SUPERLINEAR CONVERGENCE OF A SYMMETRIC PRIMAL-DUAL PATH FOLLOWING ALGORITHM FOR SEMIDEFINITE PROGRAMMING

ZHI-QUAN LUO*, JOS F. STURM†, AND SHUZHONG ZHANG‡

Abstract. This paper establishes the superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming under the assumptions that the semidefinite program has a strictly complementary primal-dual optimal solution and that the size of the central path neighborhood tends to zero. The interior point algorithm considered here closely resembles the Mizuno-Todd-Ye predictor-corrector method for linear programming where it is known to be quadratically convergent. It is shown that when the iterates are well centered, the duality gap is reduced superlinearly after each predictor step. Indeed, if each predictor step is succeeded by r consecutive corrector steps then the predictor reduces the duality gap superlinearly with order $\frac{2}{1+2^{-r}}$. The proof relies on a careful analysis of the central path for semidefinite programming. It is shown that under the strict complementarity assumption, the primal-dual central path converges to the analytic center of the primal-dual optimal solution set, and the distance from any point on the central path to this analytic center is bounded by the duality gap.

Key words. Semidefinite programming, central path, path following, superlinear convergence.

AMS subject classifications. 90C25, 90C26, 90C60.

1. Introduction. Recently, there have been many interior point algorithms developed for semidefinite programming (SDP), see for example [1, 2, 4, 8, 12, 14, 15, 17, 21]. These algorithms differ in their choices of scaling matrix, the size of the central path neighborhoods, and stepsize rules, among others. In particular, the algorithms of Kojima-Shindoh-Hara [8] and Nesterov-Todd [14, 15] are based on the primal-dual scaling and they both can be viewed as extensions of the predictor-corrector method for linear programming [11]. It has been shown [7, 9, 14, 15, 17, 21] that these algorithms for SDP retain many important properties of the interior point algorithms for linear programming including polynomial complexity. For an overview of SDP and its applications, we refer to Vandenberghe and Boyd [19].

However, there exists considerable difficulty in extending one key property of the predictor-corrector method for linear programming to the interior point algorithms for SDP. This is the property of quadratic convergence of the duality gap (see [20] for a proof of the LCP case). In some sense, the need for superlinear convergence in solving SDP is more pronounced than that for the linear programming case. This is because for SDP there cannot exist any finite termination procedures as in the case of linear programming. Indeed, the recent papers of Kojima-Shida-Shindoh [7] and Potra-Sheng [16] are both focused on the issue of superlinear convergence for solving SDP. In particular, the latter reference provided a sufficient condition for the superlinear convergence of an infeasible path following algorithm, while the former reference [7] established the superlinear convergence of their algorithm [8] under certain key assumptions. These assumptions are: (1) SDP is nondegenerate in the sense that the Jacobian matrix of its KKT system is nonsingular; (2) SDP has a strictly comple-

*Room 225/CRL, Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, L8S 4L7, Canada (luozq@mcmail.cis.mcmaster.ca). Supported by the Natural Sciences and Engineering Research Council of Canada, Grant No. OPG0090391, during a research leave to the Econometric Institute, Erasmus University Rotterdam.

†Econometric Institute, Erasmus University Rotterdam, The Netherlands (sturm@few.eur.nl).

‡Econometric Institute, Erasmus University Rotterdam, The Netherlands (zhang@few.eur.nl).

mentary optimal solution; (3) the iterates converge tangentially to the central path in the sense that the size of the central path neighborhood in which the iterates reside must tend to zero. Among these three assumptions for superlinear convergence, (2) is inevitable since it is needed even in the case of LCP (see [20]). Assumption (3) is needed to ensure the duality gap is reduced superlinearly after each predictor step for *all* points in the central path neighborhood. In reference [7], an example was given which showed that, without the tangential convergence assumption, the duality gap is reduced only linearly after one predictor step for certain points in the central path neighborhood.

Our goal in this paper is to establish the superlinear convergence of a symmetric path following algorithm for SDP under only assumptions (2) and (3) (i.e., without the nondegeneracy assumption). In particular, we consider the primal-dual path following algorithm of Nesterov-Todd [14, 15] (later discovered independently by Sturm and Zhang [17] using a V-space notion). In this paper we adopt the framework of [17] since it greatly facilitates the subsequent analysis. We show that this symmetric primal-dual path following algorithm has an order of convergence that is asymptotically quadratic (i.e., sub-quadratic). Indeed, for any given constant positive integer r, the algorithm can be set so that the duality gap decreases superlinearly with order $\frac{2}{1+2^{-r}}$ after one predictor (affine scaling) step followed by (at most) r corrector steps. The cornerstone in our bid to establish this superlinear convergence result is a bound on the distance from any point on the central path to the optimal solution set (see Section 3). Specifically, it is shown that, under the strict complementarity assumption, the primal-dual central path converges to the analytic center of the optimal solution set, and that the distance to this analytic center from any point on the central path can be bounded above by the duality gap. These properties of the central path are algorithm-independent and are likely to be useful in the analysis of other interior point algorithms for SDP.

The organization of this paper is as follows. At the end of this section, we describe some basic notation to be used in this paper. In Section 2, we will discuss some fundamental background notions, and we will make two assumptions concerning the solution set of the SDP. In Section 3 we will analyze the limiting behavior of the primal-dual central path. In Section 4, the notion of V-space for SDP is reviewed and a path following algorithm in the spirit of [17] is introduced. The superlinear convergence of this algorithm is established in Section 5. Finally, some concluding remarks are given in Section 6.

Notation. The space of symmetric $n \times n$ matrices will be denoted $\mathcal{S}$. Given matrices X and Y in $\Re^{m_1 \times m_2}$, the standard inner product is defined by

$$X \bullet Y = \operatorname{tr} X^T Y,$$

where $\operatorname{tr}(\cdot)$ denotes the trace of a matrix. The notation $X \perp Y$ denotes orthogonality in the sense that $X \bullet Y = 0$. The Euclidean norm and its associated operator norm, viz. the spectral norm, are both denoted by $\|\cdot\|$. The Frobenius norm of X is $\|X\|_F = \sqrt{X \bullet X}$. If $X \in \mathcal{S}$ is positive (semi-) definite, we write $(X \succeq 0)$ $X \succ 0$. The cone of positive semi-definite matrices is denoted by $\mathcal{S}_+$ and the cone of positive definite matrices is $\mathcal{S}_{++}$. The identity matrix is denoted by I. We use the standard Landau notation ("big O" and "small o"). In particular, if $\{u(\mu) : \mu > 0\}$ and $\{w(\mu) : \mu > 0\}$ are real sequences with $w(\mu) > 0$, then $u(\mu) = O(w(\mu))$ means that $u(\mu)/w(\mu)$ is bounded, independent of μ; $u(\mu) = o(w(\mu))$ means that $\lim_{\mu \to 0} u(\mu)/w(\mu) = 0$; $u(\mu) \sim w(\mu)$ means that $\lim_{\mu \to 0} u(\mu)/w(\mu) = 1$; $u(\mu) = \Theta(w(\mu))$ whenever $u(\mu)/w(\mu)$

and $w(\mu)/u(\mu)$ are both bounded. For a symmetric positive definite matrix, we use "O" and "Θ" to denote the order of all its eigenvalues. Hence, for $U(\mu) \in \mathcal{S}_{++}$, the notation $U(\mu) = \Theta(w(\mu))$ signifies the existence of $\Gamma > 0$ such that

$$\frac{1}{\Gamma} I \preceq \frac{1}{w(\mu)} U(\mu) \preceq \Gamma I, \qquad \text{for all } \mu > 0.$$

Therefore, the condition $U(\mu) = \Theta(w(\mu))$ implies that $\|U(\mu)\| = \Theta(w(\mu))$ and the diagonal entries

$$U_{11}(\mu),\ U_{22}(\mu),\ \ldots,\ U_{nn}(\mu)$$

are all $\Theta(w(\mu))$ too.

2. Problem formulation. A semidefinite programming (SDP) problem is given as

$$\begin{array}{lll} \text{minimize} & C \bullet X & \\ \text{subject to} & A^{(i)} \bullet X = b_i, \quad \text{for } i = 1, 2, \ldots, m, & \text{(P)} \\ & X \succeq 0 & \end{array}$$

where $C \in \mathcal{S}$, $A^{(1)}, A^{(2)}, \ldots, A^{(m)} \in \mathcal{S}$ and $b \in \Re^m$. The decision variable is $X \in \mathcal{S}$. The corresponding dual program can be formulated as

$$\begin{array}{lll} \text{maximize} & b^T y & \\ \text{subject to} & \sum_{i=1}^m y_i A^{(i)} + Z = C, & \text{(D)} \\ & Z \succeq 0. & \end{array}$$

Denote the feasible sets of (P) and (D) by $\mathcal{F}_P$ and $\mathcal{F}_D$ respectively, i.e.

$$\mathcal{F}_P := \{X \in \mathcal{S} :\ A^{(i)} \bullet X = b_i,\ i = 1, 2, \ldots, m,\ X \succeq 0\},$$

and

$$\mathcal{F}_D := \{Z \in \mathcal{S} :\ Z = C - \sum_{i=1}^m y_i A^{(i)} \text{ for some } y \in \Re^m,\ Z \succeq 0\}.$$

We make the following assumptions throughout this paper.

