs or to the work of Mehrotra and Özevin for SSDPs. The purpose of this paper is to derive a class of volumetric barrier decomposition algorithms for SSDPs, and to prove polynomial complexity of certain members of the class.

1. INTRODUCTION

Ariyawansa and Zhu [5] have recently proposed a new class of optimization problems termed stochastic semidefinite programs (SSDPs). SSDPs may be viewed as an extension of two-stage stochastic (linear) programs with recourse (SLPs) [7, 8, 9, 17, 20]. Alternatively, SSDPs may be viewed as an extension of (deterministic) semidefinite programs (SDPs) [1, 23, 24, 25]. See [5] for details on these relations and an application of SSDPs. Zhao [26] has derived a decomposition algorithm for SLPs based on a logarithmic barrier and proved its polynomial complexity. Mehrotra and Özevin [18] have extended the work of Zhao [26] to the case of SSDPs to derive a polynomial logarithmic barrier decomposition algorithm for SSDPs. The work of Mehrotra and Özevin [18] takes the viewpoint that SSDPs are extensions of SLPs and utilizes the work of Zhao [26].

An alternative to the logarithmic barrier is the volumetric barrier of Vaidya [21] (see also [2, 3, 4]). It has been observed [10] that certain cutting plane algorithms [16] for SLPs based on the volumetric barrier perform better in practice than those based on the logarithmic barrier. The authors know of no work based on volumetric

Received by the editor July 27, 2009 and, in revised form, May 14, 2010.

2010 *Mathematics Subject Classification*. Primary 90C15, 90C51, 49M27.

Key words and phrases. Stochastic linear programming, semidefinite programming, stochastic semidefinite programming, volumetric-barrier, self-concordance, decomposition.

The work of the first author was supported in part by the U.S. Army Research Office under Grant DAAD 19-00-1-0465 and under Award W911NF-08-1-0530.

The work of the second author was supported in part by the ASU West MGIA Grant 2007.

barriers analogous to that of Zhao [26] for SLPs or to the work of Mehrotra and Özevin [18] for SSDPs.

The purpose of this paper is to derive a class of decomposition algorithms for SSDPs based on a volumetric barrier, and to prove polynomial complexity of short step [3, 13] and long step [3, 13] members of the class.

While there is no work based on volumetric barriers for SLPs analogous to the work of Zhao [26], Anstreicher [4] has proved important results on volumetric barriers for SDPs. The present paper utilizes the work of Zhao [26] for SLPs, the work of Anstreicher [4] for SDPs, and the relationship of SSDPs to SLPs and SDPs described in [5], to derive volumetric barrier decomposition algorithms for SSDPs.

We begin by introducing our notation and then defining a SSDP in primal and dual standard forms. We let \mathbb{R}^+ denote the set of positive real numbers. All vectors in this paper are column vectors. We use superscript “T” to denote transposition. The i^{th} unit vector is denoted by e_i . Let $\mathbb{R}^{m \times n}$ and $\mathbb{R}^{n \vee n}$ denote the vector spaces of real $m \times n$ matrices and real symmetric $n \times n$ matrices respectively. For $U, V \in \mathbb{R}^{n \vee n}$ we write $U \succeq 0$ ($U \succ 0$) to mean that U is positive semidefinite (positive definite), and we use $U \succeq V$ or $V \preceq U$ to mean that $U - V \succeq 0$. For $U, V \in \mathbb{R}^{m \times n}$ we write $U \bullet V := \text{trace}(U^T V)$ to denote the Frobenius inner product between U and V . For $A \in \mathbb{R}^{m \times n}$, we use $\|A\|_2$ to denote the spectral norm of A . We use $\det(A)$ to denote the determinant of $A \in \mathbb{R}^{n \times n}$.

Following [5], we define a SSDP with recourse in primal standard form based on deterministic data $A_i \in \mathbb{R}^{n_1 \vee n_1}$ for $i = 1, 2, \dots, m_1$, $b \in \mathbb{R}^{m_1}$ and $C \in \mathbb{R}^{n_1 \vee n_1}$; and random data $T_j \in \mathbb{R}^{n_1 \vee n_1}$ and $W_j \in \mathbb{R}^{n_2 \vee n_2}$ for $j = 1, 2, \dots, m_2$, $d \in \mathbb{R}^{m_2}$, and $D \in \mathbb{R}^{n_2 \vee n_2}$ that depend on an underlying outcome ω in an event space Ω with a known probability function P . Given this data, a SSDP with recourse in primal standard form is

$$(1) \quad \begin{array}{ll} \text{minimize} & C \bullet X + \mathbb{E}[\mathcal{Q}(X, \omega)] \\ \text{subject to} & A_i \bullet X = b_i, \quad i = 1, 2, \dots, m_1, \\ & X \succeq 0, \end{array}$$

where $X \in \mathbb{R}^{n_1 \vee n_1}$ is the first-stage decision variable, $\mathcal{Q}(X, \omega)$ is the minimum of the problem

$$(2) \quad \begin{array}{ll} \text{minimize} & D(\omega) \bullet Y \\ \text{subject to} & T_j(\omega) \bullet X + W_j(\omega) \bullet Y = d_j(\omega), \quad j = 1, 2, \dots, m_2, \\ & Y \succeq 0, \end{array}$$

where $Y \in \mathbb{R}^{n_2 \vee n_2}$ is the second-stage variable, and

$$(3) \quad \mathbb{E}[\mathcal{Q}(X, \omega)] = \int_{\Omega} \mathcal{Q}(X, \omega) P(d\omega).$$

Also following [5], we define a SSDP with recourse in dual standard form based on deterministic data $A_i \in \mathbb{R}^{n_1 \vee n_1}$ for $i = 1, 2, \dots, m_1$, $b \in \mathbb{R}^{m_1}$ and $C \in \mathbb{R}^{n_1 \vee n_1}$; and random data $d \in \mathbb{R}^{m_2}$, $W_i \in \mathbb{R}^{n_2 \vee n_2}$ for $i = 1, 2, \dots, m_1$, $T_i \in \mathbb{R}^{n_2 \vee n_2}$ for $i = 1, 2, \dots, m_2$, and $D \in \mathbb{R}^{n_2 \vee n_2}$ that depend on an underlying outcome ω in an event space Ω with a known probability function P . Given this data, a SSDP with recourse in dual standard form is

$$(4) \quad \begin{array}{ll} \text{maximize} & b^T y + \mathbb{E}[Q(y, \omega)] \\ \text{subject to} & \sum_{i=1}^{m_1} y_i A_i \preceq C, \end{array}$$

where $y \in \mathbb{R}^{m_1}$ is the first-stage variable, $Q(y, \omega)$ is the maximum of the problem

$$(5) \quad \begin{aligned} & \text{maximize} && d(\omega)^\top x \\ & \text{subject to} && \sum_{i=1}^{m_1} y_i W_i(\omega) + \sum_{i=1}^{m_2} x_i T_i(\omega) \preceq D(\omega), \end{aligned}$$

where $x \in \mathbb{R}^{m_2}$ is the second-stage variable, and

$$(6) \quad \mathbb{E}[Q(y, \omega)] = \int_{\Omega} Q(y, \omega) P(d\omega).$$

See [5] for a justification for referring to problems (1)–(3) and (4)–(6) as primal and dual problems respectively.

We now examine the SSDP (4)–(6) when the event space Ω is finite. Let $\{(d^{(k)}, (W_i^{(k)} : i = 1, 2, \dots, m_1), (T_i^{(k)} : i = 1, 2, \dots, m_2), D^{(k)}) : k = 1, 2, \dots, K\}$ be the possible values of the random variables $(d(\omega), (W_i(\omega) : i = 1, 2, \dots, m_1), (T_i(\omega) : i = 1, 2, \dots, m_2), D(\omega))$ and let $p_k := P((d(\omega), (W_i(\omega) : i = 1, 2, \dots, m_1), (T_i(\omega) : i = 1, 2, \dots, m_2), D(\omega)) = (d^{(k)}, (W_i^{(k)} : i = 1, 2, \dots, m_1), (T_i^{(k)} : i = 1, 2, \dots, m_2), D^{(k)}))$ be the associated probability for $k = 1, 2, \dots, K$. Then problem (4)–(6) becomes

$$(7) \quad \begin{aligned} & \text{maximize} && b^\top y + \sum_{k=1}^K p_k Q^{(k)}(y) \\ & \text{subject to} && \sum_{i=1}^{m_1} y_i A_i \preceq C, \end{aligned}$$

where $y \in \mathbb{R}^{m_1}$ is the first-stage variable, $Q^{(k)}(y)$ is the maximum of the problem

$$(8) \quad \begin{aligned} & \text{maximize} && (d^{(k)})^\top x^{(k)} \\ & \text{subject to} && \sum_{i=1}^{m_1} y_i W_i^{(k)} + \sum_{i=1}^{m_2} x_i^{(k)} T_i^{(k)} \preceq D^{(k)}, \end{aligned}$$

where $x^{(k)} \in \mathbb{R}^{m_2}$ is the second-stage variable, for $k = 1, 2, \dots, K$.

We notice that the constraints in (7) and (8) are negative semidefinite while the common practice in the SDP literature is to use positive semidefinite constraints. So for convenience we redefine $d^{(k)} := p_k d^{(k)}$ for $k = 1, 2, \dots, K$, and rewrite problem (7)–(8) as follows:

$$(9) \quad \begin{aligned} & \text{minimize} && b^\top y + \sum_{k=1}^K Q^{(k)}(y) \\ & \text{subject to} && \sum_{i=1}^{m_1} y_i A_i - C \succeq 0, \end{aligned}$$

where for $k = 1, 2, \dots, K$, $Q^{(k)}(y)$ is the minimum of

$$(10) \quad \begin{aligned} & \text{minimize} \quad (d^{(k)})^\top x^{(k)} \\ & \text{subject to} \quad \sum_{i=1}^{m_1} y_i W_i^{(k)} + \sum_{i=1}^{m_2} x_i^{(k)} T_i^{(k)} - D^{(k)} \succeq 0. \end{aligned}$$

In the rest of this paper our attention will be on problem (9)–(10), and from now on when we use the term stochastic semidefinite program (SSDP) in this paper we mean problem (9)–(10).

The paper is organized as follows. In the next section we state some mathematical preliminaries. In §3 we introduce a volumetric barrier for the SSDP (9)–(10). In §4 we show that the set of barrier functions for positive values of the barrier parameter comprises a self-concordant family [19]. Based on this property a class of volumetric barrier decomposition algorithms is presented in §5. A convergence and complexity analysis of this class of algorithms is presented in §6. And the last section contains some concluding remarks.

2. PRELIMINARIES

In this section we introduce some further notation, and in order to make this paper self-contained, state some results from linear algebra and matrix calculus which we borrow from [4] (see also [12, 14, 15]).

