January 29, 1996

SUPERLINEAR CONVERGENCE OF A SYMMETRIC PRIMAL-DUAL PATH FOLLOWING ALGORITHM FOR SEMIDEFINITE PROGRAMMING[1]

Zhi-Quan Luo[2], Jos F. Sturm[3] and Shuzhong Zhang[3]

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 which 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 $\frac{2}{1+2^{-2r}}$. 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 classification: 90C25, 90C26, 90C60.

---

[1] The research of the first author is supported by the Natural Sciences and Engineering Research Council of Canada, Grant No. OPG0090391, and was performed during a research leave to the Econometric Institute, Erasmus University Rotterdam.

[2] Room 225/CRL, Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, L8S 4L7, Canada.

[3] Econometric Institute, Erasmus University Rotterdam, The Netherlands.

1 Introduction

Recently, there have been many interior point algorithms developed for semidefinite programming (SDP), see for example [1, 2, 5, 9, 11, 13, 17]. 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-Shida-Hara [5] and Nesterov-Todd [11] are based on the primal-dual scaling and they both can be viewed as extensions of the predictor-corrector method for linear programming [8]. It has been shown [4, 6, 11, 13, 17] 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 Vanderberghe and Boyd [15].

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 [16] 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-Shidoh [4] and Potra-Sheng [12] 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 [4] established the superlinear convergence of their algorithm [5] 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 complementary 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 [16]). 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 the reference [4], 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 the only assumptions of (2) and (3) (i.e., without the nondegeneracy assumption). In particular, we consider the primal-dual path following algorithm of Nesterov-Todd [11] (later discovered independently by Sturm and Zhang [13] using a V-space notion). In this paper we adopt the framework of [13] 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 $\frac{2}{1+2^{-2r}}$ 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 [13] 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 X and Y in $\Re^{n\times n}$, the standard inner product is defined by

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

where 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 "big O" and "small o" notation. In particular, $w(\mu) = O(\mu)$ with $\mu > 0$ means that there is a positive constant Γ, possibly dependent on problem data but independent of μ, such that $w(\mu) \le \Gamma\mu$ for all μ; $w(\mu) = o(\mu)$ for $\mu \to 0$ means that $\lim_{\mu\to 0} w(\mu)/\mu = 0$. Moreover, we write $w(\mu) = \Theta(\mu)$ whenever we have both $w(\mu) = O(\mu)$ and $\mu = O(w(\mu))$. For a positive definite matrix, we use "O" and "Θ" to denote the order of all its eigenvalues. Hence, for $W(\mu) \in \mathcal{S}_{++}$, the notation $W(\mu) = \Theta(\mu)$ signifies the existence of $\Gamma > 0$ such that

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

2 Problem formulation

A semidefinite programming (SDP) problem is given as

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

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

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

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

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

We make the following assumptions throughout this paper.

Assumption 1 *There exist positive definite solutions $X \in \mathcal{F}_P$ and $Z \in \mathcal{F}_D$ for* (P) *and* (D) *respectively.*

Assumption 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*

$$\begin{cases} X^* Z^* = 0, \\ X^* + Z^* \succ 0. \end{cases}$$

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, 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 [5, 14] for a proof of the existence and uniqueness of μ-centers. The *central path* of the problem (P) is the curve

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

To be consistent with the above definition of the central path, we define the analytic center of $\mathcal{F}_P^*$ as the unique solution X^a of the 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 unique solution Z^a of the system

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

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 ([7]). It is also known for linear programming that the central path does *not* approach (X^a, Z^a) tangentially to the optimal solution set, viz. it is shown in [10] that

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

In the following we shall extend these result to the semidefinite programs (P) and (D).

The following lemma 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. We have

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

where we used the property $(X(\mu) - X(1)) \perp (Z(\mu) - Z(1))$ in the second equality. Since $X(1) \succ 0$ and $Z(1) \succ 0$, we have

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

Q.E.D.