ASSUMPTION 2.1. *There exist positive definite solutions $X \in \mathcal{F}_P$ and $Z \in \mathcal{F}_D$ for* (P) *and* (D) *respectively.*

ASSUMPTION 2.2. *There exists a pair of strictly complementary primal-dual optimal solutions for* (P) *and* (D). *Specifically, there exists $(X^*, Z^*) \in \mathcal{F}_P \times \mathcal{F}_D$ such that*

$$\left\{ \begin{array}{l} X^* Z^* = 0, \\ X^* + Z^* \succ 0. \end{array} \right.$$

Since $X^* Z^* = Z^* X^* = 0$, we can diagonalize X^* and Z^* simultaneously. Therefore, by applying an orthonormal transformation to the problem data if necessary, we can assume without loss of generality that X^*, Z^* are both diagonal and of the form

$$X^* = \begin{bmatrix} \Lambda_B & 0 \\ 0 & 0 \end{bmatrix}, \quad Z^* = \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_N \end{bmatrix}, \tag{2.1}$$

where $\Lambda_B := \mathrm{diag}(\lambda_1, ..., \lambda_K)$, $\Lambda_N := \mathrm{diag}(\lambda_{K+1}, ..., \lambda_n)$ for some integer $0 \le K \le n$ and some positive scalars $\lambda_i > 0$, $i = 1, 2, ..., n$. Here the subscripts B and N signify the "basic" and "nonbasic" subspaces (following the terminology of linear programming). Throughout this paper, the decomposition of any $n \times n$ matrix X is always made with respect to the above partition B and N. In fact, we shall adhere to the following notation throughout:

$$X = \begin{bmatrix} X_B & X_U \\ X_U^T & X_N \end{bmatrix},$$

so X_U will always denote the off-diagonal block of X with size $K \times (n-K)$, etc.

Notice that $X \in \mathcal{F}_P$ is an optimal solution to (P) if and only if $XZ^* = 0$. Hence, by Assumption 2.2, the primal optimal solution set can be written as

$$\mathcal{F}_P^* := \{X \in \mathcal{F}_P : X_U = 0 \text{ and } X_N = 0\}.$$

Analogously, the dual optimal solution set is given by

$$\mathcal{F}_D^* := \{Z \in \mathcal{F}_D : Z_U = 0 \text{ and } Z_B = 0\}.$$

Given $\mu \in \Re_{++}$, the pair $(X, Z) \in \mathcal{F}_P \times \mathcal{F}_D$ is said to be the *μ-center* $(X(\mu), Z(\mu))$ if and only if

$$XZ = \mu I. \tag{2.2}$$

We refer to [8, 18] for a proof of the existence and uniqueness of μ-centers. The *central path* of the problem (P) is the curve

$$\{(X(\mu), Z(\mu)) : \mu > 0\}.$$

The logarithmic barrier function $\log\det X_B$ is strictly concave on the relative interior of the primal optimal solution set

$$\{X \in \mathcal{F}_P^* : X_B \succ 0\}.$$

Moreover, Assumption 2.1 implies that $\mathcal{F}_P^*$ and $\mathcal{F}_D^*$ are bounded. The unique maximizer X^a of $\log\det X_B$ on the relative interior of $\mathcal{F}_P^*$ is called the *analytic center* of $\mathcal{F}_P^*$. It is characterized by the Karush-Kuhn-Tucker system

$$\begin{cases} X_B^a Z_B = I, \\ \sum_{i=1}^m y_i A_B^{(i)} + Z_B = 0, \\ X^a \in \mathcal{F}_P^* \text{ and } Z_B \succ 0. \end{cases} \tag{2.3}$$

In a similar fashion, we define the analytic center of $\mathcal{F}_D^*$ as the maximizer of the logarithmic barrier $\log\det Z_N$ on the relative interior of $\mathcal{F}_D^*$; Z^a is characterized by the Karush-Kuhn-Tucker system

$$\begin{cases} X_N Z_N^a = I, \\ A^{(i)} \bullet X = 0, \qquad i = 1, 2, \ldots, m, \\ Z^a \in \mathcal{F}_D^* \text{ and } X_N \succ 0. \end{cases}$$

3. Properties of the central path. The notion of central path plays a fundamental role in the development of interior point methods for linear programming. In this section, we shall study the analytic properties of the central path in the context of semidefinite programming. These properties will be used in Section 5 where we perform convergence analysis of a predictor-corrector algorithm for SDP.

For linear programming (i.e., $A^{(i)}$'s and C are diagonal), it is known that the central path curve converges: $(X(\mu), Z(\mu)) \to (X^a, Z^a)$, as $\mu \to 0$, with (X^a, Z^a) being the analytic center of the primal and dual optimal solution sets $\mathcal{F}_P^*$ and $\mathcal{F}_D^*$ respectively ([10]). Another important property of the central path in the context of linear programming is that it never converges *tangentially* to the optimal face [13]. This means that for any point on the central path, the distance to the end of the central path is of the same order as the distance to the optimal face, viz. $O(\mu)$. The aim of this section is to establish a similar property of the central path for semidefinite programming. More specifically, we shall prove that

$$\|X(\mu) - X^a\| + \|Z(\mu) - Z^a\| = O(\mu).$$

We begin with the following lemma which shows that the set

$$\{(X(\mu), Z(\mu)) : 0 < \mu < 1\}$$

is bounded.

LEMMA 3.1. *For any $\mu > 0$ there holds*

$$\|X(\mu)\| + \|Z(\mu)\| = O(1 + \mu).$$

Proof. Since $A^{(i)} \bullet (X(\mu) - X(1)) = 0$ for $i = 1, ..., m$, and $Z(\mu) - Z(1) \in \text{Span}\{A^{(i)},\ i = 1, ..., m\}$, it follows that $(X(\mu) - X(1)) \perp (Z(\mu) - Z(1))$.. Using this property, we obtain

$$\begin{aligned} n\mu + n &= X(\mu) \bullet Z(\mu) + X(1) \bullet Z(1) \\ &= X(1) \bullet Z(\mu) + Z(1) \bullet X(\mu). \end{aligned}$$

Since $X(1) \succ 0$ and $Z(1) \succ 0$, we have

$$\|X(\mu)\| + \|Z(\mu)\| = O(X(1) \bullet Z(\mu) + Z(1) \bullet X(\mu)) = O(1 + \mu). \qquad \square$$

It follows from Lemma 3.1 that the central path has a limit point. We will now show that any limit point of the central path $\{(X(\mu), Z(\mu))\}$ is a strictly complementary optimal primal-dual pair.

LEMMA 3.2. *For any $\mu \in (0, 1)$ there holds*

$$X_B(\mu) = \Theta(1), \quad X_N(\mu) = \Theta(\mu),$$
$$Z_B(\mu) = \Theta(\mu), \quad Z_N(\mu) = \Theta(1).$$

Hence, any limit point of $\{(X(\mu), Z(\mu))\}$ as $\mu \to 0$ is a pair of strictly complementary primal-dual optimal solutions of (P) *and* (D).

Proof. Let $0 < \mu < 1$. For notational convenience, we will use X and Z to denote the matrices $X(\mu)$ and $Z(\mu)$. Let (X^*, Z^*) be the pair of strictly complementary primal-dual optimal solutions postulated by Assumption 2.2. Since $(X - X^*) \perp (Z - Z^*)$, we have

$$\begin{aligned} 0 &= (X - X^*) \bullet (Z - Z^*) \\ &= X \bullet Z - X^* \bullet Z - X \bullet Z^* \\ &= \operatorname{tr}\,(\mu I - X^* Z - X Z^*) \\ &= n\mu - \sum_{i=1}^{K} \lambda_i Z_{ii} - \sum_{i=K+1}^{n} \lambda_i X_{ii}, \end{aligned}$$

where the last step follows from (2.1). Since $\lambda_i > 0$ for all i and $X_{ii} \geq 0$ and $Z_{ii} \geq 0$ (by the positive semidefiniteness of X and Z), we obtain

$$\begin{cases} Z_{ii} = O(\mu), & i = 1, ..., K, \\ X_{ii} = O(\mu), & i = K+1, ..., n. \end{cases}$$

Since $X \succeq 0$, $Z \succeq 0$, it follows that

$$X_N = O(\mu), \quad Z_B = O(\mu). \tag{3.1}$$

From $X \succ 0$ and $Z \succ 0$ we obtain

$$X_N - X_U^T X_B^{-1} X_U \succ 0, \quad Z_B - Z_U Z_N^{-1} Z_U^T \succ 0.$$

Now consider the identities

$$\begin{aligned} \log\det X &= \log\det X_B + \log\det(X_N - X_U^T X_B^{-1} X_U), \\ \log\det Z &= \log\det Z_N + \log\det(Z_B - Z_U Z_N^{-1} Z_U^T). \end{aligned}$$

Since $\det X \det Z = \det(\mu I) = \mu^n$, it follows that $\log\det X + \log\det Z = n \log \mu$ and

$$\begin{aligned} 0 = \log\det X_B &+ \log\det\left(\frac{1}{\mu}(X_N - X_U^T X_B^{-1} X_U)\right) \\ &+ \log\det Z_N + \log\det\left(\frac{1}{\mu}(Z_B - Z_U Z_N^{-1} Z_U^T)\right). \end{aligned}$$

By the estimates (3.1) and using Lemma 3.1, we see that

$$X_B = O(1), \quad \frac{1}{\mu}(X_N - X_U^T X_B^{-1} X_U) = O(1),$$

$$Z_N = O(1), \quad \frac{1}{\mu}(Z_B - Z_U Z_N^{-1} Z_U^T) = O(1).$$

Therefore each of the four logarithm terms in the preceding equation are bounded from above as $\mu \to 0$. Since these four terms sum to zero, we must have

$$\begin{aligned} X_B &= \Theta(1), \qquad \frac{1}{\mu}(X_N - X_U^T X_B^{-1} X_U) = \Theta(1), \\ Z_N &= \Theta(1), \qquad \frac{1}{\mu}(Z_B - Z_U Z_N^{-1} Z_U^T) = \Theta(1). \end{aligned}$$

Together with (3.1), this implies

$$X_N = \Theta(\mu), \quad Z_B = \Theta(\mu).$$

This completes the proof of the lemma. □

Lemma 3.2 provides a precise result on the order of the eigenvalues of $X_B(\mu)$, $X_N(\mu)$, $Z_B(\mu)$ and $Z_N(\mu)$. We will now prove a preliminary result on the order of the off-diagonal blocks $X_U(\mu)$ and $Z_U(\mu)$.

