Journal of Math-for-Industry, Vol.2(2010B-9), pp.199–212

A polynomial-time inexact interior-point method for convex quadratic symmetric cone programming

Lu Li and Kim-Chuan Toh

Received on September 11, 2010

Abstract.
In this paper, we design an inexact primal-dual infeasible path-following algorithm for convex quadratic programming over symmetric cones. Our algorithm and its polynomial iteration complexity analysis give a unified treatment for a number of previous algorithms and their complexity analysis. In particular, our algorithm and analysis includes the one designed for linear semidefinite programming in "Math. Prog. 99 (2004), pp. 261–282". Under a mild condition on the inexactness of the search direction at each interior-point iteration, we show that the algorithm can find an ϵ-approximate solution in $O(n^2 \log(1/\epsilon))$ iterations, where n is the rank of the underlying Euclidean Jordan algebra.

Keywords. semidefinite programming, symmetric cone programming, infeasible interior point method, inexact search direction, polynomial complexity

1. Introduction

Our purpose in this paper is to propose (and establish the polynomial iteration complexity for) an inexact primal-dual path-following interior-point method (IPM) for a convex quadratic programming (CQP) problem over the product of cones of symmetric positive semidefinite matrices, second-order cones, and nonnegative orthants. As these cones are symmetric cones (which are cones of squares of some Euclidean Jordan algebras), it is convenient to study the problem in a unified manner under the framework of Euclidean Jordan algebras, and this is the framework we will adopt in this paper.

Recall that a Jordan algebra J is a finite dimensional vector space (over the field of real or complex numbers) endowed with a bilinear map $\circ : J \times J \to J$ satisfying the following properties for all $x, y \in J$: (i) $x \circ y = y \circ x$, (ii) $x \circ (x^2 \circ y) = x^2 \circ (x \circ y)$ where $x^2 = x \circ x$. Moreover, $(J, \circ)$ is called an Euclidean Jordan algebra if there exists a symmetric positive definite bilinear form $\langle \cdot, \cdot \rangle$ on J which is associative, i.e., $\langle x \circ y, z \rangle = \langle x, y \circ z \rangle$ for all $x, y, z \in J$. In other words, J has an inner product which is associative. In this paper, our presentation of the theory on Euclidean Jordan algebras will follow the book by Faraut and Korányi [5].

Let $\mathcal{J}$ be the product of p Euclidean Jordan algebras $(\mathcal{J}_i, \circ)$ with the identity element $\boldsymbol{e}_i$, i.e., $\boldsymbol{x}_i \circ \boldsymbol{e}_i = \boldsymbol{x}_i$ for all $\boldsymbol{x}_i \in \mathcal{J}_i$. Thus we have $\mathcal{J} = \{\boldsymbol{x} = (\boldsymbol{x}_i)_{i=1}^p : \boldsymbol{x}_i \in \mathcal{J}_i, i = 1, \ldots, p\}$ with $\boldsymbol{x} \circ \boldsymbol{y} = (\boldsymbol{x}_i \circ \boldsymbol{y}_i)_{i=1}^p$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{J}$, and the identity element of $\mathcal{J}$ is given by $\boldsymbol{e} = (\boldsymbol{e}_i)_{i=1}^p$. The associative inner product on $\mathcal{J}$ is induced naturally from the associative inner products of the constituent Euclidean Jordan algebras, i.e, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^p \langle \boldsymbol{x}_i, \boldsymbol{y}_i \rangle$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{J}$. We let $\mathcal{K}$ be the cone of squares of $\mathcal{J}$, i.e., $\mathcal{K} = \{\boldsymbol{x}^2 : \boldsymbol{x} \in \mathcal{J}\}$. Note that as we shall see in the next section, $\mathcal{K}$ is a symmetric cone. That is, $\mathcal{K}$ is a closed, pointed convex cone that is self-dual and its automorphism group acts transitively on its interior.

We consider the following CQP problem over symmetric cones:

$$(P) \quad \min \quad f(\boldsymbol{x}) := \tfrac{1}{2}\langle \boldsymbol{x}, \mathcal{H}(\boldsymbol{x}) \rangle + \langle \boldsymbol{c}, \boldsymbol{x} \rangle$$
$$\text{s.t.} \quad \mathcal{A}(\boldsymbol{x}) = b, \quad \boldsymbol{x} \in \mathcal{K},$$

where $\boldsymbol{c} \in \mathcal{J}$ and $b \in \mathbb{R}^m$ are given data, $\mathcal{A} : \mathcal{J} \to \mathbb{R}^m$ is a given linear map, and $\mathcal{H}$ is a given self-adjoint positive semidefinite (with respect to $\langle \cdot, \cdot \rangle$) linear operator on $\mathcal{J}$. Note that the inner product $\langle \cdot, \cdot \rangle$ will be defined explicitly in (2). The dual problem of (P) is given by

$$(D) \quad \max \quad -\tfrac{1}{2}\langle \boldsymbol{x}, \mathcal{H}(\boldsymbol{x}) \rangle + b^T y$$
$$\text{s.t.} \quad \mathcal{A}^T y + \boldsymbol{z} = \nabla f(\boldsymbol{x}) = \mathcal{H}(\boldsymbol{x}) + \boldsymbol{c}, \quad \boldsymbol{z} \in \mathcal{K}$$

where $\mathcal{A}^T$ denotes the adjoint of $\mathcal{A}$. Throughout the paper, we made the following assumptions.

Assumption 1. The problems (P) and (D) are strictly feasible, i.e., there exists $(\boldsymbol{x}, y, \boldsymbol{z})$ satisfying the linear constraints in (P) and (D) and $\boldsymbol{x}, \boldsymbol{z} \in \text{int}\mathcal{K}$, where $\text{int}\mathcal{K}$ denotes the interior of $\mathcal{K}$.

Assumption 2. The linear map $\mathcal{A}$ is surjective, which implies that $\mathcal{A}\mathcal{A}^T$ is non-singular, and the pseudo inverse of $\mathcal{A}$ is well defined as $\mathcal{A}^+ = \mathcal{A}^T(\mathcal{A}\mathcal{A}^T)^{-1}$.

The problem (P) includes the linear symmetric cone programming problems considered in [21] as a special case when $\mathcal{H} = \mathbf{0}$. In particular, it includes linear semidefinite programming (SDP) problems when $\mathcal{K}$ is the cone of symmetric positive semidefinite matrices, as well as the following linearly constrained convex quadratic programming (LCCQP): $\min\left\{\frac{1}{2}x^T Hx + c^T x : \; Ax = b, x \in \mathbb{R}^n_+\right\}$, where H is a given symmetric positive semidefinite matrix. The design and iteration complexity analysis of (exact) primal-dual IPMs for linear SDP problems have been studied intensively in the last two decades, and the development has been described in detail in the papers [19, 15, 21] and the references therein. The word "exact" here means that at each interior-point iteration, the search direction $\Delta\boldsymbol{w} := (\Delta\boldsymbol{x}, \Delta y, \Delta\boldsymbol{z})$ is computed exactly from the defining Newton-type linear system of equations $\mathcal{B}\Delta\boldsymbol{w} = \boldsymbol{h}$. In contrast, in inexact IPMs, the search direction is computed only approximately and the required accuracy in the residual vector $\boldsymbol{h} - \mathcal{B}\Delta\boldsymbol{w}$ is typically dependent on the current complementarity gap, and primal and dual infeasibilities. The main motivation for studying inexact IPMs comes from solving large problems, where the linear system of equations at each iteration is too large to be solved by a direct solver based on matrix factorizations but must be solved by an iterative linear solver such as the conjugate gradient method. The main advantage of an iterative solver is that only matrix-vector products are needed and it does not require computing and storing the entire coefficient matrix of the linear system. But for computational efficiency, the search direction is generally inexact in that it is not computed to high accuracy (an exact solution is deemed to be of machine accuracy). In order to guarantee the global polynomial convergence of an inexact IPM, the inexactness in the search direction must be controlled appropriately. That is, it must be accurate enough for the polynomial convergence to hold but at the same time the accuracy required must not be too stringent so that the iterative solver does not take unnecessarily large number of steps to compute the direction.

The application of Euclidean Jordan algebras as the basic toolbox for analyzing complexity proofs of (exact) IPMs for linear symmetric cone programming was started by Faybusovich [6] who extended earlier work of Nesterov and Todd, and Kojima et al. Tsuchiya [28] latter also used Jordan algebraic techniques to analyze primal-dual IPMs based on those of [17] for linear second-order cone programming. In [18], Muramatsu extended polynomial-time feasible IPMs for SDP to symmetric cone programming. Subsequently, Alizadeh et al. [1, 20, 21] studied primal-dual IPMs for linear symmetric cone programming extensively under the framework of Euclidean Jordan algebras.

The study of inexact IPMs was started in the 1980's when attempts were made to solve large linear programming (LP) problems. For LP and monotone linear complementarity problems, numerous papers have been devoted to the design and analysis of inexact IPMs; see for example [7, 11, 14] and the references therein. We should note that some of the inexact methods proposed earlier on may not be practically efficient as they require high accuracy on the computed search direction. For example in [14], since the iterates are required to maintain feasibility once it is achieved, this implies that the linear system must be solved to machine accuracy and the cost of solving the linear system turns out to be as expensive as in an exact algorithm. There are relatively fewer papers on the convergence analysis of inexact IPMs for LCCQP, the most recent ones we are aware of are [13, 3]. For the computational aspects of inexact IPMs for LP and LLCQP, we refer the readers to [2, 4] and the references therein.

For linear SDP, the first theoretical complexity analysis paper on inexact IPMs was published by Kojima et al. [9], wherein the algorithms only allow inexactness in the component corresponding to the complementarity equation (the third equation in (10)). Later, Zhou and Toh [30] developed an inexact IPM allowing inexactness not only in the complementarity equation but also in the primal and dual feasibilities. Furthermore, primal and dual feasibilities need not be maintained even if some iterates happen to be feasible. The latter property implies that the linear system at that particular iteration need not be solved to machine accuracy. The work in [30] was subsequently extended to convex quadratic SDP [12]. On the computational front, the readers are referred to [26, 24] for the design and implementation of efficient inexact IPMs for linear SDP, and to [25, 27] for the design of efficient inexact IPMs for convex quadratic SDP. We should mention that for efficient computation of the inexact directions, the design of efficient preconditioners for the increasingly ill-conditioned linear system at each iteration is the main challenge one must overcome in an inexact IPM.

In this paper, we design an inexact primal-dual infeasible path-following algorithm for convex quadratic symmetric cone programming problems. Our algorithm and its polynomial iteration complexity analysis give a unified treatment of a number of previous algorithms and their complexity analysis. In particular, for the case of linear SDP, our algorithm includes an analogous algorithm designed for linear SDP in [30]. It also extends the inexact IPM for convex quadratic SDP in [12] to convex quadratic symmetric programming, and hence it also includes the LCCQP problem studied in [13]. Under a mild condition on the inexactness of the search direction at each interior-point iteration, we show that our algorithm can find an ϵ-approximate solution in $O(n^2 \log(1/\epsilon))$ iterations, where n is the rank of the underlying Euclidean Jordan algebra.

The paper is organized as follows. In section 2, we provide some basic information on Euclidean Jordan algebras. In section 3, we define the infeasible central path and its corresponding neighborhood. In addition, we also establish some key lemmas that are needed for subsequent complexity analysis. In section 4, we discuss the computation of inexact search directions by an iterative linear solver. We also present our inexact primal-dual infeasible path-following algorithm and establish a polynomial complexity result for this algorithm. In section 5, we give detailed proofs on the polynomial complexity result.

2. Euclidean Jordan Algebras

In this section, we present some basic results on Euclidean Jordan algebras that are needed in the subsequent sections. Our presentation is mainly adapted from [1, 5, 20, 21].

Let $(\mathcal{J}, \circ)$ be an Euclidean Jordan algebra (assumed to have a unique unit element $\boldsymbol{e}$). Recall that an idempotent $\boldsymbol{c}$ is a nonzero element of $\mathcal{J}$ such that $\boldsymbol{c}^2 = \boldsymbol{c}$. An idempotent is said to be primitive if it is not the sum of two other idempotents. A complete system of orthogonal idempotents is a set of idempotents $\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k\}$ where $\boldsymbol{c}_i \circ \boldsymbol{c}_j = \boldsymbol{0}$ for any distinct i, j, and $\boldsymbol{c}_1 + \cdots + \boldsymbol{c}_k = \boldsymbol{e}$. For an element $\boldsymbol{x} \in \mathcal{J}$, the degree of $\boldsymbol{x}$ is the smallest integer such that $\{\boldsymbol{e}, \boldsymbol{x}, \ldots, \boldsymbol{x}^k\}$ is linearly dependent. The rank of $\mathcal{J}$ is the largest degree of $\boldsymbol{x} \in \mathcal{J}$. Let the rank of $\mathcal{J}$ be r, we can see that the maximum possible number of primitive idempotents in $\mathcal{J}$ is r. A complete system of orthogonal primitive idempotents $\{\boldsymbol{c}_1, \ldots \boldsymbol{c}_r\}$ is called a Jordan frame.

Theorem 1. [5, Theorem III.1.2] Let $\mathcal{J}$ be an Euclidean Jordan algebra with rank r and unit element $\boldsymbol{e}$. Then for any $\boldsymbol{x} \in \mathcal{J}$, there exists a Jordan frame $\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_r\}$ and real numbers $\lambda_1, \ldots, \lambda_r$ such that $\boldsymbol{x} = \lambda_1 \boldsymbol{c}_1 + \cdots + \lambda_r \boldsymbol{c}_r$. The numbers λ_i, $i = 1, \ldots, r$ (counting multiplicities) are uniquely determined by $\boldsymbol{x}$, and they are called the eigenvalues of $\boldsymbol{x}$.

For each $\boldsymbol{x} \in \mathcal{J}$, we will order its eigenvalues in a nonincreasing order, i.e., $\lambda_1 \geq \cdots \geq \lambda_r$. We said that $\boldsymbol{x} \in \mathcal{J}$ is positive semidefinite (definite) if all its eigenvalues are nonnegative (positive), and write $\boldsymbol{x} \succeq 0$ ($\boldsymbol{x} \succ 0$) if $\boldsymbol{x}$ is positive semidefinite (definite).

From the result in Theorem 1, we can define the following quantities for any $\boldsymbol{x} \in \mathcal{J}$:

$$\text{(1)} \quad \operatorname{tr}(\boldsymbol{x}) := \lambda_1 + \ldots + \lambda_r, \quad \det(\boldsymbol{x}) := \lambda_1 \ldots \lambda_r.$$

By [5, Proposition II.2.2], $\det(\boldsymbol{x} \circ \boldsymbol{y}) = \det(\boldsymbol{x}) \det(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{J}$, and $\operatorname{tr}(\boldsymbol{e}) = r$, $\det(\boldsymbol{e}) = 1$.