Proposition 1. *Let $A, B \in \mathbb{R}^{n \times n}$. Then*

- (1) $\text{trace}(AB) = \text{trace}(BA)$;
- (2) *if A is symmetric, then $\text{trace}(AB) = \text{trace}(AB^\top)$;*
- (3) *if A and B are positive semidefinite, then $A \bullet B \geq 0$, and $A \bullet B = 0$ if and only if $AB = 0$;*
- (4) *if $A \succeq 0$ and $B \succeq C$, then $A \bullet B \succeq A \bullet C$.*

Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{k \times l}$, respectively. Then we define the Kronecker product $A \otimes B \in \mathbb{R}^{mk \times nl}$ of A and B as the matrix whose (i, j) block is $a_{ij}B$ for $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$. We also define

$$A \otimes_s B := \frac{1}{2}(A \otimes B + B \otimes A).$$

For a matrix $A \in \mathbb{R}^{m \times n}$, we use $\text{vec}(A) \in \mathbb{R}^{mn}$ to denote the vector formed by “stacking” the columns of A one atop another in the natural order. We have

Proposition 2. *Let $A, B, C, D \in \mathbb{R}^{n \times n}$. Then*

- (1) $(A \otimes B)(C \otimes D) = AC \otimes BD$;
- (2) $(A \otimes_s B)(C \otimes_s D) = \frac{1}{2}(AC \otimes_s BD + AD \otimes_s BC)$;
- (3) $(A \otimes B)^\top = A^\top \otimes B^\top$;
- (4) *if A and B are nonsingular, then $A \otimes B$ is nonsingular, and $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$;*
- (5) $\text{vec}(ABC) = (C^\top \otimes A)\text{vec}(B)$;
- (6) *if A and B are positive semidefinite, then $A \otimes B$ is positive semidefinite.*

We end this section by stating the following matrix calculus results.

Proposition 3. *Let $X \in \mathbb{R}^{n \times n}$ be nonsingular, and $\det(X)$ be positive. Then*

$$\frac{\partial}{\partial x_{ij}} \ln \det(X) = [X^{-1}]_{ji}$$

and

$$\frac{\partial}{\partial x_{ij}} X^{-1} = -X^{-1} e_i e_j^T X^{-1},$$

for $i, j = 1, 2, \dots, n$.

3. A VOLUMETRIC BARRIER FOR SSDPs

In this section we formulate a volumetric barrier for SSDPs and obtain expressions for the derivatives required in the rest of the paper.

3.1. Formulation. In order to define the volumetric barrier problem for the SSDP (9)–(10), we are going to make some assumptions. First we define

$$\begin{aligned} \mathcal{F}_1 &:= \{ y \in \mathbb{R}^{m_1} : S_1(y) := \sum_{i=1}^{m_1} y_i A_i - C \succ 0 \}; \\ \mathcal{F}^{(k)}(y) &:= \{ x^{(k)} \in \mathbb{R}^{m_2} : S_2^{(k)}(y, x^{(k)}) := \sum_{i=1}^{m_1} y_i W_i^{(k)} + \sum_{i=1}^{m_2} x_i^{(k)} T_i^{(k)} - D^{(k)} \succ 0 \}; \\ \mathcal{F}_2 &:= \{ y \in \mathbb{R}^{m_1} : \mathcal{F}^{(k)}(y) \neq \emptyset, k = 1, 2, \dots, K \}; \\ \mathcal{F}_0 &:= \mathcal{F}_1 \cap \mathcal{F}_2. \end{aligned}$$

Then we make

Assumption 1. The set \mathcal{F}_0 is nonempty.

The set \mathcal{F}_1 is nonempty under Assumption 1. The logarithmic barrier [19] for \mathcal{F}_1 is the function $f_1 : \mathcal{F}_1 \rightarrow \mathbb{R}$ defined by

$$f_1(y) := -\ln \det(S_1(y)), \quad \forall y \in \mathcal{F}_1,$$

and the volumetric barrier [19, 21] for \mathcal{F}_1 is the function $V_1 : \mathcal{F}_1 \rightarrow \mathbb{R}$ defined by

$$V_1(y) := \frac{1}{2} \ln \det(\nabla^2 f_1(y)), \quad \forall y \in \mathcal{F}_1.$$

Also under Assumption 1, \mathcal{F}_2 is nonempty and for $y \in \mathcal{F}_2$, $\mathcal{F}^{(k)}(y)$ is nonempty for $k = 1, 2, \dots, K$. The logarithmic barrier [19] for $\mathcal{F}^{(k)}(y)$ is the function $f_2^{(k)} : \mathcal{F}^{(k)}(y) \rightarrow \mathbb{R}$ defined by

$$f_2^{(k)}(y, x^{(k)}) := -\ln \det(S_2^{(k)}(y, x^{(k)})), \quad \forall x^{(k)} \in \mathcal{F}^{(k)}(y), y \in \mathcal{F}_2,$$

and the volumetric barrier [19, 21] for $\mathcal{F}^{(k)}(y)$ is the function $V_2^{(k)} : \mathcal{F}^{(k)}(y) \rightarrow \mathbb{R}$ defined by

$$V_2^{(k)}(y, x^{(k)}) := \frac{1}{2} \ln \det(\nabla_{x^{(k)} x^{(k)}}^2 f_2^{(k)}(y, x^{(k)})), \quad \forall x^{(k)} \in \mathcal{F}^{(k)}(y), y \in \mathcal{F}_2.$$

Next we make

Assumption 2. For each $y \in \mathcal{F}_0$ and for $k = 1, 2, \dots, K$, problem (10) has a nonempty isolated compact set of minimizers.

We now define the volumetric barrier problem for the SSDP (9)–(10) as

$$(11) \quad \text{minimize} \quad \eta(\mu, y) := b^\top y + \sum_{k=1}^K \rho_k(\mu, y) + \mu c_1 V_1(y),$$

where for $k = 1, 2, \dots, K$ and $y \in \mathcal{F}_0$, $\rho_k(\mu, y)$ is the minimum of

$$(12) \quad \text{minimize} \quad (d^{(k)})^\top x^{(k)} + \mu c_2 V_2^{(k)}(y, x^{(k)}).$$

Here $c_1 := 225\sqrt{n_1}$ and $c_2 := 450n_2^3$ are constants, and $\mu > 0$ is the barrier parameter.

We will now show that (12) has a unique minimizer for each $y \in \mathcal{F}_0$ and for $k = 1, 2, \dots, K$ by utilizing:

Theorem 1 (Fiacco and McCormick [11, Theorem 8]). *Consider the inequality constrained problem*

$$(13) \quad \begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && g_i(x) \geq 0, \quad i = 1, 2, \dots, m, \end{aligned}$$

where the functions $f, g_1, \dots, g_m : \mathbb{R}^n \rightarrow \mathbb{R}$ are continuous. Let I be a scalar-valued function of x with the following two properties: $I(x)$ is continuous in the region $R^0 := \{x : g_i(x) > 0, i = 1, 2, \dots, m\}$, which is assumed to be nonempty; if $\{x_k\}$ is any infinite sequence of points in R^0 converging to x_B such that $g_i(x_B) = 0$ for at least one i , then $\lim_{k \rightarrow \infty} I(x_k) = +\infty$. Let s be a scalar-valued function of the single variable r with the following two properties: if $r_1 > r_2 > 0$, then $s(r_1) > s(r_2) > 0$; if $\{r_k\}$ is an infinite sequence of points such that $\lim_{k \rightarrow \infty} r_k = 0$, then $\lim_{k \rightarrow \infty} s(r_k) = 0$. Let $U : R^0 \times \mathbb{R}^+ \rightarrow \mathbb{R}$ be defined by $U(x, r) := f(x) + s(r)I(x)$. If (13) has a nonempty, isolated compact set of local minimizers and $\{r_k\}$ is a strictly decreasing infinite sequence, then the unconstrained local minimizers of $U(\cdot, r_k)$ exist for r_k small.

Lemma 1. *If Assumptions 1 and 2 hold, then for each $y \in \mathcal{F}_0$ and $k = 1, 2, \dots, K$, the Problem (12) has a unique minimizer for μ small.*

Proof. For any given $y \in \mathcal{F}_0$, $V_2^{(k)}(y, x^{(k)})$ is defined on the nonempty set $\mathcal{F}^{(k)}(y)$. The positive definite matrix $S_2^{(k)}(y, x^{(k)})$ can be factored into the product of three matrices: a unit lower triangular matrix L , a positive definite diagonal matrix M , and the transpose of L , such that $S_2^{(k)}(y, x^{(k)}) = LML^\top$. Let m_j denote the j -th diagonal element of M viewed as a function of $x^{(k)} \in \mathcal{F}^{(k)}(y)$, for $j = 1, 2, \dots, n_2$. Then m_j is continuous for $j = 1, 2, \dots, n_2$ [22]. Then the constraint $S_2^{(k)}(y, x^{(k)}) \succ 0$ can be replaced by the constraints: $m_j(x^{(k)}) > 0, \quad j = 1, 2, \dots, n_2$ [22]. So (10) can be rewritten in the form of (13). Therefore, by Theorem 1, local minimizers of (12) exist for each $y \in \mathcal{F}_0$ and $k = 1, 2, \dots, K$ for μ small. The uniqueness of the minimizer follows from the fact that $V_2^{(k)}$ is strictly convex. \square

By Lemma 1, problem (11) is well defined, and its feasible set is \mathcal{F}_0 .

3.2. Expressions for partial derivatives of η with respect to y . In order to compute the derivatives of η we need the derivatives of $\rho_k, k = 1, 2, \dots, K$, which in turn require the derivatives of $V_2^{(k)}$ and $f_2^{(k)}$ for each $k = 1, 2, \dots, K$. Some of these computations are lengthy and it is convenient to drop the superscript (k) . We do so when it does not lead to confusion.

Also, we make

Assumption 3. Matrices $T_i^{(k)}$, $i = 1, 2, \dots, m_2$ are linear independent for $k = 1, 2, \dots, K$.