LEMMA 3.3. *For $\mu \in (0,1)$, there holds*

$$\begin{aligned} &\|X_U(\mu)\| = \Theta(1)\,\|Z_U(\mu)\|, \\ &-X_U(\mu) \bullet Z_U(\mu) = \Theta(1)\,\|X_U(\mu)\|^2, \\ &\|X_U(\mu)\| = o(\sqrt{\mu}), \quad \|Z_U(\mu)\| = o(\sqrt{\mu}), \qquad \text{as } \mu \to 0. \end{aligned} \tag{3.2}$$

Proof. By the central path definition, we have

$$\mu I = \begin{bmatrix} X_B(\mu) & X_U(\mu) \\ X_U(\mu)^T & X_N(\mu) \end{bmatrix} \begin{bmatrix} Z_B(\mu) & Z_U(\mu) \\ Z_U(\mu)^T & Z_N(\mu) \end{bmatrix}.$$

Expanding the right-hand side and comparing the upper-right corner of the above identity, we have

$$0 = X_B(\mu) Z_U(\mu) + X_U(\mu) Z_N(\mu), \tag{3.3}$$

or equivalently,

$$Z_U(\mu) = -X_B(\mu)^{-1} X_U(\mu) Z_N(\mu). \tag{3.4}$$

Using $X_B(\mu) = \Theta(1)$ and $Z_N(\mu) = \Theta(1)$ (see Lemma 3.2), this implies that

$$\|Z_U(\mu)\| = \Theta(1)\,\|X_U(\mu)\|.$$

This proves the first part of the lemma.

We now prove (3.2). Pre-multiplying both sides of (3.4) by the matrix $X_U(\mu)^T$ yields

$$X_U(\mu)^T Z_U(\mu) = -X_U(\mu)^T X_B(\mu)^{-1} X_U(\mu) Z_N(\mu).$$

Now taking the trace of the above matrices, we obtain

$$\begin{aligned} X_U(\mu) \bullet Z_U(\mu) &= -\operatorname{tr} X_U(\mu)^T X_B(\mu)^{-1} X_U(\mu) Z_N(\mu) \\ &= -\operatorname{tr} Z_N(\mu)^{1/2} X_U(\mu)^T X_B(\mu)^{-1} X_U(\mu) Z_N(\mu)^{1/2} \\ &= -\Theta(1)\,\|X_U(\mu)\|^2, \end{aligned}$$

where we used the fact that $X_B(\mu) = \Theta(1)$ and $Z_N(\mu) = \Theta(1)$, as shown in Lemma 3.2. This establishes (3.2).

It remains to prove the last part of the lemma. We consider an arbitrary convergent sequence $\{(X(\mu_k), Z(\mu_k)) : k = 1, 2, ...\}$ on the central path with $\mu_k \to 0$; its limit is denoted by X^*, Z^*, so that

$$X(\mu_k) - X^* = o(1), \quad Z(\mu_k) - Z^* = o(1). \tag{3.5}$$

By Lemma 3.2, we have $X_N^* = 0$, $X_U^* = Z_U^* = 0$ and $Z_B^* = 0$. Since $(X(\mu_k) - X^*) \perp (Z(\mu_k) - Z^*)$, we have

$$0 = (X_B(\mu_k) - X_B^*) \bullet Z_B(\mu_k) + 2X_U(\mu_k) \bullet Z_U(\mu_k) + X_N(\mu_k) \bullet (Z_N(\mu_k) - Z_N^*). \tag{3.6}$$

Using (3.2) and (3.6), we obtain

$$\begin{aligned} \|X_U(\mu_k)\|^2 &= -\Theta(1)\,(X_U(\mu_k) \bullet Z_U(\mu_k)) \\ &= \Theta(1)\,((X_B(\mu_k) - X_B^*) \bullet Z_B(\mu_k) + X_N(\mu_k) \bullet (Z_N(\mu_k) - Z_N^*)) \\ &= o(\mu_k), \end{aligned} \tag{3.7}$$

where in the last step we used (3.5) and of Lemma 3.2. This implies that $\|X_U(\mu)\|^2 = o(\mu)$ holds true on the entire central path curve, for otherwise there would exist a convergent subsequence $\{(X(\mu_k), Z(\mu_k)) : k = 1, 2, ...\}$ for which

$$\liminf_{k\to\infty} \frac{\|X_U(\mu_k)\|^2}{\mu_k} > 0,$$

contradicting (3.7). The proof is complete.. □

We now use Lemma 3.2 and Lemma 3.3 to prove the convergence of the central path $\{(X(\mu), Z(\mu)) : \mu > 0\}$ to the analytic center (X^a, Z^a), and to estimate the rate at which it converges to this limit.

LEMMA 3.4. *The primal-dual central path $\{(X(\mu), Z(\mu)) : \mu > 0\}$ converges to the analytic centers (X^a, Z^a) of $\mathcal{F}_P^*$ and $\mathcal{F}_D^*$ respectively. Moreover, if we let*

$$\epsilon(\mu) := \frac{\|X_U(\mu)\|}{\sqrt{\mu}},$$

then

$$\|X_B(\mu) - X_B^a\| = O((\epsilon(\mu) + \sqrt{\mu})^2), \quad \|Z_N(\mu) - Z_N^a\| = O((\epsilon(\mu) + \sqrt{\mu})^2).$$

Proof. Suppose $0 < \mu < 1$. By expanding $X(\mu)Z(\mu) = \mu I$ and comparing the upper-left block, we obtain

$$\mu I_B = X_B(\mu)Z_B(\mu) + X_U(\mu)Z_U(\mu)^T.$$

Pre-multiplying both sides with $(\mu X_B(\mu))^{-1}$ yields

$$X_B(\mu)^{-1} = \frac{1}{\mu}Z_B(\mu) + \frac{1}{\mu}X_B(\mu)^{-1}X_U(\mu)Z_U(\mu)^T. \tag{3.8}$$

Let $\mathcal{J}$ be an index set of minimal cardinality such that

$$\mathrm{Span}\{A_B^{(i)} : i \in \mathcal{J}\} = \mathrm{Span}\{A_B^{(i)} : i = 1, 2, \ldots, m\}.$$

Since $Z_B^* = 0$, it follows from dual feasibility and (3.8) that

$$\begin{aligned} \frac{1}{\mu}Z_B(\mu) &= \sum_{i\in\mathcal{J}} \nu_i(\mu)A_B^{(i)}, \qquad \text{for some scalars } \nu_i(\mu) \\ &= X_B(\mu)^{-1} - \frac{1}{\mu}X_B(\mu)^{-1}X_U(\mu)Z_U(\mu)^T. \end{aligned} \tag{3.9}$$

From Lemma 3.2, we know that $Z_B(\mu)/\mu = \Theta(1)$. As the matrices $A_B^{(i)}$, $i \in \mathcal{J}$ are linearly independent, this implies that the sequences $\{\nu_i(\mu) : \mu \in (0,1)\}$, $i \in \mathcal{J}$, are bounded. Moreover, we have from Lemma 3.3 that

$$\lim_{\mu \to 0} \frac{1}{\mu} X_B(\mu)^{-1} X_U(\mu) Z_U(\mu)^T = 0.$$

Hence, any limit X^*, ν_i^*, $i \in \mathcal{J}$ for $\mu \to 0$ satisfies the following nonlinear system of equations:

$$\text{(3.10)} \qquad \begin{cases} X_B^{-1} - \sum_{i \in \mathcal{J}} \nu_i A_B^{(i)} = 0, \\ A_B^{(i)} \bullet X_B = b_i, \qquad i \in \mathcal{J}. \end{cases}$$

Moreover, since $Z_B(\mu)/\mu = \Theta(1)$ and $X_B(\mu) = \Theta(1)$ for $\mu \in (0,1)$, we have

$$\sum_{i \in \mathcal{J}} \nu_i^* A_B^{(i)} \succ 0, \qquad X_B^* \succ 0.$$

By (2.3), this means that $X^* = X^a$, the analytic center of $\mathcal{F}_P^*$ and hence

$$X(\mu) - X^a = o(1), \quad \text{as } \mu \to 0.$$

Using the linear independence of the matrices $A_B^{(i)}$, $i \in \mathcal{J}$ and using the fact that X_B^a is positive definite, it can be checked that the Jacobian (with respect to the variables X_B and ν_i, $i \in \mathcal{J}$) of the nonlinear system (3.10) is nonsingular at the solution X_B^a, ν_i^*, $i \in \mathcal{J}$. Hence we can apply the classical inverse function theorem to the above nonlinear system at the point: $X_B = X_B^a$, $\nu_i = \nu_i^*$, $i \in \mathcal{J}$, to obtain

$$\text{(3.11)} \; \|X_B(\mu) - X_B^a\| = O\left(\|X_B(\mu)^{-1} - \sum_{i \in \mathcal{J}} \nu_i(\mu) A_B^{(i)}\| + \sum_{i \in \mathcal{J}} |A_B^{(i)} \bullet X_B(\mu) - b_i| \right).$$