In general, for a given $\boldsymbol{x} \in \mathcal{J}$, we can define $f(\boldsymbol{x})$ for any real valued continuous function $f(\cdot)$ which is defined on an open set containing the set of eigenvalues $\Lambda(\boldsymbol{x}) := \{\lambda_1, \ldots, \lambda_r\}$ as follows:

$$f(\boldsymbol{x}) := f(\lambda_1)\boldsymbol{c}_1 + \cdots + f(\lambda_r)\boldsymbol{c}_r.$$

It is easy to see that the following identities are well-defined:

$$\begin{aligned}
\boldsymbol{x}^{-1} &:= \lambda_1^{-1}\boldsymbol{c}_1 + \cdots + \lambda_r^{-1}\boldsymbol{c}_r, \text{ if } \lambda_i \neq 0 \; \forall \, i \\
\boldsymbol{x}^{1/2} &:= \lambda_1^{1/2}\boldsymbol{c}_1 + \cdots + \lambda_r^{1/2}\boldsymbol{c}_r, \text{ if } \lambda_i \geq 0 \; \forall \, i \\
\|\boldsymbol{x}\|_2 &:= \max\{|\lambda_1|, \ldots, |\lambda_r|\} \\
\log\det(\boldsymbol{x}) &:= \log(\lambda_1) + \cdots + \log(\lambda_r), \text{ if } \lambda_i > 0 \; \forall \, i.
\end{aligned}$$

In [5, Proposition III.4.2], it is shown that $\nabla \log\det(\boldsymbol{x}) = \boldsymbol{x}^{-1}$ and $\nabla^2 \log\det(\boldsymbol{x}) = \mathbf{Q}(\boldsymbol{x}^{-1})$. Note that $(\boldsymbol{x}^{1/2})^2 = \boldsymbol{x}$ and $\boldsymbol{x}^{-1} \circ \boldsymbol{x} = \boldsymbol{e}$. If $\boldsymbol{x}^{-1}$ is well defined, we said that $\boldsymbol{x}$ is invertible.

By [5, Proposition III.1.5], we may define the following inner product on $\mathcal{J}$:

$$\text{(2)} \qquad \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \operatorname{tr}(\boldsymbol{x} \circ \boldsymbol{y}) \quad \forall \, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{J}.$$

From the associativity of $\operatorname{tr}(\cdot)$ [5, Proposition II.4.3], we know that $\langle \cdot, \cdot \rangle$ is also associative, i.e., $\langle \boldsymbol{x} \circ \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \circ \boldsymbol{z} \rangle$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{J}$. The norm $\| \cdot \|$ induced by the inner product $\langle \cdot, \cdot \rangle$ is given by

$$\text{(3)} \qquad \|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = (\lambda_1^2 + \cdots + \lambda_r^2)^{1/2}.$$

For any linear operator $\mathcal{B}$ defined on $\mathcal{J}$, we let $\|\mathcal{B}\|_2$ be the operator norm induced by $\| \cdot \|$.

As "$\circ$" is a bilinear map, for any $\boldsymbol{x} \in \mathcal{J}$, one can define the linear map $L(\boldsymbol{x}) : \mathcal{J} \to \mathcal{J}$ such that $L(\boldsymbol{x})\boldsymbol{y} = \boldsymbol{x} \circ \boldsymbol{y}$, and from the property of "$\circ$", one can verify that $L(\boldsymbol{x})L(\boldsymbol{x}^2) = L(\boldsymbol{x}^2)L(\boldsymbol{x})$ and $\langle L(\boldsymbol{x})\boldsymbol{z}, \boldsymbol{y} \rangle = \langle \boldsymbol{z}, L(\boldsymbol{x})\boldsymbol{y} \rangle$. In addition to $L(\boldsymbol{x})$, there is another linear map $\mathbf{Q}(\boldsymbol{x})$ (called the quadratic representation) associated with $\boldsymbol{x}$ which is defined by

$$\mathbf{Q}(\boldsymbol{x}) := 2L^2(\boldsymbol{x}) - L(\boldsymbol{x}^2).$$

From the definition of $\mathbf{Q}(\boldsymbol{x})$, it is clear that $\langle \mathbf{Q}(\boldsymbol{x})\boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{y}, \mathbf{Q}(\boldsymbol{x})\boldsymbol{z} \rangle$. Thus we can see that both $L(\boldsymbol{x})$ and $\mathbf{Q}(\boldsymbol{x})$ are self-adjoint linear operators in $\mathcal{J}$. Moreover, they satisfy the properties stated in the following lemma.

Lemma 1. [21, Lemma 12] Given the spectral decomposition $\boldsymbol{x} = \lambda_1 \boldsymbol{c}_1 + \ldots + \lambda_r \boldsymbol{c}_r$ in a rank r Euclidean Jordan algebra, we have that:

1. The operators $L(\boldsymbol{x})$ and $\mathbf{Q}(\boldsymbol{x})$ commute and share a common system of eigenvectors.
2. The eigenvalues of $L(\boldsymbol{x})$ are given by $\{(\lambda_i + \lambda_j)/2 : 1 \leq i, j \leq r\}$. Hence $\|L(\boldsymbol{x})\|_2 = \|\boldsymbol{x}\|_2$. Also, $\boldsymbol{x} \succeq 0$ ($\succ 0$) iff $L(\boldsymbol{x})$ is positive semidefinite (positive definite).
3. The eigenvalues of $\mathbf{Q}(\boldsymbol{x})$ are given by $\{\lambda_i \lambda_j : 1 \leq i, j \leq r\}$. Hence $\|\mathbf{Q}(\boldsymbol{x})\|_2 = \|\boldsymbol{x}\|_2^2$.
4. $\|\boldsymbol{x} \circ \boldsymbol{y}\| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$.

The last inequality is obvious since $\|\boldsymbol{x} \circ \boldsymbol{y}\| = \|L(\boldsymbol{x})\boldsymbol{y}\| \leq \|L(\boldsymbol{x})\|_2 \|\boldsymbol{y}\| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$. From the fact that $\langle \cdot, \cdot \rangle$ is an inner product, we also have the Cauchy-Schwarz inequality

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|.$$

Next we present some useful identities needed in the remaining part of the paper.

Lemma 2. The quadratic representation $\mathbf{Q}$ satisfies the following properties:

1. $\mathbf{Q}(\mathbf{Q}(\boldsymbol{x})\boldsymbol{y}) = \mathbf{Q}(\boldsymbol{x})\mathbf{Q}(\boldsymbol{y})\mathbf{Q}(\boldsymbol{x})$ [5, Prop. II.3.3].
2. $\mathbf{Q}(\boldsymbol{x}^{-1/2})\boldsymbol{x} = \boldsymbol{e}$, $\mathbf{Q}(\boldsymbol{x})^{-1} = \mathbf{Q}(\boldsymbol{x}^{-1})$, $\mathbf{Q}(\boldsymbol{x})\boldsymbol{x}^{-1} = \boldsymbol{x}$ [5, Prop. II.3.1].

3. $\mathrm{tr}(\mathbf{Q}(\boldsymbol{x})\boldsymbol{y}) = \langle \boldsymbol{x}^2, \boldsymbol{y}\rangle$.
4. $\det(\mathbf{Q}(\boldsymbol{x})\boldsymbol{y}) = \det(\boldsymbol{x}^2)\det(\boldsymbol{y})$ [5, Prop. III.4.2].

Note that from part 1 of the above lemma, we have

$$\mathbf{Q}(\boldsymbol{x}^2) = \mathbf{Q}(\boldsymbol{x})^2, \quad \mathbf{Q}(\boldsymbol{x}^{1/2}) = \mathbf{Q}(\boldsymbol{x})^{1/2} \text{ if } \boldsymbol{x} \succeq 0.$$

Recall that for an Euclidean Jordan algebra $\mathcal{J}$, its cone of squares is the set

$$\mathcal{K} := \left\{\boldsymbol{x}^2 \mid \boldsymbol{x} \in \mathcal{J}\right\}.$$

With $\mathcal{K}$ defined, we can now state one of the most celebrated result in Euclidean Jordan algebras on the Jordan algebraic characterization of symmetric cones.

Theorem 2. [5, Theorem III.2.1, Theorem III.3.1] A cone is symmetric if and only if it is the cone of squares of an Euclidean Jordan algebra. Furthermore, $\mathcal{K} = \{\boldsymbol{x} \in \mathcal{J} : x \succeq 0\}$ and $\mathrm{int}\mathcal{K} = \{\boldsymbol{x} \in \mathcal{J} : x \succ 0\}$.

For $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{J}$, we write $\boldsymbol{x} \succeq \boldsymbol{y}$ ($\boldsymbol{x} \succ \boldsymbol{y}$) if $\boldsymbol{x} - \boldsymbol{y} \in \mathcal{K}$ ($\mathrm{int}\mathcal{K}$).

Next we state a lemma that will be used frequently in the next few sections.

Lemma 3. [21, Proposition 21, Lemma 30] Let $\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{p} \in \mathrm{int}\mathcal{K}$. Define $\widehat{\boldsymbol{x}} = \mathbf{Q}(\boldsymbol{p})\boldsymbol{x}$ and $\widehat{\boldsymbol{z}} = \mathbf{Q}(\boldsymbol{p}^{-1})\boldsymbol{z}$, then

1. $\mathbf{Q}(\boldsymbol{x}^{1/2})\boldsymbol{z}$ and $\mathbf{Q}(\boldsymbol{z}^{1/2})\boldsymbol{x}$ have the same spectrum.
2. $\mathbf{Q}(\boldsymbol{x}^{1/2})\boldsymbol{z}$ and $\mathbf{Q}(\widehat{\boldsymbol{x}}^{1/2})\widehat{\boldsymbol{z}}$ have the same spectrum.
3. $\lambda_{\max}(\boldsymbol{x} \circ \boldsymbol{z}) \geq \lambda_{\max}(\mathbf{Q}(\boldsymbol{x}^{1/2})\boldsymbol{z})$, with equality holding if $L(\boldsymbol{x}), L(\boldsymbol{z})$ commute.
4. $\lambda_{\min}(\boldsymbol{x} \circ \boldsymbol{z}) \leq \lambda_{\min}(\mathbf{Q}(\boldsymbol{x}^{1/2})\boldsymbol{z})$, with equality holding if $L(\boldsymbol{x}), L(\boldsymbol{z})$ commute; see also [22, Thm. 4].

Finally we give a brief description of the three basic symmetric cones focused in this paper.

Semidefinite cone. Let $\mathcal{J} = \mathcal{S}^n$, the vector space of $n \times n$ real symmetric matrices. Define $\circ$ by

$$X \circ Y := \frac{1}{2}(XY + YX).$$

Then $(\mathcal{S}^n, \circ)$ is an Euclidean Jordan algebra with the unit element being the identity matrix I. The cone of squares is $\mathcal{S}^n_+$, the set of all positive semidefinite matrices in $\mathcal{S}^n$. For any $X \in S^n$, it has an eigenvalue decomposition given by

$$X = \lambda_1 \boldsymbol{q}_1\boldsymbol{q}_1^T + \ldots + \lambda_n \boldsymbol{q}_n\boldsymbol{q}_n^T,$$

where the set $\{\boldsymbol{q}_1\boldsymbol{q}_1^T, \ldots, \boldsymbol{q}_n\boldsymbol{q}_n^T\}$ is a Jordan frame, and we have $\det(X) = \lambda_1 \cdots \lambda_n$ and $\mathrm{tr}(X) = \lambda_1 + \cdots + \lambda_n$. Thus

$$\langle X, Y\rangle := \mathrm{tr}(X \circ Y) = \mathrm{tr}(XY).$$

Also, we have

$$(4) \qquad L(X) = X \circledast I, \quad \mathbf{Q}(X) = X \circledast X$$

where $G \circledast H$ denotes the symmetric Kronecker product of two $n \times n$ matrices G and H defined by

$$(5) \quad (G \circledast H)(M) := \frac{1}{2}(GMH^T + HMG^T) \quad \forall\, M \in \mathcal{S}^n.$$

We refer the reader to the appendix of [23] for some of its properties.

Second-order cone. Let $\mathcal{J} = \mathbb{R}^n$, where we write any vector $\boldsymbol{x} \in \mathbb{R}^n$ in the form $\boldsymbol{x} = (x_0; \bar{\boldsymbol{x}})$ with $x_0 \in \mathbb{R}$ and $\bar{\boldsymbol{x}} \in \mathbb{R}^{n-1}$. Define $\boldsymbol{x} \circ \boldsymbol{y} := (\boldsymbol{x}^T\boldsymbol{y}; x_0\bar{\boldsymbol{y}} + y_0\bar{\boldsymbol{x}})$. The unit element is $\boldsymbol{e} := (1; \mathbf{0}) \in \mathbb{R}^n$. The cone of squares is given by

$$\mathcal{Q}^n = \{\boldsymbol{x} \in \mathbb{R}^n : \|\bar{\boldsymbol{x}}\| \leq x_0\},$$

which is a second-order cone. Note that $\|\bar{\boldsymbol{x}}\|$ is the usual Euclidean norm of a vector. For any $\boldsymbol{x} \in \mathbb{R}^n$, it has the following eigenvalue decomposition

$$\boldsymbol{x} = (x_0 + \|\bar{\boldsymbol{x}}\|)\boldsymbol{c}_1 + (x_0 - \|\bar{\boldsymbol{x}}\|)\boldsymbol{c}_2,$$

where

$$\boldsymbol{c}_1 = \begin{cases} (\frac{1}{2};\ \frac{\bar{\boldsymbol{x}}}{2\|\bar{\boldsymbol{x}}\|}) & \bar{\boldsymbol{x}} \neq \mathbf{0} \\ (\frac{1}{2};\ \frac{1}{2}; \mathbf{0}\,) & \bar{\boldsymbol{x}} = \mathbf{0} \end{cases}$$

and $\boldsymbol{c}_2 = \boldsymbol{e} - \boldsymbol{c}_1$. Thus, we have $\lambda_{1,2} = x_0 \pm \|\bar{\boldsymbol{x}}\|$, and $\mathrm{tr}(\boldsymbol{x}) = 2x_0$ and $\det(\boldsymbol{x}) = x_0^2 - \|\bar{\boldsymbol{x}}\|^2$. For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle := \mathrm{tr}(\boldsymbol{x} \circ \boldsymbol{y}) = 2\boldsymbol{x}^T\boldsymbol{y}.$$