Let $\mathcal{T} \in \mathbb{R}^{n_2^2 \times m_2}$ be the matrix whose i^{th} column is $\text{vec}(T_i) \in \mathbb{R}^{n_2^2}$ for $i = 1, 2, \dots, m_2$. Then the Hessian matrix $H := \nabla_{xx}^2 f_2(y, x)$ can be represented in the form (See also [1].)

$$H := \nabla_{xx}^2 f_2(y, x) = \mathcal{T}^\top [S_2^{-1} \otimes S_2^{-1}] \mathcal{T}.$$

Note that by Proposition 2 and Assumption 3, H is positive definite. We have (See also [4].):

$$\begin{aligned} \frac{\partial V_2(y, x)}{\partial x_i} &= -(\mathcal{T}H^{-1}\mathcal{T}^\top) \bullet (S_2^{-1}T_i S_2^{-1} \otimes_s S_2^{-1}) \\ &= -P \bullet (S_2^{-1/2}T_i S_2^{-1/2} \otimes_s I), \end{aligned}$$

for $i = 1, 2, \dots, m_2$, and

$$(14) \quad \begin{aligned} \frac{\partial V_2(y, x)}{\partial y_i} &= -(\mathcal{T}H^{-1}\mathcal{T}^\top) \bullet (S_2^{-1}W_i S_2^{-1} \otimes_s S_2^{-1}) \\ &= -P \bullet (S_2^{-1/2}W_i S_2^{-1/2} \otimes_s I), \end{aligned}$$

for $i = 1, 2, \dots, m_1$, where

$$P = P(S_2) = (S_2^{-1/2} \otimes S_2^{-1/2}) \mathcal{T} (\mathcal{T}^\top (S_2^{-1} \otimes S_2^{-1}) \mathcal{T})^{-1} \mathcal{T}^\top (S_2^{-1/2} \otimes S_2^{-1/2})$$

is the orthogonal projection onto the range of $(S_2^{-1/2} \otimes S_2^{-1/2}) \mathcal{T}$;

$$\nabla_{xy}^2 V_2(y, x) = \frac{\partial^2}{\partial y \partial x} V_2(y, x) = 2Q^{xy} + R^{xy} - 2T^{xy},$$

where

$$\begin{aligned} Q_{i,j}^{xy} &= (\mathcal{T}H^{-1}\mathcal{T}^\top) \bullet (S_2^{-1}T_i S_2^{-1}W_j S_2^{-1} \otimes_s S_2^{-1}), \\ R_{i,j}^{xy} &= (\mathcal{T}H^{-1}\mathcal{T}^\top) \bullet (S_2^{-1}T_i S_2^{-1} \otimes_s S_2^{-1}W_j S_2^{-1}), \\ T_{i,j}^{xy} &= (\mathcal{T}H^{-1}\mathcal{T}^\top) \bullet (S_2^{-1}T_i S_2^{-1} \otimes_s S_2^{-1}) \mathcal{T}H^{-1}\mathcal{T}^\top (S_2^{-1}W_j S_2^{-1} \otimes_s S_2^{-1}); \end{aligned}$$

$$\nabla_{yx}^2 V_2(y, x) = \frac{\partial^2}{\partial x \partial y} V_2(y, x) = 2Q^{yx} + R^{yx} - 2T^{yx},$$

where

$$\begin{aligned} Q_{i,j}^{yx} &= (\mathcal{T}H^{-1}\mathcal{T}^\top) \bullet (S_2^{-1}W_i S_2^{-1}T_j S_2^{-1} \otimes_s S_2^{-1}), \\ R_{i,j}^{yx} &= (\mathcal{T}H^{-1}\mathcal{T}^\top) \bullet (S_2^{-1}W_i S_2^{-1} \otimes_s S_2^{-1}T_j S_2^{-1}), \\ T_{i,j}^{yx} &= (\mathcal{T}H^{-1}\mathcal{T}^\top) \bullet (S_2^{-1}W_i S_2^{-1} \otimes_s S_2^{-1}) \mathcal{T}H^{-1}\mathcal{T}^\top (S_2^{-1}T_j S_2^{-1} \otimes_s S_2^{-1}); \end{aligned}$$

$$\nabla_{xx}^2 V_2(y, x) = \frac{\partial^2}{\partial x \partial x} V_2(y, x) = 2Q^{xx} + R^{xx} - 2T^{xx},$$

where

$$\begin{aligned} Q_{i,j}^{xx} &= (\mathcal{T}H^{-1}\mathcal{T}^\top) \bullet (S_2^{-1}T_i S_2^{-1}T_j S_2^{-1} \otimes_s S_2^{-1}), \\ R_{i,j}^{xx} &= (\mathcal{T}H^{-1}\mathcal{T}^\top) \bullet (S_2^{-1}T_i S_2^{-1} \otimes_s S_2^{-1}T_j S_2^{-1}), \\ T_{i,j}^{xx} &= (\mathcal{T}H^{-1}\mathcal{T}^\top) \bullet (S_2^{-1}T_i S_2^{-1} \otimes_s S_2^{-1}) \mathcal{T}H^{-1}\mathcal{T}^\top (S_2^{-1}T_j S_2^{-1} \otimes_s S_2^{-1}); \end{aligned}$$

and

$$(15) \quad \nabla_{yy}^2 V_2(y, x) = \frac{\partial^2}{\partial y \partial y} V_2(y, x) = 2Q^{yy} + R^{yy} - 2T^{yy},$$

where

$$\begin{aligned} Q_{i,j}^{yy} &= (\mathcal{T}H^{-1}\mathcal{T}^\top) \bullet (S_2^{-1}W_iS_2^{-1}W_jS_2^{-1} \otimes_s S_2^{-1}), \\ R_{i,j}^{yy} &= (\mathcal{T}H^{-1}\mathcal{T}^\top) \bullet (S_2^{-1}W_iS_2^{-1} \otimes_s S_2^{-1}W_jS_2^{-1}), \\ T_{i,j}^{yy} &= (\mathcal{T}H^{-1}\mathcal{T}^\top) \bullet (S_2^{-1}W_iS_2^{-1} \otimes_s S_2^{-1})\mathcal{T}H^{-1}\mathcal{T}^\top(S_2^{-1}W_jS_2^{-1} \otimes_s S_2^{-1}). \end{aligned}$$

Now we define $\varphi_k : \mathbb{R}^+ \times \mathcal{F}_0 \times \mathcal{F}^{(k)}(y) \rightarrow \mathbb{R}$ by

$$\varphi_k(\mu, y, x) := d^\top x + \mu c_2 V_2(y, x).$$

Then by (12) we have

$$\rho_k(\mu, y) = \min_{x \in \mathcal{F}^{(k)}(y)} \varphi_k(\mu, y, x)$$

and

$$\rho_k(\mu, y) = \varphi_k(\mu, y, x)|_{x=\bar{x}} = \varphi_k(\mu, y, \bar{x}),$$

where \bar{x} is the minimizer of (12). We notice that \bar{x} is a function of y and is defined by

$$(16) \quad \frac{\partial}{\partial x} \varphi_k(\mu, y, x)|_{x=\bar{x}} = 0.$$

Now we are ready to calculate the first and second order derivatives of ρ_k with respect to y . We have

$$\begin{aligned} \nabla_y \rho_k(\mu, y) &= \left(\frac{\partial}{\partial y} \varphi_k(\mu, y, x) + \frac{\partial}{\partial x} \varphi_k(\mu, y, x) \cdot \frac{\partial x}{\partial y} \right) |_{x=\bar{x}} \\ &= \frac{\partial}{\partial y} \varphi_k(\mu, y, x)|_{x=\bar{x}} + \frac{\partial}{\partial x} \varphi_k(\mu, y, x)|_{x=\bar{x}} \cdot \frac{\partial x}{\partial y} |_{x=\bar{x}} \\ &= \frac{\partial}{\partial y} \varphi_k(\mu, y, x)|_{x=\bar{x}} \\ &= \mu c_2 \nabla_y V_2(y, x)|_{x=\bar{x}} \\ &= \mu c_2 \nabla_y V_2(y, \bar{x}) \end{aligned}$$

and

$$\begin{aligned} \nabla_{yy}^2 \rho_k(\mu, y) &= \nabla_y (\nabla_y \rho_k(\mu, y)) \\ &= \nabla_y (\mu c_2 \nabla_y V_2(y, \bar{x})) \\ &= \mu c_2 \nabla_{yy}^2 V_2(y, \bar{x}) + \mu c_2 \nabla_{\bar{x}} \nabla_y V_2(y, \bar{x}) \cdot \frac{\partial \bar{x}}{\partial y} \\ &= \mu c_2 \nabla_{yy}^2 V_2(y, \bar{x}) + \mu c_2 \nabla_y \nabla_{\bar{x}} V_2(y, \bar{x}) \cdot \frac{\partial \bar{x}}{\partial y} \\ &= \mu c_2 \nabla_{yy}^2 V_2(y, \bar{x}). \end{aligned}$$

Note that we use the fact that $\nabla_{\bar{x}} V_2(y, \bar{x}) = \nabla_x V_2(y, x)|_{x=\bar{x}} = -\frac{1}{\mu c_2} d$ by (16), which implies $\nabla_y \nabla_{\bar{x}} V_2(y, \bar{x}) = 0$. Similarly, we have $\nabla_{yy}^2 \nabla_{\bar{x}} V_2(y, \bar{x}) = 0$, which gives us

$$\begin{aligned} \nabla_{yyy}^3 \rho_k(\mu, y) &= \nabla_y (\nabla_{yy}^2 \rho_k(\mu, y)) \\ &= \nabla_y (\mu c_2 \nabla_{yy}^2 V_2(y, \bar{x})) \\ &= \mu c_2 \nabla_{yyy}^3 V_2(y, \bar{x}) + \mu c_2 \nabla_{yy}^2 \nabla_{\bar{x}} V_2(y, \bar{x}) \cdot \frac{\partial \bar{x}}{\partial y} \\ &= \mu c_2 \nabla_{yyy}^3 V_2(y, \bar{x}). \end{aligned}$$

In summary we have

$$\begin{aligned}
 \nabla_y \rho_k(\mu, y) &= \mu c_2 \nabla_y V_2^{(k)}(y, \bar{x}^{(k)}), \\
 \nabla_{yy}^2 \rho_k(\mu, y) &= \mu c_2 \nabla_{yy}^2 V_2^{(k)}(y, \bar{x}^{(k)}), \\
 \nabla_{yyy}^3 \rho_k(\mu, y) &= \mu c_2 \nabla_{yyy}^3 V_2^{(k)}(y, \bar{x}^{(k)}),
 \end{aligned}
 \tag{17}$$

and

$$\begin{aligned}
 \nabla_y \eta(\mu, y) &= b + \mu c_1 \nabla_y V_1(y) + \sum_{k=1}^K \mu c_2 \nabla_y V_2^{(k)}(y, \bar{x}^{(k)}), \\
 \nabla_{yy}^2 \eta(\mu, y) &= \mu c_1 \nabla_{yy}^2 V_1(y) + \sum_{k=1}^K \mu c_2 \nabla_{yy}^2 V_2^{(k)}(y, \bar{x}^{(k)}),
 \end{aligned}
 \tag{18}$$

where $\nabla_y V_2^{(k)}(y, \bar{x}^{(k)})$, $\nabla_{yy}^2 V_2^{(k)}(y, \bar{x}^{(k)})$, and $\nabla_{yyy}^3 V_2^{(k)}(y, \bar{x}^{(k)})$ are calculated in (14), (15) and (22), respectively.

4. CHARACTERISTICS OF η : A SELF-CONCORDANT FAMILY

4.1. Self-concordance of $\eta(\mu, \cdot)$.

Definition 1 (Nesterov and Nemirovskii [19, Definition 2.1.1]). Let G be an open nonempty convex subset of \mathbb{R}^n , and let F be a \mathcal{C}^3 , convex mapping from G to \mathbb{R} . Then F is called α -self-concordant on G with the parameter $\alpha > 0$ if for every $x \in G$ and $\xi \in \mathbb{R}^n$, the following inequality holds:

$$|D^3 F(x)[\xi, \xi, \xi]| \leq 2\alpha^{-1/2} (D^2 F(x)[\xi, \xi])^{3/2}.$$

An α -self-concordant function F on G is called strongly α -self-concordant if F tends to infinity for any sequence approaching a boundary point of G .

We note that in the definition above the set G is assumed to be open. However, relatively openness would be sufficient to apply the definition. See also [19, Item A, Page 57]. We now show that $\rho_k(\mu, \cdot)$ is μ -self-concordant on \mathcal{F}_0 , for $k = 1, 2, \dots, K$. It is clear that \mathcal{F}_0 is open.