By (3.9) and the definition of $\epsilon(\mu)$, we obtain from Lemma 3.3

$$\left\| X_B(\mu)^{-1} - \sum_{i \in \mathcal{J}} \nu_i(\mu) A_B^{(i)} \right\| = \left\| \frac{1}{\mu} X_B(\mu)^{-1} X_U(\mu) Z_U(\mu)^T \right\| = O(\epsilon(\mu)^2).$$

Also we have from $X(\mu) \in \mathcal{F}_P$

$$\begin{aligned} \left| A_B^{(i)} \bullet X_B(\mu) - b_i \right| &= \left| 2A_U^{(i)} \bullet X_U(\mu) + A_N^{(i)} \bullet X_N(\mu) \right| \\ &= O\left(\epsilon(\mu)\sqrt{\mu} + \mu\right), \qquad \text{for } i \in \mathcal{J}. \end{aligned}$$

Substituting the above two bounds into (3.11) yields

$$\|X_B(\mu) - X_B^a\| = O((\epsilon(\mu) + \sqrt{\mu})^2).$$

It can be shown by an analogous argument that

$$\|Z_N(\mu) - Z_N^a\| = O((\epsilon(\mu) + \sqrt{\mu})^2).$$

The proof is complete. $\square$

Lemma 3.4 only provides a rough sketch of the convergence behavior of the central path as $\mu \to 0$. Our goal is to characterize this convergence behavior more precisely.

THEOREM 3.5. *Let $\mu \in (0,1)$. There holds*

$$X_B(\mu) = \Theta(1), \quad Z_N(\mu) = \Theta(1), \quad X_N(\mu) = \Theta(\mu), \quad Z_B(\mu) = \Theta(\mu), \tag{3.12}$$

and

$$\|X(\mu) - X^a\| = O(\mu), \quad \|Z(\mu) - Z^a\| = O(\mu). \tag{3.13}$$

Proof. The estimate (3.12) is already known from Lemma 3.2, so we only need to prove (3.13). By Lemma 3.3 and Lemma 3.4, it is sufficient to show that

$$\|X_U(\mu)\| = O(\mu).$$

Suppose to the contrary that there exists a sequence

$$\{(X(\mu_k), Z(\mu_k)) : \ k = 1, 2, ...\} \tag{3.14}$$

with $\|X_U(\mu_k)\| > 0$ for all k and

$$\lim_{k\to\infty} \frac{\mu_k}{\|X_U(\mu_k)\|} = 0. \tag{3.15}$$

With the notation of Lemma 3.4, the condition (3.15) implies

$$\epsilon(\mu_k) + \sqrt{\mu_k} \ \sim \ \epsilon(\mu_k) = \frac{\|X_U(\mu_k)\|}{\sqrt{\mu_k}}. \tag{3.16}$$

By virtue of Lemma 3.2, we can choose the subsequence (3.14) such that

$$\lim_{k\to\infty} \frac{Z_B(\mu_k)}{\mu_k}$$

exists, and using Lemma 3.4 and relation (3.16), we can also assume the existence of

$$\Delta_B^x(\infty) := \lim_{k\to\infty} \frac{\mu_k}{\|X_U(\mu_k)\|^2}(X_B(\mu_k) - X_B^a).$$

From the existence of the above limits, we obtain

$$\lim_{k\to\infty} \frac{(X_B(\mu_k) - X_B^a) \bullet Z_B(\mu_k)}{\|X_U(\mu_k)\|^2} = \lim_{k\to\infty} \frac{\Delta_B^x(\infty) \bullet Z_B(\mu_k)}{\mu_k}. \tag{3.17}$$

Notice that the hypothesis (3.15) implies that

$$\lim_{k\to\infty} \frac{\mu_k}{\|X_U(\mu_k)\|^2}(X(\mu_k) - X^a) = \begin{bmatrix} \Delta_B^x(\infty) & 0 \\ 0 & 0 \end{bmatrix}.$$

Using also $Z_B^a = 0$, we thus obtain for any $k = 1, 2, \ldots$ that

$$\begin{aligned} \frac{\Delta_B^x(\infty) \bullet Z_B(\mu_k)}{\mu_k} &= \frac{\Delta_B^x(\infty) \bullet (Z_B(\mu_k) - Z_B^a)}{\mu_k} \\ &= \lim_{j\to\infty} \frac{\mu_j}{\mu_k \|X_U(\mu_j)\|^2}((X(\mu_j) - X^a) \bullet (Z(\mu_k) - Z^a)) \\ &= 0, \end{aligned}$$

where the last step is due to the orthogonality condition $(X(\mu_j)-X^a) \perp (Z(\mu_k)-Z^a)$ for all j and k. Therefore,

$$0 = \lim_{k\to\infty} \frac{\Delta_B^x(\infty) \bullet Z_B(\mu_k)}{\mu_k} = \lim_{k\to\infty} \frac{(X_B(\mu_k) - X_B^a) \bullet Z_B(\mu_k)}{\|X_U(\mu_k)\|^2}, \tag{3.18}$$

where we used (3.17). Analogously, it can be shown that

$$\lim_{k\to\infty} \frac{X_N(\mu_k) \bullet (Z_N(\mu_k) - Z_N^a)}{\|X_U(\mu_k)\|^2} = 0. \tag{3.19}$$

Since $(X(\mu_k) - X^a) \perp (Z(\mu_k) - Z^a)$, we have from (3.18) and (3.19) that

$$\begin{aligned} 0 &= \lim_{k\to\infty} \frac{(X(\mu_k) - X^a) \bullet (Z(\mu_k) - Z^a)}{\|X_U(\mu_k)\|^2} \\ &= \lim_{k\to\infty} 2\frac{X_U(\mu_k) \bullet Z_U(\mu_k)}{\|X_U(\mu_k)\|^2}, \end{aligned}$$

which clearly contradicts (3.2). The proof is complete. □

Theorem 3.5 characterizes completely the limiting behavior of the primal-dual central path as $\mu \to 0$. We point out that this limiting behavior was well understood in the context of linear programming and the monotone horizontal linear complementarity problem, see Güler [3] and Monteiro and Tsuchiya [13] respectively. Notice that under a *Nondegeneracy Assumption* (the reader is refered to Alizadeh, Haeberly and Overton [2] or Kojima, Shida and Shindoh [7] for a discussion on this concept), i.e. the Jacobian of the nonlinear system

$$\begin{cases} A^{(i)} \bullet X = b_i, \qquad i = 1, 2, \ldots, m, \\ \sum_{i=1}^m y_i A^{(i)} + Z = C, \\ XZ + ZX = 0, \\ X \in \mathcal{S},\ Z \in \mathcal{S} \end{cases}$$

is nonsingular at (X^a, y^a, Z^a), the estimates (3.13) follow immediately from the application of the classical inverse function theorem. Thus, the real contribution of Theorem 3.5 lies in establishing these estimates in the absence of the nondegeneracy assumption.

It is known that in the case of linear programming the proof of quadratic convergence of predictor-corrector interior point algorithms required an error bound result of Hoffman. This error bound states that the distance from any vector $x \in \Re^n$ to a polyhedral set $\mathcal{P} := \{x : Ax \le a\}$ can be bounded in terms of the "amount of constraint violation" at x, namely $\|[Ax - a]_+\|$, where $[\cdot]_+$ denotes the positive part of a vector. More precisely, Hoffman's error bound ([5]) states that there exists some constant $\tau > 0$ such that

$$\text{dist}(x, \mathcal{P}) \le \tau \|[Ax - a]_+\|, \qquad \forall x \in \Re^n.$$

Unfortunately, this error bound no longer holds for linear systems over the cone of positive semidefinite matrices (see the example below). In fact, much of the difficulty in the local analysis of interior point algorithms for SDP can be attributed to this lack of an analog of Hoffman's error bound result (see the analysis of [7, 16]). Specifically, without such an error bound result, it is difficult to estimate the distance from the current iterates to the optimal solution set. In essence, what we have established in Theorem 3.5 is an error bound result along the central path. In other words, although a Hoffman type error bound cannot hold over the entire feasible set of (P), it nevertheless still holds true on the restricted region "near the central path". One consequence of this restriction to the central path is that we will need to require the iterates to stay "sufficiently close" to the central path in order to establish the superlinear convergence of the algorithm. Such a requirement on the iterates was called "tangential convergence to the optimal solution set" by Kojima *et. al.* [7]. Notice that the analysis in this reference required the additional nondegeneracy assumption to establish their superlinear convergence result. In contrast, this assumption is no longer needed in our analysis because Theorem 3.5 holds without the nondegeneracy assumption.