Also, we have

$$(6) \quad L(\boldsymbol{x}) = \begin{bmatrix} x_0 & \bar{\boldsymbol{x}}^T \\ \bar{\boldsymbol{x}} & x_0 I \end{bmatrix}$$

$$(7) \quad \mathbf{Q}(\boldsymbol{x}) = 2L^2(\boldsymbol{x}) - L(\boldsymbol{x}^2) = 2\boldsymbol{x}\boldsymbol{x}^T - \det(\boldsymbol{x})R.$$

where R is a diagonal matrix with the $(1, 1)$ entry equal to 1 and the remaining diagonal entries equal to -1. If $\det(\boldsymbol{x}) \neq 0$, we have

$$(8) \qquad \boldsymbol{x}^{-1} = \lambda_1^{-1}\boldsymbol{c}_1 + \lambda_2^{-1}\boldsymbol{c}_2 = \frac{1}{\det(\boldsymbol{x})}R\boldsymbol{x}$$

$$(9) \qquad \mathbf{Q}(\boldsymbol{x}^{-1}) = \frac{1}{\det(\boldsymbol{x})^2}R\mathbf{Q}(\boldsymbol{x})R.$$

Nonnegative orthants. Let $\mathcal{J} = \mathbb{R}^n$, the n-dimensional real vector space with $\boldsymbol{x} \circ \boldsymbol{y} := (x_iy_i)_{i=1}^n$. The unit element of $\mathcal{J}$ is $\boldsymbol{e} = (1, \ldots, 1) \in \mathbb{R}^n$, and the cone of squares is $\mathbb{R}^n_+ = \{\boldsymbol{x} : x_i \geq 0\ \forall\, i = 1, \ldots, n\}$. Every $\boldsymbol{x} \in \mathbb{R}^n$ has the eigenvalue decomposition

$$\boldsymbol{x} = x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n,$$

where $\boldsymbol{e}_i$ is the ith unit vector in $\mathbb{R}^n$. Thus we have $\mathrm{tr}(\boldsymbol{x}) = x_1 + \cdots + x_n$, $\det(\boldsymbol{x}) = x_1 \cdots x_n$, and $\langle \boldsymbol{x}, \boldsymbol{y}\rangle := \mathrm{tr}(\boldsymbol{x} \circ \boldsymbol{y}) = \boldsymbol{x}^T\boldsymbol{y}$. Also, we have

$$L(\boldsymbol{x}) = \mathrm{diag}(\boldsymbol{x}),\ \mathbf{Q}(\boldsymbol{x}) = \mathrm{diag}(\boldsymbol{x}^2),\ \mathbf{Q}(\boldsymbol{x}^{-1}) = \mathrm{diag}(\boldsymbol{x}^{-2}).$$

3. An infeasible central path and its neighborhood

The perturbed Karush-Kuhn-Tucker (KKT) optimality conditions for the problems (P) and (D) are given as follows:
(10)
$$\begin{pmatrix} -\nabla f(\boldsymbol{x}) + \mathcal{A}^T y + \boldsymbol{z} \\ \mathcal{A}(\boldsymbol{x}) - b \\ \boldsymbol{x} \circ \boldsymbol{z} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \nu \boldsymbol{e} \end{pmatrix}, \quad \boldsymbol{x}, \boldsymbol{z} \in \mathcal{K},$$

where ν is a positive parameter that is to be driven to zero explicitly. The last equation of (10) is a relaxation of the complementarity conditions $\boldsymbol{x} \circ \boldsymbol{z} = 0$, $\boldsymbol{x}, \boldsymbol{z} \in \mathcal{K}$. When $\nu = 0$, (10) gives the optimal conditions for (P) and (D). Let $n = \mathrm{tr}(\boldsymbol{e})$. For a nonzero ν, (10) is the optimality condition for the log-determinant problems, that is, adding the log barrier terms $-\nu \log\det \boldsymbol{x}$ and $\nu \log\det \boldsymbol{z}$ to (P) and (D) respectively. Just like the case of a linear SDP, linearizing the third equation in (10) may not lead to an element in $\mathcal{J}$. Thus it is necessary to "symmetrize" that equation before linearizing it. That is, given an invertible $\boldsymbol{p} \in \mathcal{K}$, the last equation of (10) is replaced by

$$H_{\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{z}) := \mathbf{Q}(\boldsymbol{p})\boldsymbol{x} \circ \mathbf{Q}(\boldsymbol{p}^{-1})\boldsymbol{z} = \nu \boldsymbol{e}. \tag{11}$$

It has been shown in [21, Lemma 28] that for $\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{p} \in \mathcal{J}$, if $\boldsymbol{x}, \boldsymbol{z} \succ 0$ and $\boldsymbol{p}$ invertible, then $H_{\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{z}) = \nu \boldsymbol{e}$ if and only if $\boldsymbol{x} \circ \boldsymbol{z} = \nu \boldsymbol{e}$.

In this paper, for $\boldsymbol{x}, \boldsymbol{z} \succ 0$, we only consider $\boldsymbol{p}$ that is in the commutative class defined by

$$\mathcal{C}(\boldsymbol{x}, \boldsymbol{z}) = \{\boldsymbol{p} \in \mathrm{int}\mathcal{K} \mid L(\mathbf{Q}(\boldsymbol{p})\boldsymbol{x}),\ L(\mathbf{Q}(\boldsymbol{p}^{-1})\boldsymbol{z}) \text{ commute}\}.$$

Note that if $\boldsymbol{p} \in \mathcal{C}(\boldsymbol{x}, \boldsymbol{z})$ and we let $\widehat{\boldsymbol{x}} = \mathbf{Q}(\boldsymbol{p})\boldsymbol{x}$, $\widehat{\boldsymbol{z}} = \mathbf{Q}(\boldsymbol{p}^{-1})\boldsymbol{z}$, then

$$H_{\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{z}) = \widehat{\boldsymbol{x}} \circ \widehat{\boldsymbol{z}} = \mathbf{Q}(\widehat{\boldsymbol{x}}^{1/2})\widehat{\boldsymbol{z}}. \tag{12}$$

The class $\mathcal{C}(\boldsymbol{x}, \boldsymbol{z})$ includes the common choices: $\boldsymbol{p} = \boldsymbol{z}^{1/2}, \boldsymbol{p} = \boldsymbol{x}^{-1/2}$, a.k.a. Helmberg-Kojima-Monteiro (HKM) direction [8, 10, 16], and Nesterov-Todd (NT) direction $\boldsymbol{p} = \boldsymbol{w}^{-1/2}$ where $\boldsymbol{w}$ is the NT scaling element satisfying $\mathbf{Q}(\boldsymbol{w})\boldsymbol{z} = \boldsymbol{x}$ [23]. For the NT scaling element, an explicit form of $\boldsymbol{p}$ is given by

$$\begin{aligned} \boldsymbol{p} &= \left[\mathbf{Q}(\boldsymbol{x}^{1/2})\left(\mathbf{Q}(\boldsymbol{x}^{1/2})\boldsymbol{z}\right)^{-1/2}\right]^{-1/2} \\ &= \left[\mathbf{Q}(\boldsymbol{z}^{-1/2})\left(\mathbf{Q}(\boldsymbol{z}^{1/2})\boldsymbol{x}\right)^{1/2}\right]^{-1/2}. \end{aligned} \tag{13}$$

We can easily verify that

$$\mathbf{Q}(\boldsymbol{w})\boldsymbol{z} = \mathbf{Q}(\boldsymbol{p}^{-2})\boldsymbol{z} = \mathbf{Q}\big(\mathbf{Q}(\boldsymbol{x}^{1/2})(\mathbf{Q}(\boldsymbol{x}^{1/2})\boldsymbol{z})^{-1/2}\big)\boldsymbol{z}$$
$$= \mathbf{Q}(\boldsymbol{x}^{1/2})\mathbf{Q}\big((\mathbf{Q}(\boldsymbol{x}^{1/2})\boldsymbol{z})^{-1/2}\big)(\mathbf{Q}(\boldsymbol{x}^{1/2})\boldsymbol{z}) = \mathbf{Q}(\boldsymbol{x}^{1/2})\boldsymbol{e} = \boldsymbol{x}.$$

The last two equalities above use the properties of $\mathbf{Q}(\cdot)$ stated in Lemma 2. Moreover, for the NT scaling element, we have

$$\mathbf{Q}(\boldsymbol{p}^{-1})\boldsymbol{z} = \mathbf{Q}(\boldsymbol{p})\mathbf{Q}(\boldsymbol{w})\boldsymbol{z} = \mathbf{Q}(\boldsymbol{p})\boldsymbol{x}. \tag{14}$$

In this paper, we choose $\boldsymbol{p}$ to be the NT scaling element rather than any other $\boldsymbol{p} \in \mathcal{C}(\boldsymbol{x}, \boldsymbol{z})$ as considered in [29]. The main reason for considering only the NT scaling element is that it simplifies the complexity analysis and also gives the best iteration complexity. In addition, it is employed in practical computations since it has certain desirable properties that allow one to design efficient preconditioners for the augmented system (33a) for computing search directions; see [25] for details.

Let $L = \|\mathcal{H}\|_2$. Note that L is a Lipschitz constant of the gradient of $f(\boldsymbol{x})$ defined in (P), i.e.,

$$\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| = \|\mathcal{H}(\boldsymbol{x}) - \mathcal{H}(\boldsymbol{y})\| \le L\|\boldsymbol{x} - \boldsymbol{y}\|. \tag{15}$$

Let $(\boldsymbol{x}_0, y_0, \boldsymbol{z}_0)$ be an initial point such that

$$\boldsymbol{x}_0 = \boldsymbol{z}_0 = \rho \boldsymbol{e}, \tag{16}$$

where $\rho > 0$ is a given constant. For given positive constants $\gamma_p \le \gamma_d$ such that $\gamma_d + L\gamma_p \in (0,1)$, the constant ρ is chosen to be sufficiently large so that for some solution $(\boldsymbol{x}_*, y_*, \boldsymbol{z}_*)$ to (P) and (D), the following conditions hold:

$$(1-\gamma_p)\boldsymbol{x}_0 \succ \boldsymbol{x}_* \succeq 0,\ (1-(\gamma_d + L\gamma_p))\boldsymbol{z}_0 \succ \boldsymbol{z}_* \succeq 0, \tag{17}$$

$$\mathrm{tr}(\boldsymbol{x}_*) + \mathrm{tr}(\boldsymbol{z}_*) \ \le\ n\rho. \tag{18}$$

Recall that $n = \mathrm{tr}(\boldsymbol{e})$.

We define

$$\mu_0 \ = \ \langle \boldsymbol{x}_0, \boldsymbol{z}_0 \rangle / n \ = \rho^2, \tag{19}$$

$$R_0^p \ = \ \mathcal{A}(\boldsymbol{x}_0) - b, \tag{20}$$

$$R_0^d \ = \ -\nabla f(\boldsymbol{x}_0) + \mathcal{A}^T y_0 + \boldsymbol{z}_0. \tag{21}$$

For $\theta, \nu \in (0,1]$, the following infeasible KKT system has a unique solution under Assumptions 1 and 2:

$$\begin{pmatrix} -\nabla f(\boldsymbol{x}) + \mathcal{A}^T y + \boldsymbol{z} \\ \mathcal{A}(\boldsymbol{x}) - b \\ H_{\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{z}) \end{pmatrix} = \begin{pmatrix} \theta R_0^d \\ \theta R_0^p \\ \nu\mu_0 \boldsymbol{e} \end{pmatrix}, \ \boldsymbol{x}, \boldsymbol{z} \succ 0. \tag{22}$$

Define the infeasible central path as:

$$\mathcal{P} = \{(\theta, \nu, \boldsymbol{x}, y, \boldsymbol{z}) | \theta, \nu > 0, \boldsymbol{x}, \boldsymbol{z} \in \mathrm{int}\mathcal{K}, y \in \mathbb{R}^m, (22) \text{ holds}\}.$$

The primary idea of a primal-dual infeasible path-following algorithm is to generate a sequence of points $(\boldsymbol{x}^k, y^k, \boldsymbol{z}^k)$ such that $(\theta^k, \nu^k, \boldsymbol{x}^k, y^k, \boldsymbol{z}^k) \in \mathcal{P}$ and $(\boldsymbol{x}^k, y^k, \boldsymbol{z}^k)$ converges to a solution of (P) and (D) when θ^k and ν^k are driven to 0. In practice of course, the points are never exactly on the central path $\mathcal{P}$ but lie in some neighborhood of $\mathcal{P}$. In our inexact primal-dual infeasible path-following algorithm, we consider the following neighborhood of $\mathcal{P}$. Choose a constant $\gamma \in (0,1)$ in addition to γ_p and γ_d, we define the neighborhood to be:

$$\mathcal{N} = \left\{ \begin{array}{l} (\theta, \nu, \boldsymbol{x}, y, \boldsymbol{z}) \in (0,1] \times (0,1] \times \mathrm{int}\mathcal{K} \times \mathbb{R}^m \times \mathrm{int}\mathcal{K} \\ \theta \le \nu,\ \mathcal{A}(\boldsymbol{x}) - b = \theta(R_0^p + \xi^p),\ \|\mathcal{A}^+\xi^p\| \le \gamma_p\rho, \\ -\nabla f(\boldsymbol{x}) + \mathcal{A}^T y + \boldsymbol{z} = \theta(R_0^d + \xi^d),\ \|\xi^d\| \le \gamma_d\rho, \\ \|\mathbf{Q}(\boldsymbol{x}^{1/2})\boldsymbol{z} - \nu\mu_0\boldsymbol{e}\|_2 \le \gamma\nu\mu_0. \end{array} \right\}.$$

Let $\theta_0 = \nu_0 = 1$. Then (16) implies that $(\theta_0, \nu_0, \boldsymbol{x}_0, y_0, \boldsymbol{z}_0) \in \mathcal{N}$.

From the definition of $\mathcal{N}$ and Lemma 3, it is easy to see that we have the following lemma.

Lemma 4. Suppose $(\theta, \nu, \boldsymbol{x}, y, \boldsymbol{z}) \in \mathcal{N}$ and $\boldsymbol{p} \in \mathcal{C}(\boldsymbol{x}, \boldsymbol{z})$. Then

$$(23) \qquad (1-\gamma)\nu\mu_0 \boldsymbol{e} \preceq H_{\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{z}) \preceq (1+\gamma)\nu\mu_0 \boldsymbol{e}$$

$$(24) \qquad (1-\gamma)\nu\mu_0 \;\le\; \langle \boldsymbol{x}, \boldsymbol{z}\rangle/n \;\le\; (1+\gamma)\nu\mu_0.$$

Proof. (a) Define $\widehat{\boldsymbol{x}} = \mathbf{Q}(\boldsymbol{p})\boldsymbol{x}$ and $\widehat{\boldsymbol{z}} = \mathbf{Q}(\boldsymbol{p}^{-1})\boldsymbol{z}$, and $\mu = \nu\mu_0$. By Lemma 3 and (14), $H_{\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{z}) = \mathbf{Q}(\widehat{\boldsymbol{x}}^{1/2})\widehat{\boldsymbol{z}}$ and $\mathbf{Q}(\boldsymbol{x}^{1/2})\boldsymbol{z}$ have the same spectrum. Hence

$$\begin{aligned} &|\lambda_{\min}(H_{\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{z})) - \mu| = |\lambda_{\min}(\mathbf{Q}(\boldsymbol{x}^{1/2})\boldsymbol{z}) - \mu| \\ = \;& |\lambda_{\min}(\mathbf{Q}(\boldsymbol{x}^{1/2})\boldsymbol{z} - \mu\boldsymbol{e})| \le \|\mathbf{Q}(\boldsymbol{x}^{1/2})\boldsymbol{z} - \mu\boldsymbol{e}\|_2 \le \gamma\mu. \end{aligned}$$

This implies that $\lambda_{\min}(H_{\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{z}) - (1-\gamma)\mu\boldsymbol{e})$ $= \lambda_{\min}(H_{\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{z})) - (1-\gamma)\mu \ge 0$, and hence $H_{\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{z}) \succeq (1-\gamma)\mu\boldsymbol{e}$. The proof of the other right-hand side partial order in (23) is similar, and we shall omit it.