Theorem 2. For any fixed $\mu > 0$, $\rho_k(\mu, \cdot)$ is μ -self-concordant on \mathcal{F}_0 , for $k = 1, 2, \dots, K$.

In order to prove Theorem 2 we need some intermediate results which we now obtain.

Proposition 4. Let (y, x) be such that $S_2(y, x) \succ 0$. Then we have

$$0 \preceq Q^{yy} \preceq \nabla_{yy}^2 V_2(y, x).$$

Proof. (See also [4].) Let $\xi \in \mathbb{R}^{m_1}$, $\xi \neq 0$. We have

$$\xi^\top Q^{yy} \xi = \sum_{i,j} Q_{ij}^{yy} \xi_i \xi_j = \mathcal{T} H^{-1} \mathcal{T} \bullet (S_2^{-1} B S_2^{-1} B S_2^{-1} \otimes_s S_2^{-1}) = P \bullet (\bar{B}^2 \otimes_s I),$$

where $B := B(\xi) := \sum_{i=1}^{m_1} \xi_i W_i$ and $\bar{B} := S_2^{-1/2} B S_2^{-1/2}$. Similarly, we have

$$\xi^\top R^{yy} \xi = \mathcal{T} H^{-1} \mathcal{T} \bullet (S_2^{-1} B S_2^{-1} \otimes_s S_2^{-1} B S_2^{-1}) = P \bullet (\bar{B} \otimes \bar{B})$$

and

$$\begin{aligned} \xi^\top T^{yy} \xi &= \mathcal{T}H^{-1}\mathcal{T} \bullet (S_2^{-1}BS_2^{-1} \otimes_s S_2^{-1})\mathcal{T}H^{-1}\mathcal{T}(S_2^{-1}BS_2^{-1} \otimes_s S_2^{-1}) \\ &= P \bullet (\bar{B} \otimes_s I)P(\bar{B} \otimes_s I). \end{aligned}$$

Since I , P and \bar{B}^2 are all positive semidefinite, we immediately have $Q^{yy} \succeq 0$ from Proposition 1 and Proposition 2. In addition, P is a projection implies that

$$(20) \quad (\bar{B} \otimes_s I)P(\bar{B} \otimes_s I) \preceq (\bar{B} \otimes_s I)(\bar{B} \otimes_s I) = \frac{1}{2}[(\bar{B}^2 \otimes_s I) + (\bar{B} \otimes \bar{B})].$$

We conclude that

$$P \bullet (\bar{B} \otimes_s I)P(\bar{B} \otimes_s I) \leq \frac{1}{2}P \bullet [(\bar{B}^2 \otimes_s I) + (\bar{B} \otimes \bar{B})],$$

which is exactly $\xi^\top T^{yy} \xi \leq \frac{1}{2} \xi^\top (Q^{yy} + R^{yy}) \xi$. Since ξ is arbitrary, we have shown that $T^{yy} \preceq \frac{1}{2}(Q^{yy} + R^{yy})$, which together with $Q^{yy} \succeq 0$ implies

$$0 \preceq Q^{yy} \preceq \nabla_{yy}^2 V_2(y, x). \quad \square$$

Proposition 5. For any $\xi \in \mathbb{R}^{m_1}$, let $\bar{B} := S_2^{-1/2}(\sum_{i=1}^{m_1} \xi_i W_i)S_2^{-1/2}$. Let (y, x) be such that $S_2(y, x) \succ 0$. Then

$$(21) \quad \begin{aligned} \xi^\top Q^{yy} \xi &\geq \frac{1}{2n_2^3} \|\bar{B}\|_2^2, \\ \text{i.e., } \|\bar{B}\|_2 &\leq \sqrt{2}n_2^{3/2}(\xi^\top Q^{yy} \xi)^{1/2}. \end{aligned}$$

Proof. Let $\lambda_1, \lambda_2, \dots, \lambda_{n_2}$ be the eigenvalues of \bar{B} with corresponding orthonormal eigenvectors v_1, v_2, \dots, v_{n_2} . Without loss of generality (scaling ξ as needed, and reordering indices), we may assume that $1 = |\lambda_1| \geq |\lambda_2| \geq \dots \geq \lambda_{n_2}$. Then $\bar{B}^2 \otimes_s I$ has a full set of orthonormal eigenvectors $v_i \otimes v_j$ with corresponding eigenvalues $(1/2)(\lambda_i^2 + \lambda_j^2)$, for $i, j = 1, 2, \dots, n_2$ (See also [15, Theorem 4.4.5]). We have

$$\begin{aligned} \xi^\top Q^{yy} \xi &= P \cdot \frac{1}{2} \sum_{i,j=1}^{n_2} (\lambda_i^2 + \lambda_j^2)(v_i \otimes v_j)(v_i \otimes v_j)^\top \\ &= \frac{1}{2} \sum_{i,j} (\lambda_i^2 + \lambda_j^2)(v_i \otimes v_j)^\top P(v_i \otimes v_j). \end{aligned}$$

Since P is a projection onto an m_2 -dimensional space, we have

$$P = \sum_{l=1}^{m_2} u_l u_l^\top,$$

where u_1, u_2, \dots, u_{m_2} are the orthonormal eigenvectors of P corresponding to the nonzero eigenvalues of P . Consider u_k for some k , we have

$$u_k = \sum_{i,j=1}^{n_2} c_{ij}(v_i \otimes v_j),$$

for some constants c_{ij} , for $i, j = 1, 2, \dots, n_2$, and

$$\begin{aligned} 1 &= \|u_k\|_2 = \left\| \sum_{i,j=1}^{n_2} c_{ij}(v_i \otimes v_j) \right\|_2 \\ &\leq \sum_{i,j=1}^{n_2} \|c_{ij}(v_i \otimes v_j)\|_2 \\ &= \sum_{i,j=1}^{n_2} |c_{ij}|. \end{aligned}$$

Thus there exist i_k, j_k such that

$$|c_{i_k j_k}| \geq \frac{1}{n_2}.$$

Hence,

$$\begin{aligned} \xi^\top Q^{yy} \xi &= \frac{1}{2} \sum_{i,j} (\lambda_i^2 + \lambda_j^2) (v_i \otimes v_j)^\top \left(\sum_{l=1}^{m_2} u_l u_l^\top \right) (v_i \otimes v_j) \\ &= \frac{1}{2} \sum_{i,j} (\lambda_i^2 + \lambda_j^2) \sum_{l=1}^{m_2} (v_i \otimes v_j)^\top u_l u_l^\top (v_i \otimes v_j) \\ &= \frac{1}{2} \sum_{i,j} (\lambda_i^2 + \lambda_j^2) \sum_{l=1}^{m_2} \|u_l^\top (v_i \otimes v_j)\|_2^2 \\ &\geq \frac{1}{2} \sum_{i,j} (\lambda_i^2 + \lambda_j^2) \|u_k^\top (v_{i_k} \otimes v_{j_k})\|_2^2 \\ &= \frac{1}{2} \sum_{i,j} (\lambda_i^2 + \lambda_j^2) |c_{i_k j_k}|^2 \\ &\geq \frac{1}{2} \sum_{i,j} (\lambda_i^2 + \lambda_j^2) \frac{1}{n_2^4} \\ &\geq \frac{1}{2n_2^4} \sum_j (\lambda_1^2 + \lambda_j^2) \\ &\geq \frac{1}{2n_2^4} \sum_j \lambda_1^2 \\ &= \frac{1}{2n_2^4} \sum_j \|\bar{B}\|_2^2 \\ &= \frac{1}{2n_2^3} \|\bar{B}\|_2^2. \end{aligned} \quad \square$$

Let (y, x) be such that $S_2(y, x) \succ 0$, and $\xi \in \mathbb{R}^{m_1}$. We immediately obtain

$$\begin{aligned} \frac{\partial}{\partial y_i} \xi^\top Q^{yy} \xi &= 2\mathcal{T}H^{-1}\mathcal{T}^\top (S_2^{-1}W_i S_2^{-1} \otimes_s S_2^{-1}) \mathcal{T}H^{-1}\mathcal{T}^\top \\ &\quad \cdot (S_2^{-1}BS_2^{-1}BS_2^{-1} \otimes_s S_2^{-1}) \\ &\quad + \mathcal{T}H^{-1}\mathcal{T} \cdot \frac{\partial}{\partial y_i} (S_2^{-1}BS_2^{-1}BS_2^{-1} \otimes_s S_2^{-1}), \end{aligned}$$

where $B := B(\xi) := \sum_{i=1}^{m_1} \xi_i W_i$, and

$$\begin{aligned} & \frac{\partial}{\partial y_i} (S_2^{-1} B S_2^{-1} B S_2^{-1} \otimes_s S_2^{-1}) \\ &= -(S_2^{-1} W_i S_2^{-1} B S_2^{-1} B S_2^{-1} + S_2^{-1} B S_2^{-1} W_i S_2^{-1} B S_2^{-1} + S_2^{-1} B S_2^{-1} B S_2^{-1} W_i S_2^{-1}) \\ & \otimes_s S_2^{-1} - S_2^{-1} B S_2^{-1} B S_2^{-1} \otimes_s S_2^{-1} W_i S_2^{-1}. \end{aligned}$$

We conclude that the first directional derivative of $\xi^T Q^{yy} \xi$ with respect to y , in the direction ξ , is given by

$$\begin{aligned} \nabla_y \xi^T Q^{yy} \xi [\xi] &= \sum_{i=1}^{m_1} \xi_i \frac{\partial}{\partial y_i} \xi^T Q^{yy} \xi \\ &= 2P \cdot (\bar{B} \otimes_s I) P (\bar{B}^2 \otimes_s I) - 3P \cdot (\bar{B}^3 \otimes_s I) - P \cdot (\bar{B}^2 \otimes_s \bar{B}), \end{aligned}$$

where $\bar{B} := S^{-1/2} B S^{-1/2}$, and P is defined as before. Similarly, we obtain

$$\begin{aligned} \nabla_y \xi^T R^{yy} \xi [\xi] &= 2P \cdot (\bar{B} \otimes_s I) P (\bar{B} \otimes_s \bar{B}) - 4P \cdot (\bar{B}^2 \otimes_s \bar{B}), \\ \nabla_y \xi^T T^{yy} \xi [\xi] &= 4P \cdot (\bar{B} \otimes_s I) P (\bar{B} \otimes_s I) P (\bar{B} \otimes_s I) - 4P \cdot (\bar{B} \otimes_s I) P (\bar{B}^2 \otimes_s I) \\ & \quad - 2P \cdot (\bar{B} \otimes_s I) P (\bar{B} \otimes_s \bar{B}). \end{aligned}$$

Combining the previous results, we obtain the third-order directional derivative of $V_2(y, x)$ with respect to y as:

$$\begin{aligned} \nabla_{yyy}^3 V_2(y, x) [\xi, \xi, \xi] &= 12P \cdot (\bar{B} \otimes_s I) P (\bar{B}^2 \otimes_s I) \\ & \quad - 6P \cdot (\bar{B}^3 \otimes_s I) - 6P \cdot (\bar{B}^2 \otimes_s \bar{B}) \\ (22) \quad & \quad + 6P \cdot (\bar{B} \otimes_s I) P (\bar{B} \otimes_s \bar{B}) \\ & \quad - 8P \cdot (\bar{B} \otimes_s I) P (\bar{B} \otimes_s I) P (\bar{B} \otimes_s I). \end{aligned}$$

In the proof of Theorem 2 we need to bound $\nabla_{yyy}^3 V_2(y, x) [\xi, \xi, \xi]$. We now obtain such a bound.