Example. Consider the following linear system over the cone of positive semidefinite matrices in $\Re^{2\times 2}$:

$$X_{11} = 0, \quad X_{22} = 1, \quad X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \succeq 0.$$

Clearly, there is exactly one solution X^* to the above linear system, namely

$$X^* := \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

For each $\epsilon > 0$, consider the matrix

$$X(\epsilon) := \begin{bmatrix} \epsilon^2 & \epsilon \\ \epsilon & 1 \end{bmatrix}.$$

Clearly, $X(\epsilon) \succeq 0$. The amount of constraint violation is equal to ϵ^2. However, the distance $\|X(\epsilon) - X^*\|_F = \Theta(\epsilon)$. Thus, there cannot exist some fixed $\tau > 0$ such that $\|X(\epsilon) - X^*\| \le \tau\epsilon^2$, for all $\epsilon > 0$.

4. A polynomial predictor-corrector algorithm. In [14], Nesterov and Todd developed a theoretical foundation for interior point methods for self-scaled cones. This framework was then used in [15] to generalize primal-dual interior point methods from linear programming to conic convex programming over self-scaled cones. The results of Nesterov and Todd can be specialized to semidefinite programming, since the cone $\mathcal{S}_+$ of positive semi-definite matrices is self-scaled. In fact, the main results in Sturm and Zhang [17] can be viewed as such a specialization. However, unlike Nesterov and Todd [14, 15], Sturm and Zhang [17] adopt the V-space framework, thus generalizing the approach of Kojima et al. [6]. We begin by summarizing some of the results on V-space path following for SDP that were obtained in [17].

Let $(X, Z) \in \mathcal{F}_P \times \mathcal{F}_D$ with $X \succ 0$, $Z \succ 0$. Then, there exists a unique positive definite matrix $D \in \mathcal{S}_{++}$ such that ([17, Lemma 2.1])

$$X = DZD. \tag{4.1}$$

Let L be such that

$$LL^T = D, \tag{4.2}$$

and let $V := L^T ZL$. It follows that

$$V = L^{-1}XL^{-T} = L^T ZL.$$

The quantity

$$\delta(X,Z) := \left\| I - \frac{1}{\mu} L^{-1}XZL \right\|_F$$

serves as a centrality measure, with $\mu := X \bullet Z/n$. Considering the identity

$$\delta(X,Z) = \left\| I - \frac{1}{\mu} L^{-1}XZL \right\|_F = \left\| I - \frac{1}{\mu} X^{1/2}ZX^{1/2} \right\|_F,$$

it becomes apparent that $\delta(X,Z)$ is the same centrality measure used in many other papers, including [8] and [12]. It is easy to see that the central path is the set of solutions (X,Z) with $\delta(X,Z) = 0$ or, equivalently, those solutions for which $V = \sqrt{\mu} I$. Moreover, we have

$$(1 + \delta(X,Z))I \succeq \frac{1}{\mu} V^2 \succeq (1 - \delta(X,Z))I. \tag{4.3}$$

In V-space path following, we want to drive the V-iterates towards the origin by Newton's method, in such a way that the iterates reside in a cone around the identity matrix. Before stating the Newton equation, we need to introduce the linear space $\mathcal{A}(L)$,

$$\mathcal{A}(L) := \text{Span}\{L^T A^{(i)} L : \ i = 1, 2, \ldots, m\}$$

and its orthogonal complement in $\mathcal{S}$

$$\mathcal{A}^{\perp}(L) := \{\Delta X \in \mathcal{S} : \ (L^T A^{(i)} L) \bullet \Delta X = 0 \text{ for } i = 1, 2, \ldots, m\}.$$

A Newton direction for obtaining a $(\gamma\mu)$-center, for some $\gamma \in [0,1]$, is the solution $(\Delta X, \Delta Z)$ of the following system of linear equations ([17], equation (17) therein):

$$\begin{cases} \Delta X + D\Delta Z D = \gamma\mu Z^{-1} - X \\ \Delta X \in \mathcal{A}^{\perp}(I), \quad \Delta Z \in \mathcal{A}(I). \end{cases} \tag{4.4}$$

For $\gamma = 0$, we denote the solution of (4.4) by $(\Delta X^p, \Delta Z^p)$, the predictor direction. For $\gamma = 1$, the solution is denoted by $(\Delta X^c, \Delta Z^c)$, the corrector direction. If we let

$$\Delta\bar{X} := L^{-1}\Delta X L^{-T}, \qquad \Delta\bar{Z} := L^T \Delta Z L,$$

then we can rewrite (4.4) as

$$\begin{cases} \Delta\bar{X} + \Delta\bar{Z} = \gamma\mu V^{-1} - V \\ \Delta\bar{X} \in \mathcal{A}^{\perp}(L), \quad \Delta\bar{Z} \in \mathcal{A}(L). \end{cases}$$

It follows from orthogonality that

$$\left\| \Delta\bar{X}^p \right\|_F^2 + \left\| \Delta\bar{Z}^p \right\|_F^2 = \|V\|_F^2 = n\mu. \tag{4.5}$$

The corrector direction does not change the duality gap,

$$(X + \Delta X^c) \bullet (Z + \Delta Z^c) = X \bullet Z, \tag{4.6}$$

whereas

$$(X + t\Delta X^p) \bullet (Z + t\Delta Z^p) = (1-t) X \bullet Z, \tag{4.7}$$

for any $t \in \Re$, see equation (16) of [17].

Algorithm SDP(ϵ)

Given $(X^0, Z^0) \in \mathcal{F}_P \times \mathcal{F}_D$ with $\delta(X^0, Z^0) \le \frac{1}{4}$.
Parameter ϵ, $0 < \epsilon \le (X^0 \bullet Z^0)/n$ and positive integer r.
Let $k = 0$.
`REPEAT` (main iteration)
  Let $X = X^k$, $Z = Z^k$ and $\mu_k = X \bullet Z/n$.
  Predictor: compute $(\Delta X^p, \Delta Z^p)$ from (4.4) with $\gamma = 0$.
  Compute the largest step t_k such that for all $0 \le t \le t_k$ there holds
    $\delta(X + t\Delta X^p, Z + t\Delta Z^p) \le \min(1/2, ((1-t)\mu_k/\epsilon)^{2^{-r}})$.
  Let $X' := X + t_k \Delta X^p$, $Z' := Z + t_k \Delta Z^p$ and $\beta_k = \min(\frac{1}{4}, (1-t_k)\mu_k/\epsilon)$.
  Corrector:
  `FOR` $i = 1$ `to` r `DO`
    Let $X = X'$, $Z = Z'$.
    `IF` $\delta(X, Z) \le \beta_k$ `THEN` exit loop.
    Compute $(\Delta X^c, \Delta Z^c)$ from (4.4) with $\gamma = 1$.
    Set $X' = X + \Delta X^c$, $Z' = Z + \Delta Z^c$.
  `END FOR`
  $X^{k+1} = X'$, $Z^{k+1} = Z'$
  Set $k = k + 1$.
`UNTIL` convergence.

Interestingly, each corrector step reduces $\delta(\cdot,\cdot)$ at a quadratic rate as stated in the following result:

LEMMA 4.1. *If* $\delta(X, Z) \le \frac{1}{2}$ *then*

$$\delta(X + \Delta X^c, Z + \Delta Z^c) \le \delta(X, Z)^2.$$

Proof. It follows from Lemma 4.5 in [17] that

$$X + \Delta X^c \succ 0, \quad Z + \Delta Z^c \succ 0.$$

Hence, the desired result is an immediate consequence of Lemma 4.4 in [17]. ∎

Lemma 4.1 implies that for any $k \ge 1$, we have

$$\delta(X^k, Z^k) \le \beta_{k-1}. \tag{4.8}$$

Also, it follows from (4.6) and (4.7) that for any $k > 1$

$$\begin{aligned} \delta(X^k, Z^k) &\le \beta_{k-1} \le (1 - t_{k-1})\mu_{k-1}/\epsilon \\ &= \mu_k/\epsilon = O(\mu_k). \end{aligned} \tag{4.9}$$

Furthermore, if $\beta_k = 1/4$, then only one (instead of r) corrector step is needed to recenter the iterate (see [17]). In other words, the iterations of Algorithm SDP(ϵ) are identical to those of the primal-dual predictor-corrector algorithm of [17], for all k with

$$\frac{\mu_k}{\epsilon} \geq \frac{1}{4}.$$

We can therefore conclude from Theorem 5.2 in [17] that the algorithm yields $\mu_k \leq \frac{\epsilon}{4}$ for all $k \geq \Gamma\sqrt{n}\log(\mu_0/\epsilon)$, where Γ is a universal constant, independent of the problem data. Thus, we have the following polynomial complexity result.

THEOREM 4.2. *For each $0 < \epsilon < (X_0 \bullet Z_0)/n$, Algorithm SDP($\epsilon$) will generate an iterate $(X^k, Z^k) \in \mathcal{F}_P \times \mathcal{F}_D$ with $(X^k \bullet Z^k)/n \leq \epsilon/4$ in at most $\Gamma\sqrt{n}\log(\mu_0/\epsilon)$ predictor-corrector steps, where Γ is a constant that is independent of the problem data.*

In addition to having polynomial complexity, Algorithm SDP(ϵ) also possesses a superlinear rate of convergence. We prove this in the next section.

5. Convergence analysis. We begin by establishing the global convergence of Algorithm SDP(ϵ). Notice that Algorithm SDP(ϵ) chooses the predictor step length t_k to be the largest step such that for all $0 \leq t \leq t_k$ there holds

$$\delta(X + t\Delta X^p, Z + t\Delta Z^p) \leq \min\left(\frac{1}{2}, ((1-t)\mu/\epsilon)^{2^{-r}}\right). \tag{5.1}$$

It was shown in [17] (equation (21) therein) that

$$(1-t)\delta(X + t\Delta X^p, Z + t\Delta Z^p) \leq (1-t)\delta(X, Z) + t^2 \left\|\Delta\bar{X}^p\Delta\bar{Z}^p\right\|_F/\mu. \tag{5.2}$$

Using (4.5), we thus obtain for $0 \leq t < 1$ that

$$\delta(X + t\Delta X^p, Z + t\Delta Z^p) \leq \delta(X, Z) + \frac{nt^2}{2(1-t)}. \tag{5.3}$$

Combining (5.1) and (5.3), we can easily establish the global convergence of Algorithm SDP(ϵ).