The inequalities in (24) follows from (23) and the fact that

$$\begin{aligned} &\mathrm{tr}(H_{\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{z})) = \langle \widehat{\boldsymbol{x}}, \widehat{\boldsymbol{z}}\rangle = \langle \mathbf{Q}(\boldsymbol{p})\boldsymbol{x}, \mathbf{Q}(\boldsymbol{p}^{-1})\boldsymbol{z}\rangle \\ = \;& \langle \boldsymbol{x}, \mathbf{Q}(\boldsymbol{p})\mathbf{Q}(\boldsymbol{p}^{-1})\boldsymbol{z}\rangle = \langle \boldsymbol{x}, \boldsymbol{z}\rangle. \end{aligned}$$

□

Next, we present two lemmas that are needed for the iteration complexity analysis in section 4.

Lemma 5. For any r_p and r_d satisfying $\|r_d\| \le \gamma_d\rho$ and $\|\mathcal{A}^+ r_p\| \le \gamma_p\rho$, there exists $(\widetilde{\boldsymbol{x}}, \widetilde{y}, \widetilde{\boldsymbol{z}})$ that satisfies the following conditions:

$$(25) \quad -\nabla f(\widetilde{\boldsymbol{x}}) + \mathcal{A}^T\widetilde{y} + \widetilde{\boldsymbol{z}} = R_0^d + r_d\,,$$

$$(26) \quad \mathcal{A}(\widetilde{\boldsymbol{x}}) - b = R_0^p + r_p\,,$$

$$(27) \quad (1-\gamma_p)\rho\boldsymbol{e} \preceq \widetilde{\boldsymbol{x}} \preceq (1+\gamma_p)\rho\boldsymbol{e}\,,$$

$$(28) \quad [1-(\gamma_d + L\gamma_p)]\rho\boldsymbol{e} \preceq \widetilde{\boldsymbol{z}} \preceq [1+(\gamma_d + L\gamma_p)]\rho\boldsymbol{e}\,.$$

Proof. Take

$$\widetilde{\boldsymbol{x}} = \boldsymbol{x}_0 + \mathcal{A}^+ r_p,\; \widetilde{y} = y_0,\; \widetilde{\boldsymbol{z}} = \boldsymbol{z}_0 + r_d + \nabla f(\widetilde{\boldsymbol{x}}) - \nabla f(\boldsymbol{x}_0)\,,$$

then (25)–(27) are readily shown. To show (28), we only need to establish the following inequality:

$$\begin{aligned} \|r_d + \nabla f(\widetilde{\boldsymbol{x}}) - \nabla f(\boldsymbol{x}_0)\| &\le \|r_d\| + \|\nabla f(\widetilde{\boldsymbol{x}}) - \nabla f(\boldsymbol{x}_0)\| \\ &\le (\gamma_d + L\gamma_p)\rho. \end{aligned}$$

□

Lemma 6. Given the initial conditions (16), (17) and (18), for any $(\theta, \nu, \boldsymbol{x}, y, \boldsymbol{z}) \in \mathcal{N}$, we have

$$\theta\mathrm{tr}(\boldsymbol{x}) \le \frac{6\nu\rho n}{1-(\gamma_d + L\gamma_p)}, \quad \theta\mathrm{tr}(\boldsymbol{z}) \le \frac{6\nu\rho n}{1-\gamma_p}.$$

Proof. This proof is adapted from that of Lemma 2 in [30]. For $(\theta, \nu, \boldsymbol{x}, y, \boldsymbol{z}) \in \mathcal{N}$, we have

$$\begin{aligned} -\nabla f(\boldsymbol{x}) + \mathcal{A}^T y + \boldsymbol{z} = \theta(R_0^d + r_d), \quad \|r_d\| \le \gamma_d\rho, \\ \mathcal{A}(\boldsymbol{x}) - b = \theta(R_0^p + r_p), \quad \|\mathcal{A}^+ r_p\| \le \gamma_p\rho. \end{aligned}$$

By Lemma 5, there exists $(\widetilde{\boldsymbol{x}}, \widetilde{y}, \widetilde{\boldsymbol{z}})$ satisfies conditions (25)–(28). Also, the solution $(\boldsymbol{x}_*, y_*, \boldsymbol{z}_*)$ to (P) and (D) considered in (17) satisfies the following equations:

$$\begin{aligned} \mathcal{A}(\boldsymbol{x}_*) - b = 0, \\ -\nabla f(\boldsymbol{x}_*) + \mathcal{A}^T y_* + \boldsymbol{z}_* = 0. \end{aligned}$$

Let

$$\bar{\boldsymbol{x}} = (1-\theta)\boldsymbol{x}_* + \theta\widetilde{\boldsymbol{x}} - \boldsymbol{x},\; \bar{y} = (1-\theta)y_* + \theta\widetilde{y} - y,\; \bar{\boldsymbol{z}} = (1-\theta)\boldsymbol{z}_* + \theta\widetilde{\boldsymbol{z}} - \boldsymbol{z}.$$

Then we have

$$\mathcal{A}(\bar{\boldsymbol{x}}) = 0, \quad \mathcal{A}^T(\bar{y}) + \bar{\boldsymbol{z}} = \mathcal{H}\,\bar{\boldsymbol{x}}.$$

Hence $\langle \bar{\boldsymbol{x}}, \bar{\boldsymbol{z}}\rangle = \langle \bar{\boldsymbol{x}}, \mathcal{H}(\bar{\boldsymbol{x}})\rangle$. Together with the fact that $\mathcal{H}$ is positive semidefinite, we have

$$\begin{aligned} (29) &\langle (1-\theta)\boldsymbol{x}_* + \theta\widetilde{\boldsymbol{x}}, \boldsymbol{z}\rangle + \langle \boldsymbol{x}, (1-\theta)\boldsymbol{z}_* + \theta\widetilde{\boldsymbol{z}}\rangle \\ = \;& \langle (1-\theta)\boldsymbol{x}_* + \theta\widetilde{\boldsymbol{x}}, (1-\theta)\boldsymbol{z}_* + \theta\widetilde{\boldsymbol{z}}\rangle + \langle \boldsymbol{x}, \boldsymbol{z}\rangle - \langle \bar{\boldsymbol{x}}, \mathcal{H}(\bar{\boldsymbol{x}})\rangle \\ \le \;& \langle (1-\theta)\boldsymbol{x}_* + \theta\widetilde{\boldsymbol{x}}, (1-\theta)\boldsymbol{z}_* + \theta\widetilde{\boldsymbol{z}}\rangle + \langle \boldsymbol{x}, \boldsymbol{z}\rangle. \end{aligned}$$

By using (18), (24), (27), (28), (29), and the fact that $\langle \boldsymbol{x}_*, \boldsymbol{z}_*\rangle = 0$, $\langle \boldsymbol{x}_*, \boldsymbol{z}\rangle$, $\langle \boldsymbol{x}, \boldsymbol{z}_*\rangle \ge 0$, we have that

$$\begin{aligned} &\theta\rho[(1-(\gamma_d + L\gamma_p))\langle \boldsymbol{e}, \boldsymbol{z}\rangle + (1-\gamma_p)\langle \boldsymbol{e}, \boldsymbol{z}\rangle] \\ \le \;& \theta(\langle \widetilde{\boldsymbol{z}}, \boldsymbol{x}\rangle + \langle \widetilde{\boldsymbol{x}}, \boldsymbol{z}\rangle) \\ \le \;& \langle (1-\theta)\boldsymbol{x}_* + \theta\widetilde{\boldsymbol{x}}, \boldsymbol{z}\rangle + \langle \boldsymbol{x}, (1-\theta)\boldsymbol{z}_* + \theta\widetilde{\boldsymbol{z}}\rangle \\ \le \;& \langle (1-\theta)\boldsymbol{z}_* + \theta\widetilde{\boldsymbol{x}}, (1-\theta)\boldsymbol{z}_* + \theta\widetilde{\boldsymbol{z}}\rangle + \langle \boldsymbol{x}, \boldsymbol{z}\rangle \\ \le \;& \theta(1-\theta)(\langle \boldsymbol{x}_*, \widetilde{\boldsymbol{z}}\rangle + \langle \widetilde{\boldsymbol{x}}, \boldsymbol{z}_*\rangle) + \theta^2\langle \widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{z}}\rangle + \langle \boldsymbol{x}, \boldsymbol{z}\rangle \\ \le \;& \theta(1-\theta)(1+\gamma_d + L\gamma_p)\rho(\langle \boldsymbol{x}_*, \boldsymbol{e}\rangle + \langle \boldsymbol{e}, \boldsymbol{z}_*\rangle) \\ & +\theta^2(1+\gamma_p)(1+\gamma_d + L\gamma_p)\rho^2 n + (1+\gamma)\nu\mu_0 n \\ \le \;& 6\nu\rho^2 n. \end{aligned}$$

From here, the required results follow. □

Remark. $\{(\boldsymbol{x}, y, \boldsymbol{z}) \mid (\theta, \nu, \boldsymbol{x}, y, \boldsymbol{z}) \in \mathcal{N}\}$ is bounded if $\theta = \nu$, since from Lemma 6 we have $\|\boldsymbol{x}\| \le \mathrm{tr}(\boldsymbol{x}) \le O(\rho n)$ and $\|\boldsymbol{z}\| \le \mathrm{tr}(\boldsymbol{z}) \le O(\rho n)$. Suppose we generate a sequence $\{(\theta_k, \nu_k, \boldsymbol{x}_k, y_k, \boldsymbol{z}_k)\} \in \mathcal{N}$ such that

$$\nu_k \ge \theta_k,\; \forall k, \quad \text{and } 1 = \nu_0 \ge \nu_k \ge \nu_{k+1} \ge 0.$$

If $\nu_k \to 0$ as $k \to \infty$, then any limit point of the sequence $\{\boldsymbol{x}_k, y_k, \boldsymbol{z}_k\}$ is a solution of (P) and (D). In particular, if $\theta_k = \nu_k$, then the sequence $\{\boldsymbol{x}_k, \boldsymbol{z}_k\}$ is also bounded.

4. An inexact infeasible interior-point algorithm

Let $\eta_1, \eta_2 \in (0,1]$ be given constants such that $\eta_1 \geq \eta_2$. Given a current iterate $(\theta_k, \nu_k, \boldsymbol{x}_k, y_k, \boldsymbol{z}_k) \in \mathcal{N}$, we want to construct a new iterate which remains in $\mathcal{N}$ with respect to smaller θ and ν. To this end, we consider the search direction $(\Delta\boldsymbol{x}_k, \Delta y_k, \Delta\boldsymbol{z}_k)$ determined by the following linear system:

(30)
$$\begin{pmatrix} -\mathcal{H} & \mathcal{A}^T & I \\ \mathcal{A} & 0 & 0 \\ E_k & 0 & F_k \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{x}_k \\ \Delta y_k \\ \Delta\boldsymbol{z}_k \end{pmatrix} = \begin{pmatrix} -\eta_1(R_k^d + r_k^d) \\ -\eta_1(R_k^p + r_k^p) \\ R_k^c + r_k^c \end{pmatrix},$$

where for $\boldsymbol{p}_k = \boldsymbol{w}_k^{-1/2}$ ($\boldsymbol{w}_k$ is the NT scaling element satisfying $\mathbf{Q}(\boldsymbol{w}_k)\boldsymbol{z}_k = \boldsymbol{x}_k$),

$$E_k = L(\mathbf{Q}(\boldsymbol{p}_k^{-1})\boldsymbol{z}_k)\mathbf{Q}(\boldsymbol{p}_k), \quad F_k = L(\mathbf{Q}(\boldsymbol{p}_k)\boldsymbol{x}_k)\mathbf{Q}(\boldsymbol{p}_k^{-1}),$$
$$R_k^d = -\nabla f(\boldsymbol{x}_k) + \mathcal{A}^T y_k + \boldsymbol{z}_k, \quad R_k^p = \mathcal{A}(\boldsymbol{x}_k) - b$$
$$R_k^c = (1-\eta_2)\nu_k\mu_0\boldsymbol{e} - H_{\boldsymbol{p}_k}(\boldsymbol{x}_k, \boldsymbol{z}_k).$$

Note that the last equation of (30) is equivalent to

$$H_{\boldsymbol{p}_k}(\boldsymbol{x}_k, \boldsymbol{z}_k) + H_{\boldsymbol{p}_k}(\Delta\boldsymbol{x}_k, \boldsymbol{z}_k) + H_{\boldsymbol{p}_k}(\boldsymbol{x}_k, \Delta\boldsymbol{z}_k)$$
$$= (1-\eta_2)\nu\mu_0\boldsymbol{e} + r_k^c. \tag{31}$$

The search direction $(\Delta\boldsymbol{x}_k, \Delta y_k, \Delta\boldsymbol{z}_k)$ computed from (30) is an "inexact" Newton direction for the perturbed KKT system (22). On the right hand side of (30), R_k^d, R_k^p and R_k^c are the residual components for infeasibilities and complementarity, whereas the vectors r_k^d, r_k^p, r_k^c are the residual components for the inexactness in the computed search direction.

Let $\{\sigma_k\}_{k=1}^{\infty}$ be a given sequence in $(0,1]$ such that $\bar{\sigma} := \sum_{k=0}^{\infty}\sigma_k < \infty$. We require the residual components in the inexactness in (30) to satisfy the following accuracy conditions:

$$\|\mathcal{A}^+ r_k^p\| \leq \gamma_p\rho\theta_k\sigma_k, \ \|r_k^d\| \leq \gamma_d\rho\theta_k\sigma_k, \tag{32}$$
$$\|r_k^c\| \leq 0.5(1-\eta_2)\gamma\nu_k\mu_0.$$

Remark. In practice, we can solve (30) by the following procedure:

1. Compute Δy_k and $\Delta\boldsymbol{x}_k$ from the following augmented system:

$$\begin{bmatrix} -\mathcal{H} - F_k^{-1}E_k & \mathcal{A}^T \\ \mathcal{A} & 0 \end{bmatrix} \begin{bmatrix} \Delta\boldsymbol{x}_k \\ \Delta y_k \end{bmatrix}$$
$$= \begin{bmatrix} -\eta_1(R_k^d + r_k^d) - F_k^{-1}R_k^c \\ -\eta_1(R_k^p + r_k^p) \end{bmatrix} \tag{33a}$$

with the residual vectors r_k^d and r_k^p satisfying the conditions in (32).