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*

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

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

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. Since $A^{(i)} \bullet (X - X^*) = 0$, $i = 1, ..., m$, and $Z - Z^* \in \mathrm{Span}\{A^{(i)},\ i = 1, ..., m\}$, it follows that $(X - X^*) \perp (Z - Z^*)$. Therefore, we have

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

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

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

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

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

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

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

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. **Q.E.D.**

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*

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

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,

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

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

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

This proves the first part of the lemma.

We now prove (3.2). Let $\{(X(\mu_k), Z(\mu_k)) : k = 1, 2, ...\}$ be an arbitrary convergent sequence of the central path with $\mu_k \to 0$. By Lemma 3.2, the limit of this sequence satisfies strict complementarity. Let (X^*, Z^*) denote this limit point so that

$$X^* = \lim_{k\to\infty} X(\mu_k), \quad Z^* = \lim_{k\to\infty} Z(\mu_k).$$

As before, we assume without loss of generality that X^* and Z^* are diagonal. In addition, since (3.2) holds trivially when $\|X_U(\mu_k)\| = 0$, we thus assume $\|X_U(\mu_k)\| > 0$ for all k.

First, we divide both sides of (3.3) by $\|X_U(\mu_k)\|$ and let $k \to \infty$ to obtain

$$0 = X_B^* Z_U^\infty + X_U^\infty Z_N^*,$$

where X_U^∞ and Z_U^∞ are defined by

$$X_U^\infty := \lim_{k\to\infty} \frac{X_U(\mu_k)}{\|X_U(\mu_k)\|}, \quad Z_U^\infty := \lim_{k\to\infty} \frac{Z_U(\mu_k)}{\|X_U(\mu_k)\|}. \tag{3.4}$$

(If the above limits do not exist, then we define X_U^∞ and Z_U^∞ to be any two limit points of the corresponding sequences.) Since X_B^* and Z_N^* are both positive diagonal matrices, it follows that the nonzero entries of the matrices X_U^∞, Z_U^∞ must have opposite signs. By $\|X_U^\infty\| = 1$, we conclude that

$$X_U^\infty \bullet Z_U^\infty < 0. \tag{3.5}$$

This establishes (3.2) along the sequence $\{(X(\mu_k), Z(\mu_k)) : k = 1, 2, ...\}$. Since this sequence is arbitrary, we see (3.2) holds.

It remains to establish the last part of the lemma. Once again, we consider an arbitrary convergent sequence $\{(X(\mu_k), Z(\mu_k)) : k = 1, 2, ...\}$ on the central path with $\mu_k \to 0$; we continue to use the same notation X^*, Z^*, X_U^∞, Z_U^∞ defined above. Since $\|Z_U(\mu_k)\| = \Theta(\|X_U(\mu_k)\|)$, we only need to show $\|X_U(\mu_k)\| = o(\sqrt{\mu_k})$. Assume this is not the case. Using Lemma 3.2 and passing onto a subsequence if necessary, we have $\|X_U(\mu_k)\|^2 = \Theta(\mu_k)$ for all k. Since $(X(\mu_k) - X^*) \perp (Z(\mu_k) - Z^*)$, we have

$$\begin{aligned} 0 &= (X_B(\mu_k) - X_B^*) \bullet Z_B(\mu_k) + 2X_U(\mu_k) \bullet Z_U(\mu_k) \\ &\quad + X_N(\mu_k) \bullet (Z_N(\mu_k) - Z_N^*). \end{aligned}$$

Dividing both sides of this equation by $\|X_U(\mu_k)\|^2$ and taking limit yields

$$\begin{aligned} 0 &= 2X_U^\infty \bullet Z_U^\infty \\ &\quad + \lim_{k\to\infty} \left((X_B(\mu_k) - X_B^*) \bullet \frac{Z_B(\mu_k)}{\|X_U(\mu_k)\|^2} + \frac{X_N(\mu_k)}{\|X_U(\mu_k)\|^2} \bullet (Z_N(\mu_k) - Z_N^*) \right). \end{aligned}$$