THEOREM 5.1. *There holds*

$$\lim_{k\to\infty} \mu_k = 0,$$

i.e. Algorithm SDP(ϵ) is globally convergent.

Proof. Due to (4.7), the sequence $\mu_0,\ \mu_1, \ldots$ is a monotonically decreasing sequence. Hence, $\lim_{k\to\infty}\mu_k$ exists. Suppose contrary to the statement of the lemma that

$$\mu_\infty = \lim_{k\to\infty}\mu_k, \quad \mu_\infty > 0. \tag{5.4}$$

Consider $k \geq \Gamma\sqrt{n}\log(\mu_0/\epsilon)$. From Theorem 4.2, we have

$$\mu_k \leq \epsilon/4 \tag{5.5}$$

and hence, using (4.8),

$$\delta(X^k, Z^k) \leq \beta_{k-1} = \min(\frac{1}{4}, \frac{\mu_k}{\epsilon}) = \frac{\mu_k}{\epsilon}. \tag{5.6}$$

Now consider a step length $0 \le t \le 0.5\sqrt{\mu_k/(n\epsilon)}$ and note from (5.5) that

$$1 - t \ge \frac{3}{4}.$$

We obtain from (5.3) and (5.6) that

$$\begin{aligned} \delta(X^k + t(\Delta X^p)^k, Z^k + t(\Delta Z^p)^k) &\le \delta(X^k, Z^k) + \frac{nt^2}{2(1-t)} \\ &\le \frac{\mu_k}{\epsilon} + \frac{\mu_k}{6\epsilon} \\ &\le \frac{7}{6} \cdot \frac{4}{3} \cdot \frac{(1-t)\mu_k}{\epsilon} \\ &< \left(\frac{(1-t)\mu_k}{\epsilon} \right)^{2^{-r}}, \end{aligned}$$

where we used (5.5) and the fact that $r \ge 1$. By definition of t_k, this implies that

$$t_k \ge \frac{1}{2}\sqrt{\frac{\mu_k}{n\epsilon}} \ge \frac{1}{2}\sqrt{\frac{\mu_\infty}{n\epsilon}} = \Theta(1).$$

This, together with (4.7), implies that $1 - \frac{\mu_{k+1}}{\mu_k} = \Theta(1)$, which contradicts (5.4). □

Next we proceed to establish the superlinear convergence of Algorithm SDP(ϵ). In light of (4.7), we only need to show that the predictor step length t_k approaches 1. Hence we are led to bound t_k from below. For this purpose, we note from (5.2) that, for $t \in (0, 1)$,

$$\delta(X + t\Delta X^p, Z + t\Delta Z^p) \le \delta(X, Z) + \frac{1}{1-t} \left\| \Delta \bar{X}^p \Delta \bar{Z}^p \right\|_F / \mu. \tag{5.7}$$

Thus, if we can properly bound $\left\| \Delta \bar{X}^p \Delta \bar{Z}^p \right\|_F$, then we will obtain a lower bound on the predictor step length t_k.

To begin, let us consider L_μ with

$$L_\mu L_\mu^T = \frac{1}{\sqrt{\mu}} X(\mu).$$

Remark that

$$\sqrt{\mu} I = L_\mu^{-1} X(\mu) L_\mu^{-T} = L_\mu^T Z(\mu) L_\mu.$$

Now define the predictor direction starting from the solution $(X(\mu), Z(\mu))$ on the central path as follows:

$$\begin{cases} \Delta \hat{X}^p(\mu) + \Delta \hat{Z}^p(\mu) = -\sqrt{\mu} I, \\ \Delta \hat{X}^p(\mu) \in \mathcal{A}^\perp(L_\mu), \quad \Delta \hat{Z}^p(\mu) \in \mathcal{A}(L_\mu). \end{cases}$$

Let $(\hat{X}^a, \hat{Z}^a)$ be the analytic center of the optimal solution set in the L_μ-transformed space,

$$\hat{X}^a := L_\mu^{-1} X^a L_\mu^{-T}, \quad \hat{Z}^a := L_\mu^T Z^a L_\mu.$$

We will show in Lemma 5.2 below that $\Delta\hat{X}^p(\mu)$ is close to the optimal step $\hat{X}^a - \sqrt{\mu}I$ for small μ. We will bound the difference between $\Delta\hat{X}^p(\mu)$ and $\Delta\bar{X}^p$ afterwards.

LEMMA 5.2. *There holds*

$$\left\|\sqrt{\mu}I + \Delta\hat{X}^p(\mu) - \hat{X}^a\right\| + \left\|\sqrt{\mu}I + \Delta\hat{Z}^p(\mu) - \hat{Z}^a\right\| = O(\mu^{3/2}).$$

Proof. Since

$$\hat{X}^a\hat{Z}^a = L_\mu^{-1}X^aZ^aL_\mu = 0,$$

it follows that

$$(\sqrt{\mu}I - \hat{X}^a)(\sqrt{\mu}I - \hat{Z}^a) = (\sqrt{\mu}I - \hat{Z}^a)(\sqrt{\mu}I - \hat{X}^a).$$

Therefore, the matrix $(\sqrt{\mu}I - \hat{X}^a)(\sqrt{\mu}I - \hat{Z}^a)$, or equivalently, the matrix

$$L_\mu^{-1}(X(\mu) - X^a)(Z(\mu) - Z^a)L_\mu,$$

is symmetric. By the property of the F-norm, we obtain

$$\begin{aligned}
\left\|(\sqrt{\mu}I - \hat{X}^a)(\sqrt{\mu}I - \hat{Z}^a)\right\|_F &= \left\|L_\mu^{-1}(X(\mu) - X^a)(Z(\mu) - Z^a)L_\mu\right\|_F \\
&= (\ \mathrm{tr}\ (X(\mu) - X^a)(Z(\mu) - Z^a) \\
&\quad\ (X(\mu) - X^a)(Z(\mu) - Z^a))^{1/2} \\
&= O(\mu^2),
\end{aligned} \tag{5.8}$$

where the last step follows from Theorem 3.5. Now since $\hat{X}^a\hat{Z}^a = 0$ and $\Delta\hat{X}^p(\mu) + \Delta\hat{Z}^p(\mu) = -\sqrt{\mu}I$, we have

$$\begin{aligned}
(\sqrt{\mu}I - \hat{X}^a)(\sqrt{\mu}I - \hat{Z}^a) &= \mu I - \sqrt{\mu}(\hat{X}^a + \hat{Z}^a) \\
&= \sqrt{\mu}(\sqrt{\mu}I + \Delta\hat{X}^p(\mu) - \hat{X}^a) \\
&\quad + \sqrt{\mu}(\sqrt{\mu}I + \Delta\hat{Z}^p(\mu) - \hat{Z}^a).
\end{aligned}$$

As

$$\sqrt{\mu}I + \Delta\hat{X}^p(\mu) - \hat{X}^a \in \mathcal{A}^\perp(L_\mu), \qquad \sqrt{\mu}I + \Delta\hat{Z}^p(\mu) - \hat{Z}^a \in \mathcal{A}(L_\mu),$$

it follows that

$$\begin{aligned}
&\left\|\sqrt{\mu}I + \Delta\hat{X}^p(\mu) - \hat{X}^a\right\|_F^2 + \left\|\sqrt{\mu}I + \Delta\hat{Z}^p(\mu) - \hat{Z}^a\right\|_F^2 \\
&\quad = \frac{1}{\mu}\left\|(\hat{X}^a - \sqrt{\mu}I)(\hat{Z}^a - \sqrt{\mu}I)\right\|_F^2 = O(\mu^3),
\end{aligned}$$

where the last step is due to (5.8). This proves the lemma. ∎

Lemma 5.2 applies only to $(\Delta\hat{X}^p(\mu), \Delta\hat{Z}^p(\mu))$, namely the predictor directions for the points located exactly on the central path. What we need is a similar bound for $(\Delta\bar{X}^p, \Delta\bar{Z}^p)$ (obtained at points close to the central path). This leads us to bound the difference $\Delta\hat{X}^p(\mu) - \Delta\bar{X}^p$. Indeed, our next goal is to show (Lemma 5.6) that

$$\left\|\Delta\hat{X}^p(\mu) - \Delta\bar{X}^p\right\|_F = O(\sqrt{\mu}\delta(X, Z)).$$

We prove this bound by a sequence of lemmas. Let D be given by (4.1) and define

$$\bar{D} := L_\mu^{-1} D L_\mu^{-T},$$

so that $\bar{D} = I$ if $X = X(\mu)$ and $Z = Z(\mu)$. Choose L by

$$L := L_\mu \bar{D}^{1/2},$$

and notice that indeed $LL^T = D$, as stipulated by (4.2).