Proposition 6. *For any $\xi \in \mathbb{R}^{m_1}$, let $\bar{B} := S_2^{-1/2} (\sum_{i=1}^{m_1} \xi_i W_i) S_2^{-1/2}$. Let (y, x) be such that $S_2(y, x) \succ 0$. Then*

$$(23) \quad |\nabla_{yyy}^3 V_2(y, x) [\xi, \xi, \xi]| \leq 30 \|\bar{B}\|_2 \xi^T Q^{yy} \xi.$$

Proof. (See also [4].) Using the fact that

$$(\bar{B}^2 \otimes_s I) (\bar{B} \otimes_s I) = \frac{1}{2} [(\bar{B}^3 \otimes_s I) + (\bar{B}^2 \otimes_s \bar{B})],$$

we can rewrite (22) as

$$\begin{aligned} \nabla_{yyy}^3 V_2(y, x) [\xi, \xi, \xi] &= P (\bar{B} \otimes_s I) P \cdot [12(\bar{B}^2 \otimes_s I) \\ (24) \quad & \quad + 6(\bar{B} \otimes_s \bar{B}) - 8(\bar{B} \otimes_s I) P (\bar{B} \otimes_s I)] \\ & \quad - 12P \cdot (\bar{B}^2 \otimes_s I) (\bar{B} \otimes_s I). \end{aligned}$$

From (20) we have

$$12(\bar{B}^2 \otimes_s I) + 6(\bar{B} \otimes_s \bar{B}) - 8(\bar{B} \otimes_s I) P (\bar{B} \otimes_s I) \succeq 8(\bar{B}^2 \otimes_s I) + 2(\bar{B} \otimes_s \bar{B}).$$

Using the facts that $(\bar{B}^2 \otimes_s I) \succeq (\bar{B} \otimes_s \bar{B})$ and $(\bar{B} \otimes_s \bar{B}) \succeq (-\bar{B}^2 \otimes_s I)$, we obtain

$$\begin{aligned} (25) \quad 6(\bar{B}^2 \otimes_s I) &\preceq 12(\bar{B}^2 \otimes_s I) + 6(\bar{B} \otimes_s \bar{B}) - 8(\bar{B} \otimes_s I) P (\bar{B} \otimes_s I) \\ &\preceq 18(\bar{B}^2 \otimes_s I). \end{aligned}$$

Let $\lambda_1, \lambda_2, \dots, \lambda_{n_2}$ be the eigenvalues of \bar{B} . Then for $i, j = 1, 2, \dots, n_2$, the eigenvalues of $(\bar{B} \otimes_s I)$ are of the form $(1/2)(\lambda_i + \lambda_j)$ (See also [15, Theorem 4.4.5].) We have

$$-\|\bar{B}\|_2 I \preceq (\bar{B} \otimes_s I) \preceq \|\bar{B}\|_2 I$$

and

$$(26) \quad -\|\bar{B}\|_2 P \preceq P(\bar{B} \otimes_s I)P \preceq \|\bar{B}\|_2 P.$$

Using (25), (26), and the fact that $(\bar{B}^2 \otimes_s I) \succeq 0$, we obtain

$$(27) \quad \begin{aligned} &|P(\bar{B} \otimes_s I) \cdot [12(\bar{B}^2 \otimes_s I) + 6(\bar{B} \otimes \bar{B}) - 8(\bar{B} \otimes_s I)P(\bar{B} \otimes_s I)]| \\ &\leq 18\|\bar{B}\|_2 P \cdot (\bar{B}^2 \otimes_s I). \end{aligned}$$

In addition, the fact that $(\bar{B}^2 \otimes_s I)$ and $(\bar{B} \otimes_s I)$ have the same eigenvectors implies that

$$-\|\bar{B}\|_2(\bar{B}^2 \otimes_s I) \preceq (\bar{B}^2 \otimes_s I)(\bar{B} \otimes_s I) \preceq \|\bar{B}\|_2(\bar{B}^2 \otimes_s I).$$

Therefore we have

$$(28) \quad |P \cdot (\bar{B}^2 \otimes_s I)(\bar{B} \otimes_s I)| \leq \|\bar{B}\|_2 P \cdot (\bar{B}^2 \otimes_s I).$$

The conclusion follows from (24), (27) and (28). □

We can now state the proof of Theorem 2.

Proof of Theorem 2. Combining the results of (19), (21) and (23), we obtain

$$|\nabla_{yyy}^3 V_2(y, \bar{x})[\xi, \xi, \xi]| \leq 30\sqrt{2}n_2^{3/2}(\xi^\top \nabla_{yy}^2 V_2(y, \bar{x})\xi)^{3/2},$$

which combined with (17) gives us

$$\begin{aligned} |\nabla_{yyy}^3 \rho_k(y)[\xi, \xi, \xi]| &\leq 30\sqrt{2}\mu c_2 n_2^{3/2}(\nabla_{yy}^2 V_2(y, \bar{x})[\xi, \xi])^{3/2} \\ &= 2\mu^{-1/2}(c_2 \mu \nabla_{yy}^2 V_2(y, x)[\xi, \xi])^{3/2} \\ &= 2\mu^{-1/2}(\nabla_{yy}^2 \rho_k(y)[\xi, \xi])^{3/2}. \quad \square \end{aligned}$$

Corollary 1. *For any fixed $\mu > 0$, $\eta(\mu, \cdot)$ is a μ -self-concordant function on \mathcal{F}_0 .*

Proof. It is easy to verify that $\mu c_1 V_1$ is μ -self-concordant on \mathcal{F}_1 . The corollary follows from [19, Proposition 2.1.1]. □

4.2. Parameters of the self-concordant family $\eta(\mu, \cdot)$.

Definition 2 (Nesterov and Nemirovskii [19, Definition 3.1.1]). Let \mathbb{R}^+ be the set of all positive real numbers. Let G be an open nonempty convex subset of \mathbb{R}^n . Let $\mu \in \mathbb{R}^+$ and let $F_\mu : \mathbb{R}^+ \times G \rightarrow \mathbb{R}$ be a family of functions indexed by μ . Let $\alpha_1(\mu)$, $\alpha_2(\mu)$, $\alpha_3(\mu)$, $\alpha_4(\mu)$ and $\alpha_5(\mu) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be continuously differentiable functions on μ . Then the family of functions $\{F_\mu\}_{\mu \in \mathbb{R}^+}$ is called strongly self-concordant with the parameters $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$, if the following conditions hold:

- (i) F_μ is continuous on $\mathbb{R}^+ \times G$, and for fixed $\mu \in \mathbb{R}^+$, F_μ is convex on G . F_μ has three partial derivatives on G , which are continuous on $\mathbb{R}^+ \times G$ and continuously differentiable with respect to μ on \mathbb{R}^+ .
- (ii) For any $\mu \in \mathbb{R}^+$, the function F_μ is strongly $\alpha_1(\mu)$ -self-concordant.
- (iii) For any $(\mu, x) \in \mathbb{R}^+ \times G$ and any $\xi \in \mathbb{R}^n$,

$$\begin{aligned} &|\{\nabla_x F_\mu(\mu, x)[\xi]\}' - \{\ln \alpha_3(\mu)\}' \nabla_x F_\mu(\mu, x)[\xi]| \leq \alpha_4(\mu) \alpha_1(\mu)^{\frac{1}{2}} (\nabla_{xx}^2 F_\mu(\mu, x)[\xi, \xi])^{\frac{1}{2}}, \\ &|\{\nabla_{xx}^2 F_\mu(\mu, x)[\xi, \xi]\}' - \{\ln \alpha_2(\mu)\}' \nabla_{xx}^2 F_\mu(\mu, x)[\xi, \xi]| \leq 2\alpha_5(\mu) \nabla_{xx}^2 F_\mu(\mu, x)[\xi, \xi]. \end{aligned}$$

Theorem 3. *The parametric function $\eta(\mu, \cdot)$ is a strongly self-concordant family with the following parameters:*

$$\alpha_1(\mu) = \mu, \alpha_2(\mu) = \alpha_3(\mu) = 1, \alpha_4(\mu) = \mu^{-1}[(m_1c_1+m_2c_2)(1+K)]^{1/2}, \alpha_5(\mu) = \frac{1}{\mu}.$$

By Corollary 1, in order to prove Theorem 3, we only need to show that the two inequalities in Definition 2 (iii) are satisfied by $\eta(\mu, \cdot)$. We first show the validity of the second inequality in Definition 2 (iii).

Lemma 2. *For any $\mu > 0$ and $y \in \mathcal{F}_0$, the following inequality holds:*

$$|\nabla_{yy}^2 \eta'(\mu, y)[\xi, \xi]| \leq \frac{1}{\mu} \nabla_{yy}^2 \eta(\mu, y)[\xi, \xi], \quad \forall \xi \in \mathbb{R}^{m_1}.$$

Proof. Differentiating (18) with respect to μ , we obtain

$$\begin{aligned} \nabla_{yy}^2 \eta'(\mu, y) &= c_1 \nabla_y^2 V_1(y) + \sum_{k=1}^K c_2 (\nabla_{yy}^2 V_2^{(k)}(y, \bar{x}^{(k)}) + \mu \nabla_{yy\bar{x}}^3 V_2^{(k)}(y, \bar{x}^{(k)}) \cdot (\bar{x}^{(k)})') \\ &= c_1 \nabla_{yy}^2 V_1(y) + \sum_{k=1}^K c_2 \nabla_{yy}^2 V_2^{(k)}(y, \bar{x}^{(k)}), \end{aligned}$$

which is exactly $\frac{1}{\mu} \nabla_{yy}^2 \eta(\mu, y)$. The conclusion follows since $\nabla_{yy}^2 \eta(\mu, y)$ is positive semidefinite. □

For fixed (y, \bar{x}) with $S_2(y, \bar{x}) \succ 0$, let $\bar{T}_i = S_2^{-1/2} T_i S_2^{-1/2}, i = 1, 2, \dots, m_2$ and $\bar{W}_j = S_2^{-1/2} W_j S_2^{-1/2}, j = 1, 2, \dots, m_1$. We apply a Gram-Schmidt procedure to $\{\bar{T}_i\}$ to obtain $\{U_i\}$ with $\|U_i\| = 1$ for all i and $U_i \cdot U_j = 0, i \neq j$. Then the linear span of $\{U_i, i = 1, 2, \dots, m_2\}$ is equal to the span of $\{\bar{T}_i, i = 1, 2, \dots, m_2\}$. Let U be the $n_2^2 \times m_2$ matrix whose i^{th} column is $\text{vec}(U_i)$ and let $\Sigma = \sum_{k=1}^{m_2} U_k^2$. Then $P = P(S_2) = UU^T$. We have

$$\begin{aligned} \frac{\partial V_2(y, \bar{x})}{\partial y_i} &= -P \cdot (\bar{W}_i \otimes_s I) = -UU^T \cdot (\bar{W}_i \otimes_s I) \\ &= \text{trace}(U^T (\bar{W}_i \otimes_s I) U) \\ &= -\frac{1}{2} \sum_{k=1}^{m_2} \text{vec}(U_k)^T [(\bar{W}_i \otimes I) + (I \otimes \bar{W}_i)] \text{vec}(U_k) \\ (29) \quad &= -\frac{1}{2} \sum_{k=1}^{m_2} \text{vec}(U_k)^T \text{vec}(U_k \bar{W}_i + \bar{W}_i U_k) \\ &= -\frac{1}{2} \sum_{k=1}^{m_2} U_k \cdot (U_k \bar{W}_i + \bar{W}_i U_k) = -\bar{W}_i \cdot \left(\sum_{k=1}^{m_2} U_k^2 \right) = -\bar{W}_i \cdot \Sigma. \end{aligned}$$