Since $\|X_U(\mu_k)\|^2 = \Theta(\mu_k)$ and $Z_B(\mu_k) = \Theta(\mu_k)$, $X_N(\mu_k) = \Theta(\mu_k)$ (cf. Lemma 3.2), it follows

$$\frac{Z_B(\mu_k)}{\|X_U(\mu_k)\|^2} = \Theta(1), \quad \frac{X_N(\mu_k)}{\|X_U(\mu_k)\|^2} = \Theta(1).$$

Therefore, the limit in the preceding equation equals zero, implying

$$X_U^\infty \bullet Z_U^\infty = 0.$$

But this contradicts (3.5), so we must have

$$\|X_U(\mu_k)\| = o(\sqrt{\mu_k}).$$

The proof is complete. **Q.E.D.**

We now use Lemma 3.2 and Lemma 3.3 to prove that the central path $\{(X(\mu), Z(\mu)) : \mu > 0\}$ converges to (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 $(\mu X_B)^{-1}$ yields

$$X_B^{-1} = \frac{1}{\mu}Z_B(\mu) + \frac{1}{\mu}X_B^{-1}X_U(\mu)Z_U(\mu)^T. \tag{3.6}$$

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

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

As $Z_B^* = 0$, it follows from the dual feasibility and (3.6) that

$$\begin{aligned} \frac{1}{\mu}Z_B(\mu) &= \sum_{i\in\mathcal{J}} \nu_i(\mu)A_B^{(i)}, \qquad \text{for some scalars } \nu_i(\mu) \\ &= X_B^{-1} - \frac{1}{\mu}X_B^{-1}X_U(\mu)Z_U(\mu)^T. \end{aligned} \tag{3.7}$$

Now consider the following nonlinear system of equations:

$$\begin{cases} X_B^{-1} - \sum_{i\in\mathcal{J}} \nu_i A_B^{(i)} = 0, \\ A_B^{(i)} \bullet X_B = b_i, \qquad i \in \mathcal{J}. \end{cases} \tag{3.8}$$

By (2.3), we know that X_B^a is a solution of (3.8) for some ν_i^a, $i \in \mathcal{J}$. 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 (3.8) is nonsingular at the solution X_B^a, ν_i^a, $i \in \mathcal{J}$. Hence we can apply the classical inverse function theorem to the above nonlinear system at the point: $X_B = X_B^a$, $\nu_i = \nu_i^a$, $i \in \mathcal{J}$, to obtain

$$\|X_B(\mu) - X_B^a\| = O\left(\|X_B^{-1} - \sum_{i\in\mathcal{J}} \nu_i A_B^{(i)}\| + \sum_{i\in\mathcal{J}} |A_B^{(i)} \bullet X_B(\mu) - b_i|\right). \tag{3.9}$$

By (3.7) we have

$$\left\|X_B^{-1} - \sum_{i\in\mathcal{J}} \nu_i A_B^{(i)}\right\| = \left\|\frac{1}{\mu} X_B^{-1} X_U(\mu) Z_U(\mu)^T\right\| = \Theta(\epsilon(\mu)^2) \tag{3.10}$$

and from $X(\mu) \in \mathcal{F}_P$ we obtain

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

Combining this with (3.9) and (3.10) yields

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

It can be shown with an analogous argument that

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

The proof is complete. **Q.E.D.**

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

$$X_B(\mu) = \Theta(1), \quad Z_N(\mu) = \Theta(1), \quad X_N(\mu) = \Theta(\mu), \quad Z_B(\mu) = \Theta(\mu), \tag{3.11}$$

and

$$\|X(\mu) - X^a\| = O(\mu), \quad \|Z(\mu) - Z^a\| = O(\mu). \tag{3.12}$$

Proof. The estimate (3.11) is already known from Lemma 3.2, so we only need to prove (3.12). 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

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

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

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

To simplify notations, we introduce

$$\Delta_B^x(\infty) := \lim_{k\to\infty} \frac{\mu_k}{\|X_U(\mu_k)\|^2}(X_B(\mu_k) - X_B^a). \tag{3.13}$$