LEMMA 5.3. *Suppose $\delta(X, Z) \leq 1/2$. There holds*

$$\left\| L^{-1}(X(\mu) - X)L^{-T} \right\| + \left\| L^T(Z(\mu) - Z)L \right\| = O(\sqrt{\mu}\delta(X, Z)).$$

Proof. Let

$$\Delta_x(\mu) := L^{-1}(X(\mu) - X)L^{-T}, \qquad \Delta_z(\mu) := L^T(Z(\mu) - Z)L.$$

Clearly, $\Delta_x(\mu)$ and $\Delta_z(\mu)$ are symmetric and $\Delta_x(\mu) \perp \Delta_z(\mu)$. Let ρ denote the smallest eigenvalue of $\Delta_x(\mu) + \Delta_z(\mu)$, i.e.

$$\rho = \arg\max\{\bar{\rho} : \ \Delta_x(\mu) + \Delta_z(\mu) \succeq \bar{\rho} I\}.$$

Since $X \bullet Z = X(\mu) \bullet Z(\mu) = n\mu$, we have

$$\begin{aligned} \text{tr } (Z(X(\mu) - X) + X(Z(\mu) - Z)) &= \text{ tr } ((X(\mu) - X)Z + X(Z(\mu) - Z)) \\ &= - \text{ tr } ((X(\mu) - X)(Z(\mu) - Z)) - \text{ tr } XZ \\ &\quad + \text{tr } X(\mu)Z(\mu) \\ &= 0, \end{aligned}$$

where the last step follows from $(X(\mu) - X) \perp (Z(\mu) - Z)$. Recall that $V = L^{-1}XL^{-T} = L^T Z L$. Consider

$$\begin{aligned} \text{tr } (V(\Delta_x(\mu) + \Delta_z(\mu))) &= \text{ tr } (L^T Z(X(\mu) - X)L^{-T} + L^{-1}X(Z(\mu) - Z)L) \\ &= \text{ tr } (Z(X(\mu) - X) + X(Z(\mu) - Z)) \\ &= 0. \end{aligned}$$

By (4.3), the matrix V is symmetric positive definite and $V = \Theta(\sqrt{\mu})$. Diagonalize the symmetric matrix $\Delta_x(\mu) + \Delta_z(\mu) = Q^T \Lambda Q$ and consider

$$0 = \text{ tr } (V(\Delta_x(\mu) + \Delta_z(\mu))) = \text{ tr } (VQ^T\Lambda Q) = \text{ tr } (QVQ^T\Lambda).$$

Since $QVQ^T = \Theta(\sqrt{\mu})$, the diagonal entries of QVQ^T must be $\Theta(\sqrt{\mu})$. Therefore, the preceding equation implies that the diagonal matrix Λ must have a nonpositive eigenvalue and that its diagonal entries are of same order of magnitude. In other words, $\rho \leq 0$ and $\|\Lambda\| = O(|\rho|)$. This further implies

$$\|\Delta_x(\mu) + \Delta_z(\mu)\| = O(|\rho|). \tag{5.9}$$

By the definition of the central path, we have

$$\begin{aligned} \mu I &= (V + \Delta_x(\mu))(V + \Delta_z(\mu)) \\ &= \left(V + \frac{\Delta_x(\mu) + \Delta_z(\mu)}{2} + \frac{\Delta_x(\mu) - \Delta_z(\mu)}{2} \right) \times \\ &\quad \left(V + \frac{\Delta_x(\mu) + \Delta_z(\mu)}{2} - \frac{\Delta_x(\mu) - \Delta_z(\mu)}{2} \right). \end{aligned}$$

Since the left hand side matrix is symmetric, the skew-symmetric cross term must cancel when we expand the matrix product in the right hand side. It follows that

$$\mu I = \left(V + \frac{\Delta_x(\mu)+\Delta_z(\mu)}{2}\right)^2 - \frac{1}{4}\left(\Delta_x(\mu)-\Delta_z(\mu)\right)^2$$

and therefore,

$$V + \frac{\Delta_x(\mu)+\Delta_z(\mu)}{2} \succeq \sqrt{\mu} I.$$

Using (4.3), we obtain

$$|\rho| = O(\sqrt{\mu}\delta(X,Z)).$$

Combining this with (5.9) and using the fact that $\Delta_x(\mu) \perp \Delta_z(\mu)$, we have

$$\|\Delta_x(\mu)\| + \|\Delta_z(\mu)\| = O(|\rho|) = O(\sqrt{\mu}\delta(X,Z)).$$ ☐

LEMMA 5.4. *Suppose $\delta(X,Z) \le 1/2$. Then there holds*

$$\left\|\bar{D} - I\right\| = O(\delta(X,Z)).$$

Proof. Notice that

$$L_\mu^{-1} X L_\mu^{-T} = \sqrt{\mu} I + L_\mu^{-1}(X - X(\mu)) L_\mu^{-T} \tag{5.10}$$

and

$$L_\mu^T Z L_\mu = \sqrt{\mu} I + L_\mu^T (Z - Z(\mu)) L_\mu. \tag{5.11}$$

Now using

$$L_\mu^{-1} X L_\mu^{-T} = \bar{D}(L_\mu^T Z L_\mu)\bar{D},$$

we have, by pre- and post-multiplying (5.10) by $\bar{D}^{-1/2}$ and (5.11) by $\bar{D}^{1/2}$ and rearranging terms,

$$\sqrt{\mu}(\bar{D}^{-1} - \bar{D}) = L^{-1}(X(\mu) - X)L^{-T} + L^T(Z - Z(\mu))L.$$

Together with Lemma 5.3, this implies $\bar{D} = \Theta(1)$ and

$$\left\|\bar{D} - I\right\| = O(\delta(X,Z)).$$

The lemma is proved. ☐

Now, let

$$\Delta\hat{X}^p := \bar{D}^{1/2}\Delta\bar{X}^p\bar{D}^{1/2}, \quad \Delta\hat{Z}^p := \bar{D}^{-1/2}\Delta\bar{Z}^p\bar{D}^{-1/2}.$$

Notice that $(\Delta\hat{X}^p, \Delta\hat{Z}^p) \in \mathcal{A}^\perp(L_\mu) \times \mathcal{A}(L_\mu)$.

LEMMA 5.5. *Suppose $\delta(X,Z) \le 1/2$. We have*

$$\left\|\Delta\hat{X}^p - \Delta\bar{X}^p\right\| + \left\|\Delta\hat{Z}^p - \Delta\bar{Z}^p\right\| = O(\sqrt{\mu}\delta(X,Z)).$$

Proof. We have

$$\Delta\hat{X}^p = \bar{D}^{1/2}\Delta\bar{X}^p\bar{D}^{1/2} = \Delta\bar{X}^p + (\bar{D}^{1/2} - I)\Delta\bar{X}^p\bar{D}^{1/2} + \Delta\bar{X}^p(\bar{D}^{1/2} - I).$$

Now using Lemma 5.4 and (4.5), we see that

$$\left\|\Delta\hat{X}^p - \Delta\bar{X}^p\right\| = O(\sqrt{\mu}\delta(X,Z)).$$

It can be shown in an analogous way that

$$\left\|\Delta\hat{Z}^p - \Delta\bar{Z}^p\right\| = O(\sqrt{\mu}\delta(X,Z)).$$ □

Now we are ready to bound the difference between $\Delta\hat{X}^p(\mu)$ and $\Delta\bar{X}^p$.

LEMMA 5.6. *Suppose* $\delta(X,Z) \leq 1/2$. *We have*

$$\left\|\Delta\hat{X}^p(\mu) - \Delta\bar{X}^p\right\| + \left\|\Delta\hat{Z}^p(\mu) - \Delta\bar{Z}^p\right\| = O(\sqrt{\mu}\delta(X,Z)).$$

Proof. By definition of the predictor directions, we have

$$\Delta\hat{X}^p(\mu) + \Delta\hat{Z}^p(\mu) = -\sqrt{\mu}I$$

and

$$\Delta\bar{X}^p + \Delta\bar{Z}^p = -V.$$

Combining these two relations yields

$$\Delta\hat{X}^p(\mu) - \Delta\hat{X}^p + \Delta\hat{Z}^p(\mu) - \Delta\hat{Z}^p = V - \sqrt{\mu}I + \Delta\bar{X}^p - \Delta\hat{X}^p + \Delta\bar{Z}^p - \Delta\hat{Z}^p.$$

Now using Lemma 5.5 and using the fact that

$$\|V - \sqrt{\mu}I\| = \|(V + \sqrt{\mu}I)^{-1}(V^2 - \mu I)\| \leq \sqrt{\mu}\delta(X,Z),$$

we obtain

$$\left\|\Delta\hat{X}^p(\mu) - \Delta\hat{X}^p + \Delta\hat{Z}^p(\mu) - \Delta\hat{Z}^p\right\| = O(\sqrt{\mu}\delta(X,Z)).$$

Since $(\Delta\hat{X}^p(\mu) - \Delta\hat{X}^p) \perp (\Delta\hat{Z}^p(\mu) - \Delta\hat{Z}^p)$, the lemma follows from the above relation, after applying Lemma 5.5 once more. □

Combining (5.8), Lemma 5.2 and Lemma 5.6 we can now estimate the order of $\left\|\Delta\bar{X}^p\Delta\bar{Z}^p\right\|$, and hence, using (5.7), we can estimate the predictor step length t_k.