Similarly, we have

$$\begin{aligned}
 Q_{i,j}^{yy} &= P \cdot (\bar{W}_i \bar{W}_j \otimes_s I) = \text{trace}(U^\top (\bar{W}_i \bar{W}_j \otimes_s I) U) \\
 &= \sum_{k=1}^{m_2} \text{vec}(U_k)^\top (\bar{W}_i \bar{W}_j \otimes_s I) \text{vec}(U_k) \\
 &= \frac{1}{2} \sum_{k=1}^{m_2} \text{vec}(U_k)^\top \text{vec}(U_k \bar{W}_i \bar{W}_j + \bar{W}_i \bar{W}_j U_k) \\
 (30) \quad &= \frac{1}{2} \sum_{k=1}^{m_2} U_k \cdot (U_k \bar{W}_i \bar{W}_j + \bar{W}_i \bar{W}_j U_k) \\
 &= \frac{1}{2} \sum_{k=1}^{m_2} \text{trace}(U_k (U_k \bar{W}_i \bar{W}_i + \bar{W}_j \bar{W}_j U_k)) \\
 &= \sum_{k=1}^{m_2} \text{trace}(\bar{W}_i U_k^2 \bar{W}_j) = \text{trace}(\bar{W}_i \Sigma \bar{W}_j).
 \end{aligned}$$

Proposition 7. *Let (y, \bar{x}) be such that $S_2(y, \bar{x}) \succ 0$. Then*

$$(31) \quad \nabla_y V_2(y, \bar{x})^\top (\nabla_{yy}^2 V_2(y, \bar{x}))^{-1} \nabla_y V_2(y, \bar{x}) \leq m_2.$$

Proof. (See also [4].) Let \mathfrak{W} be the $n_2^2 \times m_1$ matrix whose i^{th} column is $\text{vec}(\bar{W}_i)$. Then from (30) we have

$$Q_{i,j}^{yy} = \text{trace}(\bar{W}_i \Sigma \bar{W}_j) = \text{vec}(\bar{W}_i)^\top \text{vec}(\Sigma \bar{W}_j) = \text{vec}(\bar{W}_i)^\top (I \otimes \Sigma) \text{vec}(\bar{W}_j).$$

We can then write $Q^{yy} = \mathfrak{W}^\top (I \otimes \Sigma) \mathfrak{W}$. Also, it follows from (29) that $\nabla_y V_2(y, \bar{x})^\top = -\mathfrak{W}^\top \text{vec}(\Sigma)$. Hence,

$$\begin{aligned}
 \nabla_y V_2(y, \bar{x})^\top [Q^{yy}]^{-1} \nabla_y V_2(y, \bar{x}) &= \text{vec}(\Sigma)^\top \mathfrak{W} (\mathfrak{W}^\top (I \otimes \Sigma) \mathfrak{W})^{-1} \mathfrak{W}^\top \text{vec}(\Sigma) \\
 &= \text{vec}(\Sigma^{1/2})^\top (I \otimes \Sigma^{1/2}) \mathfrak{W} (\mathfrak{W}^\top (I \otimes \Sigma) \mathfrak{W})^{-1} \mathfrak{W}^\top \\
 &\quad (I \otimes \Sigma^{1/2}) \text{vec}(\Sigma^{1/2}) \\
 &\leq \text{vec}(\Sigma^{1/2})^\top \text{vec}(\Sigma^{1/2}) = \text{trace}(\Sigma) = m_2,
 \end{aligned}$$

since $\Sigma = \sum_{k=1}^{m_2} U_k^2$, and $\text{trace}(U_k^2) = U_k \cdot U_k = 1$ for each k . In addition, $Q^{yy} \preceq \nabla_{yy}^2 V_2(y, \bar{x})$ implies $(\nabla_{yy}^2 V_2(y, \bar{x}))^{-1} \preceq (Q^{yy})^{-1}$. So the result follows. \square

It can be easily verified that (31) is equivalent to the following inequality:

$$(32) \quad |\nabla_y V_2(y, \bar{x})[\xi]| \leq \sqrt{m_2 \nabla_{yy}^2 V_2(y, \bar{x})[\xi, \xi]}, \quad \forall \xi \in \mathbb{R}^{m_1}.$$

Now we show the validity of the first inequality in Definition 2 (iii).

Lemma 3. *For any $\mu > 0$ and $y \in \mathcal{F}_0$, we have*

$$|\nabla_y \eta'(\mu, y)^\top[\xi]| \leq \sqrt{\frac{(m_1 c_1 + m_2 c_2)(1 + K)}{\mu}} \nabla_{yy}^2 \eta(\mu, y)[\xi, \xi], \quad \forall \xi \in \mathbb{R}^{m_1}.$$

Proof. We have

$$\begin{aligned}
 \nabla_y \eta'(\mu, y) &= c_1 \nabla_y V_1(y) + \sum_{k=1}^K c_2 (\nabla_y V_2^{(k)}(y, \bar{x}^{(k)}) + \mu \nabla_{y\bar{x}}^2 V_2^{(k)}(y, \bar{x}^{(k)}) \cdot (\bar{x}^{(k)})') \\
 &= c_1 \nabla_y V_1(y) + \sum_{k=1}^K c_2 \nabla_y V_2^{(k)}(y, \bar{x}^{(k)}).
 \end{aligned}$$

Anstreicher [4, Theorem 4.4] has shown that

$$\nabla_y V_1(y)^\top (\nabla_{yy}^2 V_1(y))^{-1} \nabla_y V_1(y) \leq m_1,$$

which is equivalent to

$$(33) \quad |\nabla_y V_1(y)[\xi]| \leq \sqrt{m_1 \nabla_{yy}^2 V_1(y)[\xi, \xi]}, \quad \forall \xi \in \mathbb{R}^{m_1}.$$

Then we have that for all $\xi \in \mathbb{R}^{m_1}$,

$$\begin{aligned} & |\nabla_y \eta'(\mu, y)^\top [\xi]| \\ &= |(c_1 \nabla_y V_1(y) + \sum_{k=1}^K c_2 \nabla_y V_2^{(k)}(y, \bar{x}^{(k)}))^\top [\xi]| \\ &\leq |c_1 \nabla_y V_1(y)^\top [\xi]| + \sum_{k=1}^K |c_2 \nabla_y V_2^{(k)}(y, \bar{x}^{(k)})^\top [\xi]| \\ &\leq \sqrt{m_1 c_1^2 \nabla_{yy}^2 V_1(y)[\xi, \xi]} + \sum_{k=1}^K \sqrt{m_2 c_2^2 \nabla_{yy}^2 V_2^{(k)}(y, \bar{x}^{(k)})[\xi, \xi]} \\ &= \sqrt{(m_1 c_1) c_1 \nabla_{yy}^2 V_1(y)[\xi, \xi]} + \sum_{k=1}^K \sqrt{(m_2 c_2) c_2 \nabla_{yy}^2 V_2^{(k)}(y, \bar{x}^{(k)})[\xi, \xi]} \\ &\leq \sqrt{(m_1 c_1 + m_2 c_2)(1 + K)(c_1 \nabla_{yy}^2 V_1(y)[\xi, \xi] + \sum_{k=1}^K c_2 \nabla_{yy}^2 V_2^{(k)}(y, \bar{x}^{(k)})[\xi, \xi])} \\ &= \sqrt{\frac{(m_1 c_1 + m_2 c_2)(1 + K)}{\mu} \nabla_{yy}^2 \eta(\mu, y)[\xi, \xi]}. \quad \square \end{aligned}$$

With Lemma 2 and Lemma 3 established, we have that Theorem 3 is true.

5. A CLASS OF VOLUMETRIC BARRIER ALGORITHMS FOR SOLVING SSDPs

In §4 we have established that the parametric function $\eta(\mu, \cdot)$ is a strongly self-concordant family. In this section we introduce a class of volumetric barrier algorithms for solving (9) and (10). This class, indexed by a parameter $\gamma \in (0, 1)$, is stated formally in Algorithm 1.

Our algorithm is initialized with a starting point $y^0 \in \mathcal{F}_0$ and a starting value $\mu^0 > 0$ for the barrier parameter μ . We use δ as a measure of the proximity of the current point y to the central path, and β as a threshold for that measure. If the current y is too far away from the central path in the sense that $\delta > \beta$, we apply Newton’s method to find a point close to the central path. Then we reduce the value of μ by a factor γ and repeat the whole process until the value of μ is within the tolerance ϵ .

6. COMPLEXITY ANALYSIS

For fixed $\mu > 0$, the function $\eta(\mu, \cdot)$ possesses many nice properties. The following proposition follows directly from the definition of self-concordance and [19, Theorem 2.1.1].

Proposition 8. *For any $\mu > 0$, $y \in \mathcal{F}_0$ and Δy , we denote*

$$\delta := \sqrt{\frac{1}{\mu} \Delta y^\top \nabla_{yy}^2 \eta(\mu, y) \Delta y}. \text{ Then for } \delta < 1, \tau \in [0, 1] \text{ and any } \xi \in \mathbb{R}^{m_2} \text{ we have}$$

$$\xi^\top \nabla_{yy}^2 \eta(\mu, y + \tau \Delta y) \xi \leq (1 - \tau \delta)^{-2} \xi^\top \nabla_{yy}^2 \eta(\mu, y) \xi.$$

Algorithm 1. Volumetric Barrier Algorithm for Solving SSDP (9)–(10)

Require: $\epsilon > 0$, $\gamma \in (0, 1)$, $\theta > 0$, $\beta > 0$, $y^0 \in \mathcal{F}_0$ and $\mu^0 > 0$.