(By virtue of Lemma 3.4, we can assume the above limit exists because otherwise we can always pass onto a convergent subsequence.) From Lemma 3.3 it follows that

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

Since for each $Z \in \mathcal{F}_D$ we have

$$(X(\mu_k) - X^a) \perp (Z - Z^a) \qquad \text{for } k = 1, 2, \ldots$$

it follows

$$\Delta_B^x(\infty) \bullet Z_B = \Delta_B^x(\infty) \bullet (Z_B - Z_B^a) = 0.$$

We know from Lemma 3.2 that $Z_B(\mu_k) = \Theta(\mu_k)$ so that the above relation implies

$$\lim_{k\to\infty} \frac{(X_B(\mu_k) - X_B^a) \bullet Z_B(\mu_k)}{\|X_U(\mu_k)\|^2} = 0. \tag{3.14}$$

Analogously, it can be shown that

$$\lim_{k\to\infty} \frac{X_N(\mu_k) \bullet (Z_N(\mu_k) - Z_N^a)}{\|X_U(\mu_k)\|^2} = 0. \tag{3.15}$$

As $(X(\mu_k) - X^a) \perp (Z(\mu_k) - Z^a)$, we have from (3.14) and (3.15) 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. **Q.E.D.**

Theorem 3.1 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 Megiddo [7] and Monteiro and Tsuchiya [10] respectively. Notice that under a *Nondegeneracy Assumption* (i.e., the Jacobian of the nonlinear system (2.2) is nonsingular at (X^a, Z^a)), the estimates (3.12) follow immediately from the application of the classical inverse function theorem. Thus, the real contribution of Theorem 3.1 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 ([3]) states that there exists some constant $\tau > 0$ such that

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

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 Hoffman's error bound result (see the analysis of [4, 12]). Specifically, without such 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 3.1 is an error bound result along the central path. In other words, although 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 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.* [4]. 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 3.1 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, $X(\epsilon) \geq 0$. 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 $\|X(\epsilon) - X^*\| \leq \tau\epsilon^2$, for all $\epsilon > 0$. Instead, we have in this case $\|X(\epsilon) - X^*\| \leq \tau|X_{11}(\epsilon)|^{1/2}$; that is, the error bound holds with an exponent of $1/2$.

4 A polynomial predictor-corrector algorithm

We begin by summarizing some of the results on V-space path following for SDP that were obtained in [13].

Let $(X, Z) \in \mathcal{F}_P \times \mathcal{F}_D$ with $X \succeq 0,\ Z \succeq 0$. Then, there exists a unique positive definite matrix $D \in \mathcal{S}_{++}$ such that ([13, Lemma 2.1])

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

Let L be such that

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

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 $\mu := X \bullet Z/n$. 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 $\mathcal{A}(L)$,

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

and its orthoplement in $\mathcal{S}$

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

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 ([13], equation (17)):

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

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

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

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 [13].

Algorithm SDP

Given $(X^0, Z^0) \in \mathcal{F}_P \times \mathcal{F}_D$ with $\delta(X^0, Z^0) \le \frac{1}{4}$.
Parameter $\epsilon \le (X^0 \bullet Z^0)/n$ 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
    $\delta(X + t\Delta X^p, Z + t\Delta Z^p) \le 1/2$
  and
    $\delta(X + t\Delta X^p, Z + t\Delta Z^p) \le ((1-t)\mu_k/\epsilon)^{2^{-r}}$.
  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 [13] 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 [13]. **Q.E.D.**

Also, it follows from (4.6), (4.7) and Lemma 4.1 that for any $k > 1$

$$\begin{aligned} \delta(X^k, Z^k) &\leq \beta_{k-1} \\ &\leq (1 - t_{k-1})\mu_{k-1}/\epsilon \\ &= \mu_k/\epsilon = O(\mu_k). \end{aligned} \tag{4.8}$$

Furthermore, if $\beta_k = 1/4$, then only one corrector step (i.e., $r = 1$) is needed to recenter the iterate (see [13]). In other words, the iterations of Algorithm SDP are identical to that of the primal-dual predictor-corrector algorithm of [13], for all k with

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

We can therefore conclude from Theorem 5.2 in [13] that the algorithm yields $\mu_{\bar{k}} \leq \frac{\epsilon}{4}$ for $\bar{k} = O(\sqrt{n}\log(\mu_0/\epsilon))$. Thus, we have the following polynomial complexity result.