2. Compute $\Delta\boldsymbol{z}_k$ from

$$\Delta\boldsymbol{z}_k = -F_k^{-1}E_k\Delta\boldsymbol{x}_k + F_k^{-1}R_k^c. \tag{33b}$$

Here, we can see that $\Delta\boldsymbol{z}_k$ is obtained directly from (31) with $r_k^c = 0$. Thus, r_k^c can be ignored in the system (30). For a convex quadratic SDP problem, the dimension of the augmented system (33a) is n^2+m, which is typically a large number even for $n = 100$. The computational cost and memory requirement for solving (33a) by a direct solver is at least $\Theta((n^2+m)^3)$ and $\Theta((n^2+m)^2)$ respectively, which are prohibitively expensive for large scale problems. An iterative solver would not require the storage or manipulation of the full coefficient matrix. But the disadvantage of using an iterative solver is the demand of good preconditioners to accelerate its convergence. In practice, constructing cheap and effective preconditioners could be the most challenging task in designing an efficient inexact IPM for solving a convex quadratic SDP problem; see [25] for details.

After computing the search direction in (30), we consider the following trial iterate to determine the new iterate:

$$(\theta_k(\alpha), \nu_k(\alpha)) = ((1-\alpha\eta_1)\theta_k, (1-\alpha\eta_2)\nu_k),$$
$$(\boldsymbol{x}_k(\alpha), y_k(\alpha), \boldsymbol{z}_k(\alpha)) = (\boldsymbol{x}_k + \alpha\Delta\boldsymbol{x}_k, y_k + \alpha\Delta y_k, \boldsymbol{z}_k + \alpha\Delta\boldsymbol{z}_k)$$

where $\alpha \in [0,1]$. To find the new iterate, we need to choose an appropriate step length α_k to keep the new iterate in $\mathcal{N}$. The precise choice of α_k will be discussed shortly. Before that, we present our inexact primal-dual infeasible path-following algorithm.

Algorithm IIPF. Let $\epsilon > 0$ be a given tolerance. Let $\theta_0 = \nu_0 = 1$. Choose parameters $\eta_1, \eta_2 \in (0,1]$ with $\eta_1 \geq \eta_2$, $\gamma_p, \gamma_d \in (0,1)$ such that $\gamma_p \leq \gamma_d$ and $\gamma_d + L\gamma_p < 1$. Pick a sequence $\{\sigma_k\}_{k=1}^{\infty}$ in $(0,1]$ such that $\bar{\sigma} := \sum_{k=0}^{\infty}\sigma_k < \infty$. Choose $(\boldsymbol{x}_0, y_0, \boldsymbol{z}_0)$ satisfying (16)–(18). Note that $(\theta_0, \nu_0, \boldsymbol{x}_0, y_0, \boldsymbol{z}_0) \in \mathcal{N}$.

For $k = 0, 1, \ldots$

1. Terminate when $\nu_k < \epsilon$.
2. Find an inexact search direction $(\Delta\boldsymbol{x}_k, \Delta y_k, \Delta\boldsymbol{z}_k)$ from the linear system (30).
3. Let $\alpha_k \in [0,1]$ be chosen appropriately so that

$$(\theta_{k+1}, \nu_{k+1}, \boldsymbol{x}_{k+1}, y_{k+1}, \boldsymbol{z}_{k+1})$$
$$:= (\theta_k(\alpha_k), \nu_k(\alpha_k), \boldsymbol{x}_k(\alpha_k), y_k(\alpha_k), \boldsymbol{z}_k(\alpha_k)) \in \mathcal{N}.$$

Let $\alpha_0, \alpha_1, \ldots, \alpha_{k-1}$ be the step lengths that have already been determined in the previous k iterations. For reasons that will become apparent shortly, we assume that the step lengths α_i, $i = 0, \ldots, k-1$, are contained in the interval

$$\mathcal{I} := \big[0, \min\{1, 1/(\eta_1(1+\bar{\sigma}))\}\big]. \tag{34}$$

Let the primal and dual infeasibilities associated with $(\theta_k(\alpha), \nu_k(\alpha), \boldsymbol{x}_k(\alpha), y_k(\alpha), \boldsymbol{x}_k(\alpha))$ be

$$\begin{aligned} R_k^p(\alpha) &= \mathcal{A}(\boldsymbol{x}_k(\alpha)) - b, \\ R_k^d(\alpha) &= -\nabla f(\boldsymbol{x}_k(\alpha)) + \mathcal{A}^T y_k(\alpha) + \boldsymbol{z}_k(\alpha). \end{aligned}$$

We will show that $R_k^p(\alpha)$ and $R_k^d(\alpha)$ satisfy the first two conditions in $\mathcal{N}$ when α is restricted to be in the interval $\mathcal{I}$ given in (34).

Lemma 7. Suppose the step lengths α_i associated with the iterates $(\theta_i, \nu_i, \boldsymbol{x}_i, y_i, \boldsymbol{z}_i)$ are restricted to be in the interval $\mathcal{I}$ for $i = 0, \ldots, k-1$. Then we have

$$R_k^p(\alpha) = \theta_k(\alpha)(R_0^p + \xi_k^p(\alpha)) \tag{35}$$

$$R_k^d(\alpha) = \theta_k(\alpha)(R_0^d + \xi_k^d(\alpha)) \tag{36}$$

where

$$\|\mathcal{A}^+ \xi_k^p(\alpha)\| \le \gamma_p \rho, \quad \|\xi_k^d(\alpha)\| \le \gamma_d \rho, \quad \forall\, \alpha \in \mathcal{I}.$$

Proof. Note that $R_k^p(\alpha)$ has exactly the same form as in the inexact interior-point algorithm considered in [30] for a linear SDP. Using the result in [30], we have

$$R_k^p(\alpha) = \theta_k(\alpha)(R_0^p + \xi_k^p(\alpha)),$$

where

$$\begin{aligned} \xi_k^p(\alpha) &= \xi_k^p - \frac{\alpha\eta_1}{(1-\alpha\eta_1)\theta_k} r_k^p \\ &= -\sum_{i=0}^{k-1} \frac{\alpha_i \eta_1}{(1-\alpha_i\eta_1)\theta_i} r_i^p - \frac{\alpha\eta_1}{(1-\alpha\eta_1)\theta_k} r_k^p. \end{aligned} \tag{37}$$

The quantity $R_k^d(\alpha)$ is different from its counterpart in a linear SDP as it contains an extra term from the quadratic term in $f(\boldsymbol{x})$. Thus, we need to investigate the details. Given that the current iterate belongs to $\mathcal{N}$, we have

$$\begin{aligned} R_k^d(\alpha) &= -\nabla f(\boldsymbol{x}_k(\alpha)) + \mathcal{A}^T y_k(\alpha) + \boldsymbol{z}_k(\alpha) \\ &= -\nabla f(\boldsymbol{x}_k) + \mathcal{A}^T y_k + \boldsymbol{z}_k + \alpha[-\mathcal{H}(\Delta \boldsymbol{x}_k) + \mathcal{A}^T \Delta y_k + \Delta \boldsymbol{z}_k] \\ &= R_k^d - \alpha\eta_1(R_k^d + r_k^d) \\ &= (1-\alpha\eta_1)\theta_k(R_0^d + \xi_k^d) - \alpha\eta_1 r_k^d \\ &= (1-\alpha\eta_1)\theta_k\Big(R_0^d + \xi_k^d - \frac{\alpha\eta_1}{(1-\alpha\eta_1)\theta_k} r_k^d\Big) \\ &= \theta(\alpha)(R_0^d + \xi_k^d(\alpha)), \end{aligned}$$

where

$$\begin{aligned} \xi_k^d(\alpha) &= \xi_k^d - \frac{\alpha\eta_1}{(1-\alpha\eta_1)\theta_k} r_k^d \\ &= -\sum_{i=0}^{k-1} \frac{\alpha_i \eta_1}{(1-\alpha_i\eta_1)\theta_i} r_i^d - \frac{\alpha\eta_1}{(1-\alpha\eta_1)\theta_k} r_k^d. \end{aligned} \tag{38}$$

From (37) and (38), we see that since $\alpha_i \le \frac{1}{\eta_1(1+\bar{\sigma})}$ for $i = 1, \ldots, k-1$, we have

$$\|\mathcal{A}^+ \xi_k^p(\alpha)\| \le \gamma_p \rho, \quad \|\xi_k^d(\alpha)\| \le \gamma_d \rho, \quad \forall\, \alpha \in \mathcal{I}.$$

□

Let

$$\bar{\alpha}_k = \min\Big\{1, \frac{1}{\eta_1(1+\bar{\sigma})}, \frac{0.5(1-\eta_2)\gamma\nu_k\mu_0}{\|H_{\boldsymbol{p}_k}(\Delta \boldsymbol{x}_k, \Delta \boldsymbol{z}_k)\|}\Big\}. \tag{39}$$

Next, we check the last condition in $\mathcal{N}$. The following lemma generalizes the result of Lemma 4.2 in [29].

Lemma 8. For $(\theta_k, \nu_k, \boldsymbol{x}_k, y_k, \boldsymbol{z}_k) \in \mathcal{N}$ and $\Delta \boldsymbol{x}_k, \Delta \boldsymbol{z}_k$ satisfying (30), we have
(a)

$$\begin{aligned} &H_{\boldsymbol{p}_k}(\boldsymbol{x}_k(\alpha), \boldsymbol{z}_k(\alpha)) = (1-\alpha) H_{\boldsymbol{p}_k}(\boldsymbol{x}_k, \boldsymbol{z}_k) \\ &\quad + \alpha(1-\eta_2)\nu_k\mu_0 \boldsymbol{e} + \alpha r_k^c + \alpha^2 H_{\boldsymbol{p}_k}(\Delta \boldsymbol{x}_k, \Delta \boldsymbol{z}_k) \end{aligned}$$

(b) For all $\alpha \in [0, \bar{\alpha}_k]$,

$$(1-\gamma)\nu_k(\alpha)\mu_0 \le \lambda_i\left(\mathbf{Q}(\boldsymbol{x}_k(\alpha)^{1/2})\boldsymbol{z}_k(\alpha)\right) \le (1+\gamma)\nu_k(\alpha)\mu_0.$$

(a) For all $\alpha \in [0, \bar{\alpha}_k]$, $\boldsymbol{x}_k(\alpha) \succ 0$ and $\boldsymbol{z}_k(\alpha) \succ 0$.

Proof. (a) The proof of part (a) is quite standard and uses equation (31).
(b) First, we note that

$$\begin{aligned} &\lambda_{\min}\left(H_{\boldsymbol{p}_k}(\boldsymbol{x}_k(\alpha), \boldsymbol{z}_k(\alpha))\right) \\ &\ge (1-\alpha)(1-\gamma)\nu_k\mu_0 + \alpha(1-\eta_2)\nu_k\mu_0 - \alpha\|r_k^c\| \\ &\quad -\alpha^2 \|H_{\boldsymbol{p}_k}(\Delta \boldsymbol{x}_k, \Delta \boldsymbol{z}_k)\| \\ &= \alpha\gamma(1-\eta_2)\nu_k\mu_0 - \alpha\|r_k^c\| - \alpha^2\|H_{\boldsymbol{p}_k}(\Delta \boldsymbol{x}_k, \Delta \boldsymbol{z}_k)\| \\ &\quad +(1-\gamma)\nu_k(\alpha)\mu_0 \\ &\ge 0.5\alpha(1-\eta_2)\gamma\nu_k\mu_0 - \alpha^2\|H_{\boldsymbol{p}_k}(\Delta \boldsymbol{x}_k, \Delta \boldsymbol{z}_k)\| \\ &\quad +(1-\gamma)\nu_k(\alpha)\mu_0 \\ &\ge (1-\gamma)\nu_k(\alpha)\mu_0 \quad \text{for } \alpha \in [0\,,\, \bar{\alpha}_k]. \end{aligned} \tag{40}$$

Now let $\widehat{\boldsymbol{x}}_k(\alpha) = \mathbf{Q}(\boldsymbol{p}_k)\boldsymbol{x}_k(\alpha)$ and $\widehat{\boldsymbol{z}}_k(\alpha) = \mathbf{Q}(\boldsymbol{p}_k^{-1})\boldsymbol{z}_k(\alpha)$. From Lemma 3, we have

$$\begin{aligned} &\lambda_{\min}\left(\mathbf{Q}(\boldsymbol{x}_k(\alpha)^{1/2})\boldsymbol{z}_k(\alpha)\right) = \lambda_{\min}\left(\mathbf{Q}(\widehat{\boldsymbol{x}}_k(\alpha)^{1/2})\widehat{\boldsymbol{z}}_k(\alpha)\right) \\ &\ge \lambda_{\min}(\widehat{\boldsymbol{x}}_k(\alpha) \circ \widehat{\boldsymbol{z}}_k(\alpha)). \end{aligned}$$

By noting that $H_{\boldsymbol{p}_k}(\boldsymbol{x}_k(\alpha), \boldsymbol{z}_k(\alpha)) = \widehat{\boldsymbol{x}}_k(\alpha) \circ \widehat{\boldsymbol{z}}_k(\alpha)$ and using (40), we have

$$\lambda_{\min}\left(\mathbf{Q}(\boldsymbol{x}_k(\alpha)^{1/2})\boldsymbol{z}_k(\alpha)\right) \ge (1-\gamma)\nu_k(\alpha)\mu_0 \;\forall\, \alpha \in [0\,,\, \bar{\alpha}_k].$$

The proof that $\lambda_{\max}\left(\mathbf{Q}(\boldsymbol{x}_k(\alpha)^{1/2})\boldsymbol{z}_k(\alpha)\right) \le (1+\gamma)\nu_k(\alpha)\mu_0$ for all $\alpha \in [0\,,\, \bar{\alpha}_k]$ is similar, and we shall omit it.
(c) Let $\widehat{\boldsymbol{x}}_k(\alpha), \widehat{\boldsymbol{z}}(\alpha)$ be defined as in part (b). Since $\boldsymbol{p}_k \succ 0$, we have that $\mathbf{Q}(\boldsymbol{p}_k)\boldsymbol{x} \succ 0$ if and only if $\boldsymbol{x} \succ 0$. Thus it is sufficient for us to prove that $\widehat{\boldsymbol{x}}_k(\alpha) \succ 0$ for all $\alpha \in [0, \bar{\alpha}_k]$. Suppose that it is not. Then by the continuity of $\lambda_{\min}(\cdot)$, there exist $\alpha^* \in [0, \bar{\alpha}_k]$ such that $\lambda_{\min}(\widehat{\boldsymbol{x}}_k(\alpha^*)) = 0$ and $\widehat{\boldsymbol{x}}_k(\alpha) \succeq 0$ for all $\alpha \in [0, \alpha^*]$. Let $\boldsymbol{v}^* = \mathbf{Q}(\widehat{\boldsymbol{x}}_k(\alpha^*)^{1/2})\widehat{\boldsymbol{z}}_k(\alpha^*)$. From the proof of part (b), we have that $\lambda_{\min}(\boldsymbol{v}^*) \ge (1-\gamma)\nu_k(\alpha^*)\mu_0 > 0$ and hence $\det(\boldsymbol{v}^*) > 0$. But from Lemma

2, we have $\det(\boldsymbol{v}^*) = \det(\widehat{\boldsymbol{x}}_k(\alpha^*))\det(\widehat{\boldsymbol{z}}_k(\alpha^*)) = 0$. Thus we get a contradiction, and hence $\boldsymbol{x}_k(\alpha) \succ 0$ for all $\alpha \in [0, \bar{\alpha}_k]$. The proof that $\boldsymbol{z}_k(\alpha) \succ 0$ for all $\alpha \in [0, \bar{\alpha}_k]$ is similar. □

Lemma 9. Under the conditions in Lemmas 7 and 8, for any $\alpha \in [0, \bar{\alpha}_k]$, we have

$$(\theta(\alpha), \nu(\alpha), \boldsymbol{x}(\alpha), y(\alpha), \boldsymbol{z}(\alpha)) \in \mathcal{N}.$$

Proof. The result follows from Lemmas 7 and 8. □

Lemma 10. Suppose the conditions in (16), (17) and (18) hold. Then

$$\|H_{\boldsymbol{p}_k}(\Delta\boldsymbol{x}_k, \Delta\boldsymbol{z}_k)\| = \frac{O(1)}{(1-(\gamma_d + L\gamma_p))^2} n^2 \nu_k \mu_0. \tag{41}$$

The proof of Lemma 10 is non-trivial and we devote the next section to its proof.