LEMMA 5.7. *We have*

$$\left\|\Delta\bar{X}^p\Delta\bar{Z}^p\right\| = O(\mu(\mu + \delta(X,Z))).$$

Proof. Combining Lemma 5.6 with Lemma 5.2, we have

$$\left\|\sqrt{\mu}I + \Delta\bar{X}^p - \hat{X}^a\right\| + \left\|\sqrt{\mu}I + \Delta\bar{Z}^p - \hat{Z}^a\right\| = O(\sqrt{\mu}(\mu + \delta(X,Z))), \tag{5.12}$$

so that, using (4.5),

$$\left\| \sqrt{\mu} I - \hat{X}^a \right\| + \left\| \sqrt{\mu} I - \hat{Z}^a \right\| = O(\sqrt{\mu}). \tag{5.13}$$

Moreover,

$$\begin{aligned}\Delta \bar{X}^p \Delta \bar{Z}^p &= (\hat{X}^a - \sqrt{\mu} I)(\hat{Z}^a - \sqrt{\mu} I) + (\hat{X}^a - \sqrt{\mu} I)(\sqrt{\mu} I + \Delta \bar{Z}^p - \hat{Z}^a) \\ &\quad + (\sqrt{\mu} I + \Delta \bar{X}^p - \hat{X}^a)\Delta \bar{Z}^p.\end{aligned}$$

Applying (5.8), (5.12), (5.13) and (4.5) to the above relation yields

$$\left\| \Delta \bar{X}^p \Delta \bar{Z}^p \right\| = O(\mu(\mu + \delta(X, Z))). \qquad \square$$

THEOREM 5.8. *The iterates (X^k, Z^k) generated by Algorithm SDP(ϵ) converge to (X^a, Z^a) superlinearly with order $2/(1 + 2^{-r})$. The duality gap μ^k converges to zero at the same rate.*

Proof. From (5.7) we see that for any $t \geq 0$ satisfying

$$\beta_{k-1} + \left\| \Delta \bar{X}^p \Delta \bar{Z}^p \right\|_F / \mu_k \leq (1 - t)((1 - t)\mu_k/\epsilon)^{2^{-r}},$$

there holds

$$\delta(X + t\Delta X^p, Z + t\Delta Z^p) \leq ((1 - t)\mu/\epsilon)^{2^{-r}}.$$

This implies using (4.9) and Lemma 5.7 that

$$\begin{aligned}(1 - t_k)^{1+2^{-r}} &\leq (\beta_{k-1} + \left\| \Delta \bar{X}^p \Delta \bar{Z}^p \right\|_F / \mu_k)(\mu_k/\epsilon)^{-2^{-r}} \\ &= O(\mu_k^{1-2^{-r}}),\end{aligned}$$

so that

$$\mu_{k+1} = (1 - t_k)\mu_k = O(\mu_k^{2/(1+2^{-r})}).$$

This shows that the duality gap converges to zero superlinearly with order $2/(1+2^{-r})$. It remains to prove that the iterates converge to the analytic center with the same order. Notice that

$$\begin{aligned}\|X^k - X(\mu_k)\|_F^2 &= \operatorname{tr}\,(X^k - X(\mu_k))L^{-T}(L^T L)L^{-1}(X^k - X(\mu_k)) \\ &\leq \|L^T L\| \cdot \operatorname{tr}\,(X^k - X(\mu_k))L^{-T} L^{-1}(X^k - X(\mu_k)) \\ &= \|L^T L\| \cdot \operatorname{tr}\, L^{-1}(X^k - X(\mu_k))(X^k - X(\mu_k))L^{-T} \\ &\leq \|L^T L\|^2 \cdot \|L^{-1}(X^k - X(\mu_k))L^{-T}\|_F^2.\end{aligned} \tag{5.14}$$

However, using the definition of the F-norm and applying Lemma 5.4,

$$\|L^T L\|_F = \|L L^T\|_F = \|L_{\mu_k} \bar{D} L_{\mu_k}^T\|_F = O(\|L_{\mu_k} L_{\mu_k}^T\|_F).$$

Recall that $L_{\mu_k} L_{\mu_k}^T = \frac{1}{\sqrt{\mu_k}} X(\mu_k)$ by definition, so that using Lemma 3.1,

$$\|L^T L\|_F = O\left(\frac{1}{\sqrt{\mu_k}}\right). \tag{5.15}$$

Combining (5.14) and (5.15) with Lemma 5.3, we have

$$\|X^k - X(\mu_k)\|_F = O(\frac{1}{\sqrt{\mu_k}}\|L^{-1}(X^k - X(\mu_k))L^{-T}\|_F) = O(\delta(X^k, Z^k)) = O(\mu_k).$$

Hence, we obtain from Theorem 3.5 that

$$\|X^k - X^a\|_F = O(\mu_k).$$

Similarly, it can be shown that

$$\|Z^k - Z^a\|_F = O(\mu_k).$$

This shows that the iterates converge to the analytic center R-superlinearly, with the same order as μ_k converges to zero. □

6. Conclusions. We have shown the global and superlinear convergence of the predictor-corrector algorithm SDP, assuming only the existence of a strictly complementary solution pair. The local convergence analysis is based on Theorem 3.5, which states that $\|X(\mu)-X^a\|+\|Z(\mu)-Z^a\| = O(\mu)$. By enforcing $\delta(X^k, Z^k) \to 0$, the iterates "inherit" this property of the central path. For the generalization of the Mizuno-Todd-Ye predictor-corrector algorithm in [17], we do not enforce $\delta(X^k, Z^k) \to 0$, and hence we cannot conclude superlinear convergence for it yet. In this respect, it will be interesting to study the asymptotic behavior of the corrector steps. Finally, it is likely that our line of argument can be applied to the infeasible primal-dual path following algorithms of Kojima-Shindoh-Hara [8] and Potra-Sheng [16].

Acknowledgment. We wish to thank the referees for providing careful comments which led to substantial improvement in the presentation of the paper.

REFERENCES

[1] F. Alizadeh, *Interior point methods in semidefinite programming with applications to combinatorial optimization problems*, SIAM J. Optim. 5 (1995), pp. 13–51.

[2] F. Alizadeh, J. A. Heaberly and M. Overton, *Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results*, Technical report, Computer Science Department, New York University, New York, 1996.

[3] O. Güler, *Limiting behavior of weighted central paths in linear programming*, Math. Programming 65 (1994), pp. 347–363.

[4] C. Helmberg, F. Rendl, R. J. Vanderbei and H. Wolkowicz, *An interior-point method for semidefinite programming*, SIAM J. Optim. 6 (1996), pp. 342–361.

[5] A. J. Hoffman, *On approximate solutions of systems of linear inequalities*, Journal of Research of the National Bureau of Standards 49 (1952), pp. 263–265.

[6] M. Kojima, N. Megiddo, T. Noma and A. Yoshise, *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems*, Springer-Verlag, Berlin, 1991.

[7] ———, M. Shida and S. Shindoh, *Local convergence of predictor-corrector infeasible-interior-point algorithms for SDPs and SDLCPs*, Research Reports on Information Sciences, B-306, Dept. of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, 1995.

[8] ———, S. Shindoh, and S. Hara, *Interior-point methods for the monotone linear complementarity problem in symmetric matrices*, SIAM J. Optim. 7 (1997), pp. 86–125.

[9] C. J. Lin and R. Saigal, *An infeasible start predictor corrector method for semi-definite linear programming*, Technical report, Dept. of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, USA, 1995.

[10] N. Megiddo, *Pathways to the optimal solution set in linear programming* in Progress in Mathematical Programming, Interior-Point and Related Methods, N. Megiddo ed., Springer-Verlag, New York, 1989, pp. 131–158.

[11] S. Mizuno, M. J. Todd and Y. Ye, *On Adaptive-Step Primal-Dual Interior-Point Algorithms for Linear Programming*, Math. Oper. Res. 18 (1993), pp. 964–981.

[12] R. D. C. Monteiro, *Primal-dual path following algorithms for semidefinite programming*, Technical report, School of Industrial and Systems Engineering, Georgia Tech, Atlanta, Georgia, U.S.A., 1995. To appear in SIAM J. Optim.

[13] ——— and T. Tsuchiya, *Limiting behavior of the derivatives of certain trajectories associated with a monotone horizontal linear complementarity problem*, Math. Oper. Res. 21 (1996), pp. 793–814.

[14] Y. Nesterov and M. J. Todd, *Self-scaled barriers and interior-point methods for convex programming*, Math. Oper. Res. 22 (1997), pp. 1–42.

[15] ——— and ———, *Primal-dual interior-point methods for self-scaled cones*, Technical report 1125, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York, 1995.

[16] F. A. Potra and R. Sheng, *A superlinearly convergent primal-dual infeasible-interior-point algorithm for semidefinite programming*, Report on Computational Mathematics No. 78, Dept. of Mathematics, The University of Iowa, Iowa City, USA, 1995.

[17] J. F. Sturm and S. Zhang, *Symmetric primal-dual path following algorithms for semidefinite programming*, Technical report 9554/A, Econometric Institute, Erasmus University Rotterdam, The Netherlands, 1995.

[18] L. Vandenberghe and S. Boyd, *A primal-dual potential reduction method for problems involving matrix inequalities*, Math. Programming 69 (1995), pp. 205–236.

[19] ——— and ———, *Semidefinite programming*, SIAM Rev. 38 (1996), pp. 49–95.

[20] Y. Ye and K. Anstreicher, *On quadratic and $O(\sqrt{n}L)$ convergence of a predictor-corrector algorithm for LCP*, Math. Programming 62 (1993), pp. 537–551.

[21] Y. Zhang, *On extending primal-dual interior-point algorithms from linear programming to semidefinite programming*, Technical report, Dept. of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, Maryland, U.S.A., 1995.