$y := y^0$, $\mu := \mu^0$
while $\mu \geq \epsilon$ **do**
 for $k = 1, 2, \dots, K$ **do**
 solve (12) to obtain $\bar{x}^{(k)}$
 end for
 compute $\Delta y := -[\nabla_{yy}^2 \eta(\mu, y)]^{-1} \nabla_y \eta(\mu, y)$ using (18)
 compute $\delta(\mu, y) := \sqrt{\frac{1}{\mu} \Delta y^\top \nabla_{yy}^2 \eta(\mu, y) \Delta y}$ using (18)
 while $\delta > \beta$ **do**
 $y := y + \theta \Delta y$
 for $k = 1, 2, \dots, K$ **do**
 solve (12) to obtain $\bar{x}^{(k)}$
 end for
 compute $\Delta y := -[\nabla_{yy}^2 \eta(\mu, y)]^{-1} \nabla_y \eta(\mu, y)$ using (18)
 compute $\delta(\mu, y) := \sqrt{\frac{1}{\mu} \Delta y^\top \nabla_{yy}^2 \eta(\mu, y) \Delta y}$ using (18)
 end while
 $\mu := \gamma \mu$
end while

We also have the following lemma that describes the behavior of the Newton method as applied to $\eta(\mu, \cdot)$. This lemma is essentially a restatement of [19, Theorem 2.2.3] for the setting of the present paper.

Lemma 4. *For any $\mu > 0$ and $y \in \mathcal{F}_0$, let Δy be the Newton direction defined by $\Delta y := -[\nabla_{yy}^2 \eta(\mu, y)]^{-1} \nabla_y \eta(\mu, y)$. We denote*

$$\delta := \delta(\mu, y) = \sqrt{\frac{1}{\mu} \Delta y^\top \nabla_{yy}^2 \eta(\mu, y) \Delta y}.$$

Then the following relations hold:

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

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

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

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

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

Depending on the manner in which γ in Algorithm 1 is chosen, we have two classes of algorithms: short-step algorithms and long-step algorithms. In the next two subsections we present the complexity analysis for these two classes of algorithms.

6.1. Complexity of short step algorithms. The i^{th} iteration of the short-step algorithms is performed as follows: at the beginning of the iteration, we have $\mu^{(i-1)}$ and $y^{(i-1)}$ on hand and $y^{(i-1)}$ is close to the center path, i.e., $\delta(\mu^{(i-1)}, y^{(i-1)}) \leq$

β . After we reduce the parameter μ from $\mu^{(i-1)}$ to $\mu^i := \gamma\mu^{(i-1)}$, we have that $\delta(\mu^i, y^{(i-1)}) \leq 2\beta$. Then a full Newton step is taken to find a new point y^i with $\delta(\mu^i, y^i) \leq \beta$. We will show that in each iteration after we reduce the parameter μ , one Newton step is sufficient to restore the proximity to the central path. We assume that we can solve all the subproblems exactly and we fix the value of $\gamma := 1 - 0.1/\sqrt{(m_1c_1 + m_2c_2)(1 + K)}$. We have the following lemma and the proof of the lemma follows from [19, Theorem 3.1.1].

Lemma 5. *Let $\gamma := 1 - 0.1/\sqrt{(m_1c_1 + m_2c_2)(1 + K)}$ and $\beta := (2 - \sqrt{3})/2$. If $\delta(\mu, y) \leq \beta$, then $\delta(\gamma\mu, y) \leq 2\beta$.*

Proof. In order to apply [19, Theorem 3.1.1], we first write the metric defined by (3.1.4) in [19] for our problem as follows: For any $0 < \mu^+ < \mu$,

$$\chi_\kappa(\eta; \mu, \mu^+) := \left(1 + \frac{\sqrt{(m_1c_1 + m_2c_2)(1 + K)}}{\kappa} \right) \ln \left(\frac{\mu}{\mu^+} \right).$$

Let $\kappa := 2\beta := 2 - \sqrt{3}$. Since $\delta(\mu, y) \leq \kappa/2$, one can verify that $\mu^+ := \gamma\mu$ satisfies

$$\chi_\kappa(\eta; \mu, \mu^+) \leq \frac{1}{2} \leq 1 - \frac{\delta(\mu, y)}{\kappa}.$$

So by virtue of [19, Theorem 3.1.1], we have $\delta(\mu^+, y) \leq \kappa$. □

By inequality in Lemma 4(i) and Lemma 5 we have that in Algorithm 1, we can reduce the parameter μ by the factor $\gamma := 1 - 0.1/\sqrt{(m_1c_1 + m_2c_2)(1 + K)}$ at each iteration, and use only one Newton step for recentering if necessary. So we have the following complexity result for short-step algorithms.

Theorem 4. *Let $\beta := (2 - \sqrt{3})/2$ and $\gamma := 1 - 0.1/\sqrt{(m_1c_1 + m_2c_2)(1 + K)}$ in Algorithm 1. If $\delta(\mu^0, y^0) \leq \beta$, then short-step algorithms terminate with at most $O(\sqrt{(m_1c_1 + m_2c_2)(1 + K)}\ln(\mu^0/\epsilon))$ iterations.*

6.2. Complexity of the long step algorithms. In the long-step version of the algorithm, the factor $\gamma \in (0, 1)$ is arbitrarily chosen. It has potential for a larger decrease on the objective function value; however, several damped Newton steps might be needed for recentering.

Suppose at the beginning of the i^{th} iteration of the algorithm we have a point $y^{(i-1)}$, which is sufficiently close to $y(\mu^{(i-1)})$, where $\mu^{(i-1)}$ is the current value for the barrier parameter μ and $y(\mu^{(i-1)})$ is the solution to (11) for $\mu := \mu^{(i-1)}$. We reduce the barrier parameter from $\mu^{(i-1)}$ to $\mu^i := \gamma\mu^{(i-1)}$, where $\gamma \in (0, 1)$, and we search for a point y^i that is sufficiently close to $y(\mu^i)$. We want to determine an upper bound on the number of Newton iterations that are needed to find the point y^i .

We begin by defining $\phi(\mu) := \eta(\mu, y) - \eta(\mu, y(\mu))$, $\forall \mu > 0$. The function ϕ stands for the difference between the objective value $\eta(\mu^i, y^{(i-1)})$ at the beginning of i^{th} iteration and the minimum objective value $\eta(\mu^i, y(\mu^i))$ at the end of i^{th} iteration. Then our task is to find an upper bound on $\phi(\mu)$.

The next lemma gives us upper bounds on $\phi(\mu)$ and $\phi'(\mu)$, respectively.

Lemma 6. *let $\mu > 0$ and $y \in \mathcal{F}_0$, we denote $\tilde{\Delta}y := y - y(\mu)$ and define*

$$\tilde{\delta} := \sqrt{\frac{1}{\mu} \tilde{\Delta}y^\top \nabla_{yy}^2 \eta(\mu, y) \tilde{\Delta}y}.$$

If $\tilde{\delta} < 1$, then the following inequalities hold:

$$(34) \quad \phi(\mu) \leq \left[\frac{\tilde{\delta}}{1 - \tilde{\delta}} + \ln(1 - \tilde{\delta}) \right] \mu,$$

$$(35) \quad |\phi'(\mu)| \leq -\ln(1 - \tilde{\delta}) \sqrt{(m_1 c_1 + m_2 c_2)(1 + K)}.$$

Proof.

$$\phi(\mu) := \eta(\mu, y) - \eta(\mu, y(\mu)) := \int_0^1 \nabla_y \eta(\mu, y(\mu) + (1 - \tau)\tilde{\Delta}y)^\top \tilde{\Delta}y d\tau.$$

Since $y(\mu)$ is the optimal solution, we have

$$(36) \quad \nabla_y \eta(\mu, y(\mu)) = 0.$$

Hence,

$$\begin{aligned} \phi(\mu) &= \int_0^1 \int_0^1 \tilde{\Delta}y^\top \nabla_{yy}^2 \eta(\mu, y(\mu) + (1 - \alpha)\tilde{\Delta}y) \tilde{\Delta}y d\alpha d\tau \\ &\leq \int_0^1 \int_0^1 \frac{\mu \tilde{\delta}^2}{(1 - \tilde{\delta} + \alpha \tilde{\delta})^2} d\alpha d\tau \quad (\text{by Proposition 8}) \\ &= \left[\frac{\tilde{\delta}}{1 - \tilde{\delta}} + \ln(1 - \tilde{\delta}) \right] \mu. \end{aligned}$$

This proves (34).

For any $\mu > 0$,

$$(37) \quad \begin{aligned} \phi'(\mu) &= \eta'(\mu, y) - \eta'(\mu, y(\mu)) - \nabla \eta(\mu, y(\mu))^\top y'(\mu) \\ &= \eta'(\mu, y) - \eta'(\mu, y(\mu)). \end{aligned}$$

From Lemma 3 we have that for any $\mu > 0$ and $y \in \mathcal{F}_0$,

$$(38) \quad \nabla \eta'(\mu, y)^\top [\nabla_{yy}^2 \eta(\mu, y)]^{-1} \nabla \eta'(\mu, y) \leq \frac{1}{\mu} (m_1 c_1 + m_2 c_2)(1 + K).$$

Applying the Mean-Value Theorem we obtain

$$\begin{aligned} |\phi'(\mu)| &= \left| \int_0^1 \nabla \eta'(\mu, y(\mu) + \tau \tilde{\Delta}y)^\top \tilde{\Delta}y d\tau \right| \\ &\leq \int_0^1 \sqrt{\tilde{\Delta}y^\top \nabla_{yy}^2 \eta(\mu, y(\mu) + \tau \tilde{\Delta}y) \tilde{\Delta}y} \\ &\quad \sqrt{\nabla \eta'(\mu, y(\mu) + \tau \tilde{\Delta}y)^\top [\nabla_{yy}^2 \eta(\mu, y(\mu) + \tau \tilde{\Delta}y)]^{-1} \nabla \eta'(\mu, y(\mu) + \tau \tilde{\Delta}y) d\tau}. \end{aligned}$$

Using (38) and Proposition 8, and noting that $y(\mu) + \tau \tilde{\Delta}y = y - (1 - \tau)\tilde{\Delta}y$, we have

$$\begin{aligned} |\phi'(\mu)| &\leq \int_0^1 \frac{\tilde{\delta} \sqrt{\mu}}{1 - \tilde{\delta} + \tau \tilde{\delta}} \sqrt{\frac{(m_1 c_1 + m_2 c_2)(1 + K)}{\mu}} d\tau \\ &= -\ln(1 - \tilde{\delta}) \sqrt{(m_1 c_1 + m_2 c_2)(1 + K)}. \end{aligned}$$

This proves (35). □

Lemma 7. *Let $\mu > 0$ and $y \in \mathcal{F}_0$ be such that $\tilde{\delta} < 1$, where $\tilde{\delta}$ is as defined in Lemma 6. Let $\mu^+ := \gamma\mu$ with $\gamma \in (0, 1)$. Then*

$$\eta(\mu^+, y) - \eta(\mu^+, y(\mu^+)) \leq O(1)[(m_1c_1 + m_2c_2)(1 + K)]\mu^+.$$

Proof. We have

$$\phi''(\mu, y) = \eta''(\mu, y) - \eta''(\mu, y(\mu)) - \nabla\eta'(\mu, y(\mu))^\top y'(\mu).$$

Since $\eta(\cdot, y)$ is strictly concave in μ , the first term $\eta''(\mu, y)$ is negative. We only need to estimate the other two terms.