We are now ready to present the main result of this paper, the polynomial iteration complexity of **Algorithm** IIPF.

Theorem 3. Let $\epsilon > 0$ be a given tolerance. Suppose the conditions in (16), (17) and (18) hold. At each iteration of **Algorithm** IIPF, set the step length $\alpha_k = \bar{\alpha}_k$. Then $\nu_k \leq \epsilon$ for $k = O(n^2 \ln(1/\epsilon))$.

Proof. From (39), Lemma 9 and Lemma 10, we know that

$$\alpha_i \geq \bar{\alpha} := \min\left\{1, \frac{1}{\eta_1(1+\bar{\sigma})}, \frac{O(1)}{n^2}\right\}, \quad i = 0, \ldots, k.$$

Then we have

$$\nu_k = \prod_{i=0}^{k-1}(1-\alpha_i\eta_2) \leq (1-\bar{\alpha}\eta_2)^k \leq \varepsilon \text{ for } k = O(n^2\ln(1/\varepsilon)).$$

□

5. Proof of Lemma 10

For a given $(\theta_k, \nu_k, \boldsymbol{x}_k, y_k, \boldsymbol{z}_k) \in \mathcal{N}$, the purpose of Lemma 10 is to establish an upper bound for $\|H_{\boldsymbol{p}_k}(\Delta\boldsymbol{x}_k, \Delta\boldsymbol{z}_k)\|$. Throughout this section, we shall consider only the NT direction, where $\boldsymbol{p}_k = \boldsymbol{w}_k^{-1/2}$, with $\boldsymbol{w}_k$ satisfying $\mathbf{Q}(\boldsymbol{w}_k)\boldsymbol{z}_k = \boldsymbol{x}_k$.

To facilitate our analysis, we introduce the following notation:

$$\widehat{\boldsymbol{x}}_k = \mathbf{Q}(\boldsymbol{p}_k)\boldsymbol{x}_k, \quad \widehat{\boldsymbol{z}}_k = \mathbf{Q}(\boldsymbol{p}_k^{-1})\boldsymbol{z}_k;$$

$$\Delta\widehat{\boldsymbol{x}}_k = \mathbf{Q}(\boldsymbol{p}_k)\Delta\boldsymbol{x}_k, \quad \Delta\widehat{\boldsymbol{z}}_k = \mathbf{Q}(\boldsymbol{p}_k^{-1})\Delta\boldsymbol{z}_k;$$

$$\widehat{E}_k = E_k\mathbf{Q}(\boldsymbol{p}_k^{-1}) = L(\widehat{\boldsymbol{z}}_k), \quad \widehat{F}_k = F_k\mathbf{Q}(\boldsymbol{p}_k) = L(\widehat{\boldsymbol{x}}_k).$$

From (14) we have

$$\widehat{\boldsymbol{z}}_k = \widehat{\boldsymbol{x}}_k, \quad \widehat{E}_k = \widehat{F}_k. \tag{42}$$

Let the spectral decomposition of $\widehat{\boldsymbol{x}}_k$ and $\widehat{\boldsymbol{z}}_k$ be

$$\widehat{\boldsymbol{x}}_k = \widehat{\boldsymbol{z}}_k = \lambda_1^k \boldsymbol{c}_1^k + \ldots + \lambda_r^k \boldsymbol{c}_r^k. \tag{43}$$

From (23), we have

$$(1-\gamma)\nu_k\mu_0 \leq (\lambda_1^k)^2 \leq \cdots \leq (\lambda_r^k)^2 \leq (1+\gamma)\nu_k\mu_0. \tag{44}$$

Let $\widehat{S}_k := \widehat{F}_k\widehat{E}_k = (\widehat{E}_k)^2$. From Lemma 1, we know that the eigenvalues of $\widehat{S}_k$ are given by

$$\Lambda(\widehat{S}_k) = \left\{\frac{1}{4}(\lambda_i^k + \lambda_j^k)^2 : 1 \leq i, j \leq r\right\}.$$

From (44), we have

$$\|\widehat{S}_k\|_2 \;\leq\; (1+\gamma)\nu_k\mu_0, \quad \|\widehat{S}_k^{-1}\|_2 \;\leq\; \frac{1}{(1-\gamma)\nu_k\mu_0}. \tag{45}$$

Now we state a few lemmas, which will lead to the proof of Lemma 10.

Lemma 11. For any $\boldsymbol{u} \in \mathcal{J}$,

$$\|\mathbf{Q}(\boldsymbol{p}_k)\boldsymbol{u}\|^2 \;\leq\; \frac{1}{(1-\gamma)\nu_k\mu_0}\|\mathbf{Q}(\boldsymbol{z}_k)\|_2\|\boldsymbol{u}\|^2,$$

$$\|\mathbf{Q}(\boldsymbol{p}_k^{-1})\boldsymbol{u}\|^2 \;\leq\; \frac{1}{(1-\gamma)\nu_k\mu_0}\|\mathbf{Q}(\boldsymbol{x}_k)\|_2\|\boldsymbol{u}\|^2.$$

Proof. From Lemma 3, we know that $\mathbf{Q}(\boldsymbol{z}_k^{1/2})\boldsymbol{x}_k$, $\mathbf{Q}(\boldsymbol{x}_k^{1/2})\boldsymbol{z}_k$, $\mathbf{Q}(\widehat{\boldsymbol{x}}_k^{1/2})\widehat{\boldsymbol{z}}_k$ and $H_{\boldsymbol{p}_k}(\boldsymbol{x}_k, \boldsymbol{z}_k)$ all have the same spectrum. Thus we have

$$\begin{aligned}(1-\gamma)\nu_k\mu_0 \leq \lambda_{\min}\left(\mathbf{Q}(\boldsymbol{z}_k^{1/2})\boldsymbol{x}_k\right) &\leq \lambda_{\max}\left(\mathbf{Q}(\boldsymbol{z}_k^{1/2})\boldsymbol{x}_k\right) \\ &\leq (1+\gamma)\nu_k\mu_0\end{aligned} \tag{46}$$

$$\begin{aligned}(1-\gamma)\nu_k\mu_0 \leq \lambda_{\min}\left(\mathbf{Q}(\boldsymbol{x}_k^{1/2})\boldsymbol{z}_k\right) &\leq \lambda_{\max}\left(\mathbf{Q}(\boldsymbol{x}_k^{1/2})\boldsymbol{z}_k\right) \\ &\leq (1+\gamma)\nu_k\mu_0\end{aligned} \tag{47}$$

Let $\boldsymbol{v}_k = \mathbf{Q}(\boldsymbol{z}_k^{1/2})\boldsymbol{x}_k$. By (13), we have $\boldsymbol{p}_k^{-2} = \mathbf{Q}(\boldsymbol{z}_k^{-1/2})\boldsymbol{v}_k^{1/2}$. By (46) , we have

$$\begin{aligned}
&\|\mathbf{Q}(\boldsymbol{p}_k)\boldsymbol{u}\|^2 = \langle \boldsymbol{u},\, \mathbf{Q}(\boldsymbol{p}_k^2)\boldsymbol{u}\rangle \\
=&\; \langle \boldsymbol{u},\, \left[\mathbf{Q}(\boldsymbol{z}_k^{-1/2})\mathbf{Q}(\boldsymbol{v}_k^{1/2})\mathbf{Q}(\boldsymbol{z}_k^{-1/2})\right]^{-1}\boldsymbol{u}\rangle \text{ (Lem. 2, part 1.)} \\
=&\; \langle \boldsymbol{u},\, \mathbf{Q}(\boldsymbol{z}_k^{1/2})\mathbf{Q}(\boldsymbol{v}_k^{-1/2})\mathbf{Q}(\boldsymbol{z}_k^{1/2})\boldsymbol{u}\rangle \\
=&\; \langle \mathbf{Q}(\boldsymbol{z}_k^{1/2})\boldsymbol{u},\, \mathbf{Q}(\boldsymbol{v}_k)^{-1/2}\mathbf{Q}(\boldsymbol{z}_k^{1/2})\boldsymbol{u}\rangle \\
\leq&\; \lambda_{\max}(\mathbf{Q}(\boldsymbol{v}_k)^{-1/2})\|\mathbf{Q}(\boldsymbol{z}_k^{1/2})\boldsymbol{u}\|^2 \\
\leq&\; \frac{1}{(1-\gamma)\nu_k\mu_0}\|\mathbf{Q}(\boldsymbol{z}_k^{1/2})\boldsymbol{u}\|^2 \\
\leq&\; \frac{1}{(1-\gamma)\nu_k\mu_0}\|\mathbf{Q}(\boldsymbol{z}_k)\|_2\|\boldsymbol{u}\|^2.
\end{aligned}$$

The proof of the second inequality in the Lemma is similar and we shall omit it. □

Lemma 12.

$$\|\Delta\widehat{\boldsymbol{x}}_k\|^2+\|\Delta\widehat{\boldsymbol{z}}_k\|^2+2\langle\Delta\widehat{\boldsymbol{x}}_k,\,\Delta\widehat{\boldsymbol{z}}_k\rangle=\|\widehat{S}_k^{-1/2}(R_k^c+r_k^c)\|^2,$$

$$\|H_{\boldsymbol{p}_k}(\Delta\boldsymbol{x}_k,\Delta\boldsymbol{z}_k)\|\le\tfrac{1}{2}\Big(\|\Delta\widehat{\boldsymbol{x}}_k\|^2+\|\Delta\widehat{\boldsymbol{z}}_k\|^2\Big).$$

Proof. The last equation of (30) can be rewritten as

$$\widehat{E}_k(\Delta\widehat{\boldsymbol{x}}_k)+\widehat{F}_k(\Delta\widehat{\boldsymbol{z}}_k)=R_k^c+r_k^c. \tag{48}$$

Multiplying (48) by $\widehat{S}_k^{-1/2}$ from the left, we have

$$\Delta\widehat{\boldsymbol{x}}_k+\Delta\widehat{\boldsymbol{z}}_k=\widehat{S}_k^{-1/2}(R_k^c+r_k^c).$$

From here, the first equation in the lemma follows.

For the second inequality, by Lemma 1, we have

$$\begin{aligned}\|H_{\boldsymbol{p}_k}(\Delta\boldsymbol{x}_k,\Delta\boldsymbol{z}_k)\|&=\|\Delta\widehat{\boldsymbol{x}}_k\circ\Delta\widehat{\boldsymbol{z}}_k\|\\ &\le\|\Delta\widehat{\boldsymbol{x}}_k\|\|\Delta\widehat{\boldsymbol{z}}_k\|\le\frac{1}{2}\Big(\|\Delta\widehat{\boldsymbol{x}}_k\|^2+\|\Delta\widehat{\boldsymbol{z}}_k\|^2\Big).\end{aligned}$$

□

Lemma 13. We have

$$\|\widehat{S}_k^{-1/2}(R_k^c+r_k^c)\|^2=O(n\nu_k\mu_0).$$

Proof. From (32) and (45), we have

$$\begin{aligned}\|\widehat{S}_k^{-1/2}r_k^c\|^2&\le\|\widehat{S}_k^{-1}\|_2\|r_k^c\|^2\le\frac{0.25[(1-\eta_2)\gamma\nu_k\mu_0]^2}{(1-\gamma)\nu_k\mu_0}\\ &=\frac{[(1-\eta_2)\gamma]^2\nu_k\mu_0}{4(1-\gamma)}.\end{aligned}\tag{49}$$

Observe that from (43),

$$\Lambda(R_k^c)=(1-\eta_2)\nu_k\mu_0-\Lambda(\widehat{\boldsymbol{x}}_k)^2.$$

Thus

$$\begin{aligned}\|\widehat{S}_k^{-1/2}R_k^c\|^2&\le\|\widehat{S}_k^{-1}\|_2\|R_k^c\|^2\\ &\le\frac{1}{(1-\gamma)\nu_k\mu_0}\sum_{i=1}^r\Big((1-\eta_2)\nu_k\mu_0-(\lambda_i^k)^2\Big)^2\\ &\le\frac{n\nu_k\mu_0}{1-\gamma}(\gamma+\eta_2)^2\quad\text{(by (44))}.\end{aligned}\tag{50}$$

The required result follows from (49) and (50). This completes the proof. □

In the rest of our analysis, we introduce an auxiliary point $(\widetilde{\boldsymbol{x}}_k,\widetilde{y}_k,\widetilde{\boldsymbol{z}}_k)$ whose existence is ensured by Lemma 5. From Lemma 7, we have the following equations at the kth iteration:

$$-\nabla f(\boldsymbol{x}_k)+\mathcal{A}^Ty_k+\boldsymbol{z}_k=\theta_k(R_0^d+\xi_k^d),\ \|\xi_k^d\|\le\gamma_d\rho, \tag{51}$$

$$\mathcal{A}(\boldsymbol{x}_k)-b=\theta_k(R_0^p+\xi_k^p),\ \|\mathcal{A}^+\xi_k^p\|\le\gamma_p\rho. \tag{52}$$