Theorem 4.1 *For each $0 < \epsilon < (X_0 \bullet Z_0)/n$, Algorithm SDP will generate an iterate $(X^k, Z^k) \in \mathcal{F}_P \times \mathcal{F}_D$ with $X^k \bullet Z^k \leq \epsilon$ in at most $O(\sqrt{n}\log(\mu_0/\epsilon))$ predictor-corrector steps.*

In addition to having polynomial complexity, Algorithm SDP 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. Notice that Algorithm SDP 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 [13] (equation (21) therein) that

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

Combining (5.1) and (5.2), we can easily establish the global convergence of Algorithm SDP.

Theorem 5.1 *There holds*

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

i.e. Algorithm SDP is globally convergent.

Proof. Due to (4.7), $\mu^0, \mu^1, \ldots$ is a monotone decreasing sequence. Hence, the sequence has a limit. Suppose contrary to the statement of the lemma that

$$\mu_\infty = \lim_{k\to\infty} \mu_k, \quad \mu_\infty > 0. \tag{5.3}$$

Then, we obtain from (4.5), (5.1) and (5.2) that $t_k = \Theta(1)$. Together with (4.7) this implies that $1 - \frac{\mu_{k+1}}{\mu_k} = \Theta(1)$, which contradicts (5.3). **Q.E.D.**

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

$$\delta(X + t\Delta X^p, Z + t\Delta Z^p) \le \delta(X, Z) + \frac{1}{1-t}\left\|\Delta\bar{X}^p \Delta\bar{Z}^p\right\|_F/\mu. \tag{5.4}$$

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 5.1 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 5.1 *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

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

it follows

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

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 F-norm, we obtain

$$\begin{aligned} \left\| (\sqrt{\mu} I - \hat{X}^a)(\sqrt{\mu} I - \hat{Z}^a) \right\|_F &= \left\| L_\mu^{-1}(X(\mu) - X^a)(Z(\mu) - Z^a) L_\mu \right\|_F \\ &= \| (X(\mu) - X^a)(Z(\mu) - Z^a) \|_F \\ &= O(\mu^2), \end{aligned} \tag{5.5}$$

where the last step follows from Theorem 3.1. 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

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

As

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

it follows that

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

where last step is due to (5.5). This proves the lemma. **Q.E.D.**

Lemma 5.1 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 5.5) 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 $LL^{\mathrm{T}} = D$, as stipulated by (4.2).

Lemma 5.2 *Suppose $\delta(X,Z) \le \frac{1}{2}$. 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

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

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

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

By (4.3), the matrix V is symmetric positive definite and $V = \Theta(1)$. Diagonalize the symmetric matrix $\Delta_x(\mu) + \Delta_z(\mu) = Q^T\Lambda Q$ and consider

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

Since $QVQ^T = \Theta(1)$, the diagonal entries of QVQ^T must be $\Theta(1)$. 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, $\rho \le 0$ and $\|\Lambda\| = O(|\rho|)$. This further implies

$$\|\Delta_x(\mu) + \Delta_z(\mu)\| = O(|\rho|). \tag{5.6}$$

By the definition of the central path, we have

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

Now using the fact that the above matrix is symmetric, 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 (5.6) and using the fact that $\Delta_x(\mu) \perp \Delta_z(\mu)$, we have

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

Q.E.D.

Lemma 5.3 *Suppose $\delta(X,Z) \le 1/2$. Then there holds*

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

Proof. Notice that

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

and

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

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 the above two equations with $\bar{D}^{-1/2}$ 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 5.2, this implies $\bar{D} = \Theta(1)$ and

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

The lemma is proved.