First we differentiate (36) with respect to μ to obtain

$$y'(\mu) = -[\nabla_{yy}^2\eta(\mu, y(\mu))]^{-1}\nabla\eta'(\mu, y(\mu)).$$

Hence, from Lemma 3 we have the following estimation:

$$\begin{aligned} -\nabla\eta'(\mu, y(\mu))^\top y'(\mu) &= \nabla\eta'(\mu, y(\mu))^\top [\nabla_{yy}^2\eta(\mu, y(\mu))]^{-1}\nabla\eta'(\mu, y(\mu)) \\ (39) \qquad \qquad \qquad &\leq \frac{1}{\mu}(m_1c_1 + m_2c_2)(1 + K). \end{aligned}$$

Now we want to estimate $\eta''(\mu, y)$ for any $\mu > 0$ and $y \in \mathcal{F}_0$. First we observe that $\eta'' = \sum_{k=1}^K \rho_k''(\mu, y)$. Differentiating $\rho_k(\mu, y)$ with respect to μ we obtain $\rho_k'(\mu, y) = c_2V_2(y, \bar{x})$. Differentiating again we obtain $\rho_k''(\mu, y) = c_2\nabla_{\bar{x}}V_2(y, \bar{x})^\top \bar{x}'$. We differentiate (16) to obtain

$$\bar{x}' = -\frac{1}{\mu}[\nabla_{\bar{x}}^2V_2(y, \bar{x})]^{-1}\nabla_{\bar{x}}V_2(y, \bar{x}).$$

Hence, we have

$$\rho_k''(\mu, y) = -\frac{1}{\mu}c_2\nabla_{\bar{x}}V_2(y, \bar{x})^\top [\nabla_{\bar{x}}^2V_2(y, \bar{x})]^{-1}\nabla_{\bar{x}}V_2(y, \bar{x}).$$

By [4, Theorem 4.4] we have $-\rho_k''(\mu, y) \leq \frac{1}{\mu}c_2m_2$. Therefore,

$$(40) \qquad -\eta''(\mu, y(\mu)) = -\sum_{k=1}^K \rho_k''(\mu, y) \leq \sum_{k=1}^K \frac{1}{\mu}c_2m_2 = \frac{1}{\mu}m_2c_2K.$$

Combining (39) and (40), we have

$$(41) \qquad \phi''(\mu) \leq \frac{1}{\mu}(m_1c_1 + 2m_2c_2)(1 + K).$$

Using Lemma 6 and (41), we have

$$\begin{aligned} \phi(\mu^+) &= \phi(\mu) + (\mu^+ - \mu)\phi'(\mu) + \int_{\mu^+}^{\mu} \int_{\tau}^{\mu} \phi''(\mu) d\mu d\tau \\ &\leq \left[\frac{\tilde{\delta}}{1 - \tilde{\delta}} + \ln(1 - \tilde{\delta}) \right] \mu - \ln(1 - \tilde{\delta}) \sqrt{(m_1 c_1 + m_2 c_2)(1 + K)} (\mu^+ - \mu) \\ &\quad + (m_1 c_1 + 2m_2 c_2)(1 + K) \int_{\mu^+}^{\mu} \int_{\tau}^{\mu} \mu^{-1} d\mu d\tau \\ &\leq \left[\frac{\tilde{\delta}}{1 - \tilde{\delta}} + \ln(1 - \tilde{\delta}) \right] \mu - \ln(1 - \tilde{\delta}) \sqrt{(m_1 c_1 + m_2 c_2)(1 + K)} (\mu^+ - \mu) \\ &\quad + (m_1 c_1 + 2m_2 c_2)(1 + K) \ln \gamma^{-1} (\mu - \mu^+). \end{aligned}$$

This proves the lemma since $\tilde{\delta}$ and γ are constants. □

In the previous lemmas we require $\tilde{\delta} < 1$. However, $\tilde{\delta}$ cannot be evaluated explicitly. In the next lemma we will see that $\tilde{\delta}$ is actually proportional to δ , which can be evaluated.

Lemma 8. *For any $\mu > 0$ and $y \in \mathcal{F}_0$, let $\Delta y := -[\nabla_{yy}^2 \eta(\mu, y)]^{-1} \nabla_y \eta(\mu, y)$ and let $\tilde{\Delta} y := (y - y(\mu))$. We denote*

$$\delta := \sqrt{\frac{1}{\mu} \Delta y^\top \nabla_{yy}^2 \eta(\mu, y) \Delta y}, \quad \tilde{\delta} := \sqrt{\frac{1}{\mu} \tilde{\Delta} y^\top \nabla_{yy}^2 \eta(\mu, y) \tilde{\Delta} y}.$$

If $\delta < 1/6$, then

$$\frac{2}{3} \delta \leq \tilde{\delta} \leq 2\delta.$$

Proof. See the proof of Lemma 9 in [4]. □

Combining Lemmas 4, 7, and 8, we have the following theorem.

Theorem 5. *Let $\beta := 1/6$ and $\gamma \in (0, 1)$ be arbitrary in Algorithm 1. If $\delta(\mu^0, y^0) \leq \beta$, then long-step algorithms terminate with at most $O((m_1 c_1 + m_2 c_2)(1 + K) \ln(\mu^0/\epsilon))$ iterations.*

7. CONCLUDING REMARKS

In this paper we have presented a class of volumetric barrier decomposition algorithms for (two-stage) stochastic semidefinite programs (SSDPs) (with recourse). We have also shown that certain short-step and long-step members of the class have polynomial complexity in terms of the number of iterations with the complexity bounds depending on \sqrt{K} and K , respectively, where K is the number of realizations. This is important given the fact K can be large in applications.

The complexity of our algorithms and of those in [18] are similar. Both are $O(\sqrt{K})$ for short-step algorithms and $O(K)$ for long-step algorithms.

SSDPs generalize (two-stage) stochastic linear programs (SLPs) (with recourse). Therefore, it is possible to specialize the class of algorithms presented in this paper to SLPs. The specialization is a new class of algorithms for SLPs. Indeed, in

[6] we show that we can go further by showing how appropriate modification of the techniques utilized in the present paper leads to a class of new volumetric barrier decomposition algorithms for stochastic quadratic programs with quadratic recourse.

It would be interesting to assess the computational performance of the algorithms developed in the present paper. A forthcoming paper will report details of an implementation and results of computational experiments performed with it.

REFERENCES

- [1] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM J. Optim.*, 5:13–51, 1995. MR1315703 (95k:90065)
- [2] K. M. Anstreicher. On Vaidya’s volumetric cutting plane method for convex programming. *Mathematics of Operations Research*, 22(1):63–89, 1997. MR1436574 (98b:90103)
- [3] K. M. Anstreicher. Volumetric path following algorithms for linear programming. *Mathematical Programming*, 76(1B):245–263, 1997. MR1426406 (98a:90057)
- [4] K. M. Anstreicher. The volumetric barrier for semidefinite programming. *Mathematics of Operations Research*, 25(3):365–380, 2000. MR1855171 (2002f:90078)
- [5] K. A. Ariyawansa and Y. Zhu. Stochastic Semidefinite Programming: A New Paradigm for Stochastic Optimization. *4OR—The Quarterly Journal of the Belgian, French and Italian OR Societies*, 4(3):65–79, 2006. (An earlier version of this paper appeared as Technical Report 2004-10 of the Department of Mathematics, Washington State University, Pullman, WA 99164-3113, in October 2004.) MR2262090 (2008e:90059)
- [6] K. A. Ariyawansa and Y. Zhu. A class of volumetric barrier decomposition algorithms for stochastic quadratic programming. *Applied Mathematics and Computation*, 186(2):1683–1693, 2007. MR2319065
- [7] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, NY, 1997. MR1460264 (99b:90001)
- [8] M. A. H. Dempster. *Stochastic Programming*. Academic Press, London, U.K., 1980. MR592594 (81j:90094)
- [9] Y. Ermoliev and R. J-B. Wets. *Numerical Techniques for Stochastic Optimization*. Springer-Verlag, Berlin, Germany, 1988. MR957304 (90h:90131)
- [10] A. J. Felt. *A Computational Evaluation of Interior Point Cutting Plane Algorithms for Stochastic Programs*. Ph.D. dissertation, Department of Pure and Applied Mathematics, Washington State University, Pullman, WA, 2000.
- [11] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley, New York, NY, 1968. MR0243831 (39:5152)
- [12] A. Graham. *Kronecker Products and Matrix Calculus: With Applications*. Ellis Horwood, Chichester, U.K., 1981. MR640865 (83g:15001)
- [13] C. C. Gonzaga. Path following methods for linear programming. *SIAM Review*, 34:167–224, 1992. MR1166175 (93j:90050)
- [14] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, U.K., 1985. MR832183 (87e:15001)
- [15] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, U.K., 1991. MR1091716 (92e:15003)
- [16] P. L. Jiang. *Polynomial Cutting Plane Algorithms for Stochastic Programming and Related Problems*. Ph.D. dissertation, Department of Pure and Applied Mathematics, Washington State University, Pullman, WA, 1997.
- [17] P. Kall and S. Wallace. *Stochastic Programming*. Wiley, New York, NY, 1994. MR1315300 (96h:90002)
- [18] S. Mehrotra and M. G. Özevin. Decomposition-based interior point methods for two-stage stochastic semidefinite programming. *SIAM J. of Optimization*, 18(1):206–222, 2007. (An earlier draft of this paper appeared under the title “Two-state stochastic semidefinite programming and decomposition based interior point methods: theory,” IEMS Technical Report 2004-16, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, in December 2004.) MR2299681 (2007m:90135)

- [19] Yu. E. Nesterov and A. S. Nemirovskii. *Interior Point Polynomial Algorithms in Convex Programming*. SIAM Publications, Philadelphia, PA, 1994. MR1258086 (94m:90005)
- [20] A. Prékopa. *Stochastic Programming*. Kluwer Academic Publishers, Boston, MA, 1995. MR1375234 (97f:90001)
- [21] P. M. Vaidya. A new algorithm for minimizing convex functions over a convex set. *Math. Program., Ser. A*, 73:291–341, 1996. MR1393123 (97e:90060)
- [22] R.J. Vanderbei and H.Y. Benson. *On Formulating Semidefinite Programming Problems as Smooth Convex Nonlinear Optimization Problems*. Working Paper, Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ, 2000.
- [23] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Rev.*, 38:49–95, 1996. MR1379041 (96m:90005)
- [24] H. Wolkowicz, R. Saigal and L. Vandenberghe. *Handbook of Semidefinite Programming—Theory, Algorithms, and Applications*. Kluwer Academic Publishers, Norwell, MA, 2000. MR1778223 (2001k:90001)
- [25] M. J. Todd. Semidefinite Optimization. *ACTA Numerica*, 10:515–560, 2001. MR2009698 (2004g:90004)
- [26] G. Zhao. A log-barrier method with Benders decomposition for solving two-stage stochastic linear programs. *Math. Program., Ser. A*, 90:507–536, 2001. MR1832221 (2002g:90067)

DEPARTMENT OF MATHEMATICS, WASHINGTON STATE UNIVERSITY, PULLMAN, WASHINGTON 99164-3113

E-mail address: ari@wsu.edu

DIVISION OF MATHEMATICAL AND NATURAL SCIENCES, ARIZONA STATE UNIVERSITY, PHOENIX, AZ 85069-7100

E-mail address: Yuntao.Zhu@asu.edu