Thus by Lemma 5, there exists $(\widetilde{\boldsymbol{x}}_k,\widetilde{y}_k,\widetilde{\boldsymbol{z}}_k)$ such that

$$-\nabla f(\widetilde{\boldsymbol{x}}_k)+\mathcal{A}^T\widetilde{y}_k+\widetilde{\boldsymbol{z}}_k=R_0^d+\xi_k^d \tag{53}$$

$$\mathcal{A}(\widetilde{\boldsymbol{x}}_k)-b=R_0^p+\xi_k^p \tag{54}$$

$$(1-\gamma_p)\rho\boldsymbol{e}\preceq\widetilde{\boldsymbol{x}}_k\preceq(1+\gamma_p)\rho\boldsymbol{e}\ , \tag{55}$$

$$[1-(\gamma_d+L\gamma_p)]\rho\boldsymbol{e}\preceq\widetilde{\boldsymbol{z}}_k\preceq[1+(\gamma_d+L\gamma_p)]\rho\boldsymbol{e}. \tag{56}$$

Lemma 14. Let

$$\bar{\boldsymbol{x}}_k=\boldsymbol{x}_k-\boldsymbol{x}_*-\theta_k(\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*),\quad\bar{\boldsymbol{z}}_k=\boldsymbol{z}_k-\boldsymbol{z}_*-\theta_k(\widetilde{\boldsymbol{z}}_k-\boldsymbol{z}_*).$$

The following equations hold:

$$\langle\bar{\boldsymbol{x}}_k,\,\bar{\boldsymbol{z}}_k\rangle=\langle\bar{\boldsymbol{x}}_k,\,\mathcal{H}\bar{\boldsymbol{x}}_k\rangle, \tag{57}$$

$$\begin{aligned}&\langle\Delta\boldsymbol{x}_k+\eta_1\theta_k(\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*)+\eta_1\mathcal{A}^+r_k^p,\,\Delta\boldsymbol{z}_k+\eta_1\theta_k(\widetilde{\boldsymbol{z}}_k-\boldsymbol{z}_*)+\eta_1r_k^d\rangle\\ &=\langle\Delta\boldsymbol{x}_k+\eta_1\theta_k(\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*)+\eta_1\mathcal{A}^+r_k^p,\,\mathcal{H}\,(\Delta\boldsymbol{x}_k+\eta_1\theta_k(\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*))\rangle.\end{aligned}\tag{58}$$

Proof. By (51)–(54) and the fact that

$$\begin{aligned}\mathcal{A}\boldsymbol{x}_*-b&=0,\\ -\nabla f(\boldsymbol{x}_*)+\mathcal{A}^Ty_*+\boldsymbol{z}_*&=0,\end{aligned}$$

we have

$$\begin{aligned}\mathcal{A}\bar{\boldsymbol{x}}_k&=0\\ \mathcal{A}^T(y_k-y_*-\theta_k(\widetilde{y}_k-y_*))+\bar{\boldsymbol{z}}_k&=\mathcal{H}\,(\bar{\boldsymbol{x}}_k),\end{aligned}$$

which implies (57). Next, by (30), and (51)–(54), we have

$$\begin{aligned}&\mathcal{A}\Big(\Delta\boldsymbol{x}_k+\eta_1\theta_k(\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*)+\eta_1\mathcal{A}^+r_k^p\Big)=0\\ &\mathcal{A}^T\Big(\Delta y_k+\eta_1\theta_k(\widetilde{y}_k-y_*)\Big)+(\Delta\boldsymbol{z}_k+\eta_1\theta_k(\widetilde{\boldsymbol{z}}_k-\boldsymbol{z}_*))\\ &+\eta_1r_k^d\ =\ \mathcal{H}\,(\Delta\boldsymbol{x}_k+\eta_1\theta_k(\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*)),\end{aligned}$$

which implies (58). □

Let

$$T_1=\Big(\|\Delta\widehat{\boldsymbol{x}}_k\|^2+\|\Delta\widehat{\boldsymbol{z}}_k\|^2\Big)^{1/2} \tag{59}$$

$$T_2=\Big(\|\mathbf{Q}(\boldsymbol{p}_k)(\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*)\|^2+\|\mathbf{Q}(\boldsymbol{p}_k^{-1})(\widetilde{\boldsymbol{z}}_k-\boldsymbol{z}_*)\|^2\Big)^{1/2} \tag{60}$$

$$T_3=\Big(\|\mathbf{Q}(\boldsymbol{p}_k)\mathcal{A}^+r_k^p\|^2+\|\mathbf{Q}(\boldsymbol{p}_k^{-1})r_k^d\|^2\Big)^{1/2} \tag{61}$$

$$T_4=\|\mathbf{Q}(\boldsymbol{p}_k^{-1})\mathcal{H}\,(\mathcal{A}^+r_k^p)\|. \tag{62}$$

Then we have the following lemma.

Lemma 15.

$$T_1\ \le\ 2\eta_1(\theta_kT_2+T_3+T_4)+\sqrt{T_5},$$

where

$$\begin{aligned}T_5&=\|\widehat{S}_k^{-1/2}(R_k^c+r_k^c)\|^2\\ &+2\eta_1^2\theta_k^2\langle\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*,\,\widetilde{\boldsymbol{z}}_k-\boldsymbol{z}_*\rangle+2\eta_1^2\Big(\theta_kT_2T_3+T_3^2+\theta_kT_2T_4\Big).\end{aligned}$$

Proof. By (58), we have that

$$
\begin{aligned}
-\langle \Delta\widehat{\boldsymbol{x}}_k,\, \Delta\widehat{\boldsymbol{z}}_k\rangle \;&=\; -\langle \Delta\boldsymbol{x}_k,\, \Delta\boldsymbol{z}_k\rangle \\
=\;& \eta_1\theta_k[\langle \Delta\boldsymbol{x}_k,\, \widetilde{\boldsymbol{z}}_k-\boldsymbol{z}_*\rangle + \langle \widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*,\, \Delta\boldsymbol{z}_k\rangle] \\
&+\eta_1[\langle \Delta\boldsymbol{x}_k,\, r_k^d\rangle + \langle \mathcal{A}^+r_k^p,\, \Delta\boldsymbol{z}_k\rangle] \\
&+\eta_1^2\theta_k[\langle \widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*,\, r_k^d\rangle + \langle \mathcal{A}^+r_k^p,\, \widetilde{\boldsymbol{z}}_k-\boldsymbol{z}_*\rangle] \\
&+\eta_1^2\langle \mathcal{A}^+r_k^p,\, r_k^d\rangle + \eta_1^2\theta_k^2\langle \widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*,\, \widetilde{\boldsymbol{z}}_k-\boldsymbol{z}_*\rangle \\
&-\eta_1\langle \mathcal{A}^+r_k^p,\, \mathcal{H}\,(\Delta\boldsymbol{x}_k+\eta_1\theta_k(\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*))\rangle \\
-\langle \Delta\boldsymbol{x}_k&+\eta_1\theta_k(\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*),\, \mathcal{H}\,(\Delta\boldsymbol{x}_k+\eta_1\theta_k(\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*))\rangle.
\end{aligned}
$$

Also, we have the following inequalities:

$$
\begin{aligned}
&|\langle \Delta\boldsymbol{x}_k,\, \widetilde{\boldsymbol{z}}_k-\boldsymbol{z}_*\rangle + \langle \widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*,\, \Delta\boldsymbol{z}_k\rangle| \\
&=\; |\langle \Delta\widehat{\boldsymbol{x}}_k,\, \mathbf{Q}(\boldsymbol{p}_k^{-1})(\widetilde{\boldsymbol{z}}_k-\boldsymbol{z}_*)\rangle + \langle \mathbf{Q}(\boldsymbol{p}_k)(\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*),\, \Delta\widehat{\boldsymbol{z}}_k\rangle| \\
&\le\; T_1T_2 \\
&|\langle \Delta\boldsymbol{x}_k,\, r_k^d\rangle + \langle \mathcal{A}^+r_k^p,\, \Delta\boldsymbol{z}_k\rangle| \le T_1T_3 \\
&|\langle \widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*,\, r_k^d\rangle + \langle \mathcal{A}^+r_k^p,\, \widetilde{\boldsymbol{z}}_k-\boldsymbol{z}_*\rangle| \le T_2T_3 \\
&|\langle \mathcal{A}^+r_k^p,\, r_k^d\rangle| \le T_3^2 \\
&|\langle \mathcal{A}^+r_k^p,\, \mathcal{H}\,(\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*)\rangle| \le T_2T_4 \\
&|\langle \mathcal{A}^+r_k^p,\, \mathcal{H}\,(\Delta\boldsymbol{x}_k)\rangle| \le T_1T_4 \\
&-\langle \Delta\boldsymbol{x}_k+\eta_1\theta_k(\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*),\, \mathcal{H}\,(\Delta\boldsymbol{x}_k+\eta_1\theta_k(\widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*))\rangle \le 0.
\end{aligned}
$$

In the above, we used the Cauchy-Schwartz inequality and the fact that $ac+bd \le \sqrt{a^2+b^2}\sqrt{c^2+d^2}$ for $a,b,c,d \ge 0$.

By Lemma 12, and the above inequalities, we have

$$
\begin{aligned}
T_1^2 \;&=\; \|\widehat{S}_k^{-1/2}(R_k^c+r_k^c)\|^2 - 2\langle \Delta\widehat{\boldsymbol{x}}_k,\, \Delta\widehat{\boldsymbol{z}}_k\rangle \\
&\le\; 2\big(\eta_1\theta_kT_1T_2+\eta_1T_1T_3+\eta_1^2\theta_kT_2T_3+\eta_1^2T_3^2\big) \\
&\quad +2\big(\eta_1^2\theta_kT_2T_4+\eta_1T_1T_4\big) \\
&\quad +\|\widehat{S}_k^{-1/2}(R_k^c+r_k^c)\|^2 + 2\eta_1^2\theta_k^2\langle \widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*,\, \widetilde{\boldsymbol{z}}_k-\boldsymbol{z}_*\rangle \\
&=\; 2\eta_1T_1(\theta_kT_2+T_3+T_4)+T_5.
\end{aligned}
$$

The quadratic function $t^2-2\eta_1(\theta_kT_2+T_3+T_4)t-T_5$ has a unique positive root at

$$
t_+ = \eta_1(\theta_kT_2+T_3+T_4)+\sqrt{\eta_1^2(\theta_kT_2+T_3+T_4)^2+T_5},
$$

and it is positive for $t > t_+$, hence we must have $T_1 \le t_+ \le 2\eta_1(\theta_kT_2+T_3+T_4)+\sqrt{T_5}$. ☐

Lemma 16. We have

$$
T_3^2 \;=\; \frac{O(1)}{(1-(\gamma_d+L\gamma_p))^2}n^2\nu_k\mu_0.
$$

Proof. By (32), we have

$$
\|\mathcal{A}^+r_k^p\| \le \theta_k\gamma_p\rho, \quad \|r_k^d\| \le \theta_k\gamma_d\rho. \tag{63}
$$

By Lemma 11 and the fact that $\|\mathbf{Q}(\boldsymbol{u})\|_2 = \lambda_{\max}^2(\boldsymbol{u}) \le [\mathrm{tr}(\boldsymbol{u})]^2$ for any $\boldsymbol{u} \succeq 0$, we have

$$
\begin{aligned}
\|\mathbf{Q}(\boldsymbol{p}_k)\mathcal{A}^+r_k^p\|^2 \;&\le\; \frac{1}{(1-\gamma)\nu_k\mu_0}\|\mathbf{A}^+r_k^p\|^2\|\mathbf{Q}(\boldsymbol{z}_k)\|_2 \\
&\le\; \frac{\gamma_p^2\rho^2}{(1-\gamma)\nu_k\mu_0}\theta_k^2\|\mathbf{Q}(\boldsymbol{z}_k)\|_2 \le \frac{\gamma_p^2\rho^2}{(1-\gamma)\nu_k\mu_0}\theta_k^2[\mathrm{tr}(\boldsymbol{z}_k)]^2 \\
&=\; \frac{\gamma_p^2\rho^2}{(1-\gamma)\nu_k\mu_0}\frac{36}{(1-\gamma_p)^2}n^2\nu_k^2\rho^2 \;=\; \frac{O(1)}{(1-\gamma_p)^2}n^2\nu_k\mu_0.
\end{aligned}
$$

Note that the last equality follows from Lemma 6. Similarly, we have

$$
\begin{aligned}
\|\mathbf{Q}(\boldsymbol{p}_k^{-1})r_k^d\|^2 \;&\le\; \frac{1}{(1-\gamma)\nu_k\mu_0}\|r_k^d\|^2\|\mathbf{Q}(\boldsymbol{x}_k)\|_2 \\
&\le\; \frac{\gamma_d^2\rho^2}{(1-\gamma)\nu_k\mu_0}\theta_k^2\|\mathbf{Q}(\boldsymbol{x}_k)\|_2 \le \frac{\gamma_d^2\rho^2}{(1-\gamma)\nu_k\mu_0}\theta_k^2[\mathrm{Tr}(\boldsymbol{x}_k)]^2 \\
&=\; \frac{\gamma_d^2\rho^2}{(1-\gamma)\nu_k\mu_0}\frac{36}{(1-(\gamma_d+L\gamma_p))^2}n^2\nu_k^2\rho^2 \\
&=\; \frac{O(1)}{(1-(\gamma_d+L\gamma_p))^2}n^2\nu_k\mu_0.
\end{aligned}
$$

Again, we used Lemma 6. From here, the required result follows. ☐

Lemma 17. Under the conditions (16), (17) and (18),

$$
\langle \widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_*,\, \widetilde{\boldsymbol{z}}_k-\boldsymbol{z}_*\rangle \;\le\; 4n\mu_0.
$$

Proof. The result follows from Lemma 11 in [30] and (55)–(56). ☐

Lemma 18. Under the conditions (16), (17), and (18),

$$
\theta_k^2T_2^2 \;=\; O(n^2\nu_k\mu_0).
$$

Proof. First we note that for any invertible $\boldsymbol{u} \in \mathcal{J}$, $\mathbf{Q}(\boldsymbol{u})\boldsymbol{x} \succ 0$ if $\boldsymbol{x} \succ 0$ [5, Prop. III.2.2]. Then for any $\boldsymbol{x} \succ 0$

$$
\|\mathbf{Q}(\boldsymbol{u})\boldsymbol{x}\| \le \mathrm{tr}(\mathbf{Q}(\boldsymbol{u})\boldsymbol{x}) = \langle \mathbf{Q}(\boldsymbol{u})\boldsymbol{x},\, \boldsymbol{e}\rangle = \langle \boldsymbol{u}^2,\, \boldsymbol{x}\rangle. \tag{64}
$$