Q.E.D.

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 5.4 *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 5.3 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

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

Q.E.D.

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

Lemma 5.5 *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 5.4 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 5.4 once more. **Q.E.D.**

Combining (5.5), Lemma 5.1 and Lemma 5.5 we can now estimate the order of $\left\|\Delta \bar{X}^p \Delta \bar{Z}^p\right\|$, and hence, using (5.4), we can estimate the predictor step length t_k.

Lemma 5.6 *We have*

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

Proof. Combining Lemma 5.5 with Lemma 5.1, we have

$$\left\|\sqrt{\mu}I + \Delta \bar{X}^p - \hat{X}^a\right\| + \left\|\sqrt{\mu}I + \Delta \bar{Z}^p - \hat{Z}^a\right\| = O(\sqrt{\mu}(\mu + \delta(X, Z))), \tag{5.7}$$

so that, using (4.5),

$$\left\|\sqrt{\mu}I - \hat{X}^a\right\| + \left\|\sqrt{\mu}I - \hat{Z}^a\right\| = O(\sqrt{\mu}). \tag{5.8}$$

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 (5.5), (5.7), (5.8) and (4.5) to the above relation yields

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

Q.E.D.

Theorem 5.2 *The iterates (X^k, Z^k) generated by Algorithm SDP 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 (5.4) 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

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

This implies using (4.8) and Lemma 5.6 that

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

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

$$\|X^k - X(\mu_k)\|_F \le \|L^T L\| \cdot \|L^{-1}(X^k - X(\mu_k))L^{-T}\|_F. \tag{5.9}$$

However, using the definition of F-norm and applying Lemma 5.3,

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

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,

$$\|L^T L\|_F = O(\frac{1}{\sqrt{\mu_k}}). \tag{5.10}$$

Combining (5.9) and (5.10) with Lemma 5.2, 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 3.1 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. **Q.E.D.**

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 3.1, 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 [13], 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 [5] and Potra-Sheng [12].

References

[1] Alizadeh, F., "Interior point methods in semidefinite programming with applications to combinatorial optimization problems," *SIAM Journal on Optimization* 5 (1995) 13-51.

[2] Helmberg, C., Rendl, F., Vanderbei, R.J. and Wolkowicz, H., "An interior-point method for semidefinite programming," Manuscript, Program in Statistics and Operations Research, Princeton University, USA, 1994.

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

[4] Kojima, M., Shida, M. and Shindoh, S., "Global and local convergence of predictor-correct infeasible-interior-point algorithm for semidefinite programming," Research Reports on Information Sciences, B-308, Dept. of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, 1995.

[5] Kojima, M., Shindoh, S. and Hara, S., "Interior-point methods for the monotone linear complementarity problem in symmetric matrices," Research Reports on Information Sciences, B-282, Dept. of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, 1994.

[6] Lin, C.J. and Saigal, R., "An infeasible start predictor corrector method for semi-definite linear programming," Research Report, Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, USA, 1995.

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

[8] Mizuno, S., Todd, M.J. and Ye, Y., "On Adaptive-Step Primal-Dual Interior-Point Algorithms for Linear Programming," *Mathematics of Operations Research* 18 (1993) 964-981.

[9] Monteiro, R.D.C., "Primal-dual path following algorithms for semidefinite programming," Technical Report, School of Industrial and Systems Engineering, Georgia Tech, Atlanta, Georgia, U.S.A., 1995.

[10] Monteiro, R.D.C. and Tsuchiya, T., "Limiting behavior of the derivatives of certain trajectories associated with a monotone horizontal linear complementarity problem," Working Paper 92-28, School of Industrial and Systems Engineering, Georgia Tech, Atlanta, Georgia, U.S.A., 1992.

[11] Nesterov, Y. and Todd, M.J., "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.

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

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

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

[15] Vandenberghe, L. and Boyd, S., "Semidefinite programming," submitted to SIAM Review.

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

[17] Zhang, Y., "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.