Let $\boldsymbol{v}_k = \mathbf{Q}(\boldsymbol{z}_k^{1/2})\boldsymbol{x}_k$. By (13), we have $\boldsymbol{p}_k^{-2} = \mathbf{Q}(\boldsymbol{z}_k^{-1/2})\boldsymbol{v}_k^{1/2}$. By the fact that $0 \prec \widetilde{\boldsymbol{x}}_k-\boldsymbol{x}_* \preceq (1+\gamma_p)\rho\boldsymbol{e}$, and (64), we have

$$
\begin{aligned}
& \|\mathbf{Q}(\boldsymbol{p}_k)(\widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_*)\| \;\le\; \langle \boldsymbol{p}_k^2,\, \widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_* \rangle \\
= \;& \langle \mathbf{Q}(\boldsymbol{p}_k^{-2})\boldsymbol{p}_k^2,\, \mathbf{Q}(\boldsymbol{p}_k^2)(\widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_*) \rangle \\
= \;& \langle \boldsymbol{p}_k^{-2},\, (\mathbf{Q}(\boldsymbol{p}_k^{-2}))^{-1}(\widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_*) \rangle \\
= \;& \langle \mathbf{Q}(\boldsymbol{z}_k^{-1/2})\boldsymbol{v}_k^{1/2},\, \left[\mathbf{Q}(\boldsymbol{z}_k^{-1/2})\mathbf{Q}(\boldsymbol{v}_k^{1/2})\mathbf{Q}(\boldsymbol{z}_k^{-1/2})\right]^{-1}(\widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_*) \rangle \\
= \;& \langle \mathbf{Q}(\boldsymbol{z}_k^{-1/2})\boldsymbol{v}_k^{1/2},\, \left[\mathbf{Q}(\boldsymbol{z}_k^{1/2})\mathbf{Q}(\boldsymbol{v}_k^{-1/2})\mathbf{Q}(\boldsymbol{z}_k^{1/2})\right](\widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_*) \rangle \\
= \;& \langle \mathbf{Q}(\boldsymbol{v}_k^{-1/2})\boldsymbol{v}_k^{1/2},\, \mathbf{Q}(\boldsymbol{z}_k^{1/2})(\widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_*) \rangle \\
= \;& \langle \boldsymbol{v}_k^{-1/2},\, \mathbf{Q}(\boldsymbol{z}_k^{1/2})(\widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_*) \rangle \\
\le \;& \frac{1}{\lambda_{\min}(\boldsymbol{v}_k)^{1/2}} \langle \boldsymbol{z}_k,\, \widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_* \rangle \\
\le \;& \frac{1}{\sqrt{(1-\gamma)\nu_k\mu_0}} \langle \boldsymbol{z}_k,\, \widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_* \rangle.
\end{aligned}
$$

Similarly, from $0 \prec \widetilde{\boldsymbol{z}}_k - \boldsymbol{z}_* \preceq (1+\gamma_d + L\gamma_p)\rho\boldsymbol{e}$, we have

$$
\|\mathbf{Q}(\boldsymbol{p}_k^{-1})(\widetilde{\boldsymbol{z}}_k - \boldsymbol{z}_*)\| \;\le\; \frac{1}{\sqrt{(1-\gamma)\nu_k\mu_0}} \langle \boldsymbol{x}_k,\, \widetilde{\boldsymbol{z}}_k - \boldsymbol{z}_* \rangle.
$$

Therefore, we have

$$
\begin{aligned}
\theta_k^2 T_2^2 \;&\le\; \theta_k^2 \Big(\|\mathbf{Q}(\boldsymbol{p}_k)(\widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_*)\| + \|\mathbf{Q}(\boldsymbol{p}_k^{-1})(\widetilde{\boldsymbol{z}}_k - \boldsymbol{z}_*)\| \Big)^2 \\
&\le\; \frac{\theta_k^2}{(1-\gamma)\nu_k\mu_0} \Big(\langle \boldsymbol{z}_k,\, \widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_* \rangle + \langle \boldsymbol{x}_k,\, \widetilde{\boldsymbol{z}}_k - \boldsymbol{z}_* \rangle \Big)^2.
\end{aligned}
$$

From (57) and the facts that $\langle \boldsymbol{x}_*, \boldsymbol{z}_* \rangle = 0$, $\langle \boldsymbol{x}_k, \boldsymbol{z}_* \rangle$, $\langle \boldsymbol{x}_*, \boldsymbol{z}_k \rangle$, $\langle \widetilde{\boldsymbol{x}}_k, \boldsymbol{z}_* \rangle$, $\langle \widetilde{\boldsymbol{z}}_k, \boldsymbol{x}_* \rangle \ge 0$, we have

$$
\begin{aligned}
& \theta_k \langle \widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_*,\, \boldsymbol{z}_k \rangle + \theta_k \langle \boldsymbol{x}_k,\, \widetilde{\boldsymbol{z}}_k - \boldsymbol{z}_* \rangle \\
= \;& \langle \boldsymbol{z}_k, \boldsymbol{z}_k \rangle - \langle \boldsymbol{x}_k, \boldsymbol{z}_* \rangle - \langle \boldsymbol{x}_*, \boldsymbol{z}_k \rangle + \langle \boldsymbol{x}_*, \boldsymbol{z}_* \rangle \\
& + \theta_k \big(\langle \boldsymbol{x}_*,\, \widetilde{\boldsymbol{z}}_k - \boldsymbol{z}_* \rangle + \langle \widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_*,\, \boldsymbol{z}_* \rangle \big) \\
& + \theta_k^2 \langle \widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_*,\, \widetilde{\boldsymbol{z}}_k - \boldsymbol{z}_* \rangle \\
& - \langle \boldsymbol{x}_k - \boldsymbol{x}_* - \theta_k(\widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_*),\, \mathcal{H}(\boldsymbol{x}_k - \boldsymbol{x}_* - \theta_k(\widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_*)) \rangle \\
\le \;& \langle \boldsymbol{x}_k, \boldsymbol{z}_k \rangle + \theta_k \big(\langle \boldsymbol{x}_*, \widetilde{\boldsymbol{z}}_k \rangle + \langle \widetilde{\boldsymbol{x}}_k, \boldsymbol{z}_* \rangle \big) + \theta_k^2 \langle \widetilde{\boldsymbol{x}}_k, \widetilde{\boldsymbol{z}}_k \rangle \\
\le \;& (1+\gamma)\nu_k\mu_0 n + \theta_k(1+\gamma_d+L\gamma_p)\rho(\langle \boldsymbol{x}_*, \boldsymbol{e} \rangle + \boldsymbol{e}\boldsymbol{z}_*) \\
& + \theta_k^2(1+\gamma_p)(1+\gamma_d+L\gamma_p)\rho^2 n \\
\le \;& 8\nu_k\mu_0 n.
\end{aligned}
$$

Thus $\theta_k^2 T_2^2 = O(n^2\nu_k\mu_0)$. □

Lemma 19.

$$
T_4^2 \;=\; \frac{O(1)}{(1-(\gamma_d+L\gamma_p))^2} n^2 \nu_k \mu_0.
$$

Proof. By Lemma 11, we have

$$
\begin{aligned}
T_4^2 \;&\le\; \frac{1}{(1-\gamma)\nu_k\mu_0} \|\mathbf{Q}(\boldsymbol{x}_k)\|_2 \|\mathcal{H}(\mathcal{A}^+ r_k^p)\|^2 \\
&\le\; \frac{1}{(1-\gamma)\nu_k\mu_0} \|\mathbf{Q}(\boldsymbol{x}_k)\|_2 L^2 \|\mathcal{A}^+ r_k^p\|^2 \\
&\le\; \frac{\gamma_p^2 \rho^2 L^2}{(1-\gamma)\nu_k\mu_0} \theta_k^2 \|\mathbf{Q}(\boldsymbol{x}_k)\|_2 \\
&\le\; \frac{\gamma_p^2 L^2}{(1-\gamma)\nu_k} \frac{O(1)}{(1-(\gamma_d+L\gamma_p))^2} n^2 \nu_k^2 \rho^2 \quad \text{(by Lem. 6)} \\
&=\; \frac{O(1)}{(1-(\gamma_d+L\gamma_p))^2} n^2 \nu_k \mu_0.
\end{aligned}
$$

□

The following proof directly leads to Lemma 10.

Lemma 20.

$$
T_1^2 \;=\; \frac{O(1)}{(1-(\gamma_d+L\gamma_p))^2} n^2 \nu_k \mu_0.
$$

Proof. From Lemma 15 to Lemma 19 and the fact that $(a+b)^2 \le 2a^2 + 2b^2$, we have

$$
\begin{aligned}
T_1^2 \;&\le\; \Big(2\eta_1(\theta_k T_2 + T_3 + T_4) + \sqrt{T_5} \Big)^2 \\
&\le\; 8(\theta_k T_2 + T_3 + T_4)^2 + 2T_5 \\
&\le\; 8(\theta_k T_2 + T_3 + T_4)^2 + 2\|\widehat{S}_k^{-1/2}(R_k^c + r_k^c)\|^2 \\
&\quad + 4\theta_k^2 \langle \widetilde{\boldsymbol{x}}_k - \boldsymbol{x}_*,\, \widetilde{\boldsymbol{z}}_k - \boldsymbol{z}_* \rangle \\
&\quad + 4\theta_k T_2 T_3 + 4T_3^2 + 4\theta_k T_2 T_4 \\
&\le\; \frac{O(1)}{(1-(\gamma_d+L\gamma_p))^2} n^2 \nu_k \mu_0 + O(n\nu_k\mu_0).
\end{aligned}
$$

□

Thus, by Lemma 12 and Lemma 20, we have

$$
\|H_{\boldsymbol{p}_k}(\Delta\boldsymbol{x}_k, \Delta\boldsymbol{z}_k)\| \;\le\; \frac{1}{2} T_1^2 \;=\; \frac{O(1)}{(1-(\gamma_d+L\gamma_p))^2} n^2 \nu_k \mu_0.
$$

References

[1] F. Alizadeh and D. Goldfarb, *Second-order cone programming*, Math. Program., 95 (2003), pp. 3–51.

[2] L. Bergamaschi, J. Gondzio, and G. Zilli, *Preconditioning indefinite systems in interior point methods for optimization*, Comput. Optim. Appl., 28 (2004), pp. 149–71.

[3] S. Cafieri, M. Dapuzzo, V. De Simone, D. di Serafino, and G. Toraldo, *Convergence analysis of an inexact potential reduction method for convex quadratic programming*, J. Optim. Theory Appl., 135 (2007), pp. 355–366.

[4] J.-S. Chai and K.-C. Toh, *Preconditioning and iterative solution of symmetric indefinite linear systems arisingfrom interior point methods for linear programming*, Comput. Optim. Appl., 36 (2007), pp. 221–47.

[5] J. Faraut and A. Korányi, *Analysis on symmetric cones*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 1994. Oxford Science Publications.

[6] L. Faybusovich, *Linear systems in Jordan algebras and primal-dual interior-point algorithms*, J. Comput. Appl. Math, 86 (1997), pp. 149–175.

[7] R. W. Freund, F. Jarre, and S. Mizuno, *Convergence of a class of inexact interior-point algorithms for linear programs*, Math. Oper. Res., 24 (1999), pp. 50–71.

[8] 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.

[9] M. Kojima, M. Shida, and S. Shindoh, *Search directions in the SDP and the monotone SDLCP: generalization and inexact computation*, Math. Program., 85 (1999), pp. 51–80.

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

[11] J. Korzak, *Convergence analysis of inexact infeasible-interior-point algorithms for solving linear programming problems*, SIAM J. Optim., 11 (2000), pp. 133–148.

[12] L. Li and K.-C. Toh, *A polynomial-time inexact primal-dual infeasible path-following algorithm forconvex quadratic SDP*. accepted, Oct 2009.

[13] Z. Lu, R. D. C. Monteiro, and J. W. O'Neal, *An iterative solver-based infeasible primal-dual path-following algorithm for convex quadratic programming*, SIAM J. Optim., 17 (2006), pp. 287–310.

[14] S. Mizuno and F. Jarre, *Global and polynomial-time convergence of an infeasible-interior-point algorithm using inexact computation*, Math. Program., 84 (1999), pp. 105–122.

[15] R. Monteiro, *First- and second-order methods for semidefinite programming*, Math. Program., 97 (2003), pp. 209–244.

[16] R. D. C. Monteiro, *Primal-dual path-following algorithms for semidefinite programming*, SIAM J. Optim., 7 (1997), pp. 663–678.

[17] R. D. C. Monteiro and Y. Zhang, *A unified analysis for a class of long-step primal-dual path-following interior-point algorithms for semidefinite programming*, Math. Program., 81 (1998), pp. 281–299.

[18] M. Muramatsu, *On commutative class of search directions for linear programming oversymmetric cones*, J. Optim. Theory Appl., 112 (2000), pp. 595–625.

[19] A. Nemirovski and M. Todd, *Interior-point methods for optimization*, Acta Numerica, 17.

[20] S. H. Schmieta and F. Alizadeh, *Associative and Jordan algebras, and polynomial time interior-point algorithms for symmetric cones*, Math. Oper. Res., 26 (2001), pp. 543–564.

[21] ——, *Extension of primal-dual interior point algorithms to symmetric cones*, Math. Program., 96 (2003), pp. 409–438.

[22] J. F. Sturm, *Similarity and other spectral relations forsymmetric cones*, Linear Algebra Appl., 312 (2000), pp. 135–154.

[23] M. J. Todd, K.-C. Toh, and R. H. Tütüncü, *On the Nesterov-Todd direction in semidefinite programming*, SIAM J. Optim., 8 (1998), pp. 769–796.

[24] K.-C. Toh, *Solving large scale semidefinite programs via an iterative solver on the augmented systems*, SIAM J. Optim., 14 (2003), pp. 670–698.

[25] ——, *An inexact primal-dual path following algorithm for convex quadratic SDP*, Math. Program., 112 (2008), pp. 221–254.

[26] K.-C. Toh and M. Kojima, *Solving some large scale semidefinite programs via the conjugate residual method*, SIAM J. Optim., 12 (2002), pp. 669–691.

[27] K.-C. Toh, R. H. Tütüncü, and M. J. Todd, *Inexact primal-dual path-following algorithms for a special class of convex quadratic SDP and related problems*, Pac. J. Optim., 3 (2007), pp. 135–164.

[28] T. Tsuchiya, *A convergence analysis of the scaling-invariantprimal-dual path-following algorithms for second-ordercone programming*, Optim. Methods and Software, 11 & 12 (1999), pp. 141–182.

[29] Y. Zhang, *On extending some primal-dual interior-point algorithms from linear programming to semidefinite programming*, SIAM J. Optim., 8 (1998), pp. 365–386.

[30] G. Zhou and K.-C. Toh, *Polynomiality of an inexact infeasible interior point algorithm for semidefinite programming*, Math. Program., 99 (2004), pp. 261–282.

Lu Li
Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076.
E-mail: lilu(at)nus.edu.sg

Kim-Chuan Toh
Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076.
E-mail: mattohkc(at)nus.edu.sg