Polynomial Convergence of Infeasible-Interior-Point Methods over Symmetric Cones.

Bharath Kumar Rangarajan *

Feb 18, 2004

Abstract

We establish polynomial-time convergence of infeasible-interior-point methods for conic programs over symmetric cones using a wide neighborhood of the central path. The convergence is shown for a commutative family of search directions used in Schmieta and Alizadeh [9]. These conic programs include linear and semidefinite programs. This extends the work of Rangarajan and Todd [8], which established convergence of infeasible-interior-point methods for self-scaled conic programs using the NT direction.

1 Introduction

There is an extensive literature on the analysis of interior-point methods (IPMs) for conic programming. In conic programs, a linear function is minimised over the intersection of an affine space and a closed convex cone. The foundation for solving these problems using IPMs was laid by Nesterov and Nemirovskii [6]. These methods were primarily either primal or dual based. Later, Nesterov and Todd [7] introduced symmetric primal-dual interior-point algorithms on a special class of cones called self-scaled cones, which allowed a symmetric treatment of the primal and the dual. Self-scaled cones are precisely the same as symmetric cones, which have been characterised using Jordan algebras (see Guler [3] and also Faraut and Koranyi [1]). Faybusovich [2] analysed an interior-point algorithm over the symmetric cones using this characterisation of symmetric cones.

Nonnegative orthants, second-order cones, and positive semidefinite cones are important special cases of symmetric cones. Monteiro and Zhang [5] gave a unified analysis of feasible-IPMs for semidefinite programs that used the so-called commutative class of search directions. These search directions include the popular directions such as the NT (Nesterov-Todd), the XS and the SX directions. As we shall see, symmetric cones, when described using Jordan algebras, bear a striking resemblance to the cone of real symmetric positive semidefinite matrices. This resemblance was exploited by Schmieta and Alizadeh [9], who extended Monteiro-Zhang's analysis to feasible-IPMs over symmetric cones.

Infeasible-IPMs, unlike feasible-IPMs, do not require that the iterates be feasible to the relevant linear systems, but only be in the interior of the cone constraints. As such infeasible points are easy to obtain, infeasible-IPM are an attractive choice for practical implementations. At the same time, the analysis of infeasible-IPMs

*School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853, USA (bharath@orie.cornell.edu). Research supported in part by NSF through grant DMS-0209457 and ONR through grant N00014-02-1-0057.

presents significant difficulties due to the non-orthogonality of search directions. Zhang [10] analysed the convergence of an infeasible-interior-point algorithm for semidefinite programming using the XS and SX search directions. Rangarajan and Todd [8] established convergence of an infeasible-IPM for self-scaled cones using the Nesterov-Todd direction for a wide neighborhood of the central path.

In this paper, we show the convergence of an infeasible-IPM on symmetric cones for the commutative class of search directions. In the process a Lyapunov lemma in this setting is established. To our knowledge this is the first time an infeasible-interior-point method has been analysed for the NT-method using the $\mathcal{N}_{-\infty}$ neighborhood for both semidefinite programming and conic programs over symmetric cones. The complexity result obtained here for symmetric cones using the NT direction compares with the best bound obtained for linear programs. Besides the work of Schmieta and Alizadeh, our main tool is the analysis of an NT-based infeasible-IPM for self-scaled conic programming in Rangarajan and Todd [8].

This paper is organized as follows: We start with an introduction to the theory of Jordan algebras. Next we outline the basics of interior-point theory that leads to the algorithm and present its analysis. We present some conclusions in the final section.

2 Euclidean Jordan Algebras

Characterization of symmetric cones using Jordan algebras (see Theorem 2.3) forms the fundamental basis for our analysis. This section covers the basic results in Jordan algebras, closely following Schmieta and Alizadeh [9] in presentation. For a comprehensive treatment of Jordan algebras, the reader is referred to Faraut and Koranyi [1]. For the purposes of illustration, we use the space of real symmetric matrices, which yields the cone of positive semidefinite matrices. In this case, the analysis in Section 3 specialises to the case of semidefinite programming.

Definition 2.1 *Let $\mathcal{J}$ be an n-dimensional vector space over the field of real numbers along with the bilinear map $\bullet : (x, y) \mapsto x \bullet y \in \mathcal{J}$. Then $(\mathcal{J}, \bullet)$ is a Euclidean Jordan algebra with identity if for all $x, y \in \mathcal{J}$*

1. $x \bullet y = y \bullet x$ (Commutativity).
2. $x \bullet (y \bullet x^2) = (x \bullet y) \bullet x^2$ *where* $x^2 = x \bullet x$ (Jordan Identity).
3. *There exists a symmetric positive definite quadratic form $\mathcal{Q}$ on $\mathcal{J}$ such that $\mathcal{Q}(x \bullet y, z) = \mathcal{Q}(x, y \bullet z)$.*
4. *There exists an identity element $e \in \mathcal{J}$, i.e., e such that $e \bullet x = x \bullet e$ for all $x \in \mathcal{J}$.*

Definition 2.2 *If $\mathcal{J}$ is a Euclidean Jordan algebra, then its cone of squares is the set*

$$\mathcal{K}(\mathcal{J}) := \{x^2 : x \in \mathcal{J}\}.$$

Symmetric cones are cones that are self-dual and homogeneous: their automorphism groups act transitively on their interiors. Symmetric cones are also precisely the class of self-scaled cones introduced by Nesterov and Todd in [7] (see also Faybusovich [2] and Guler [3]). The following theorem relates symmetric cones and Euclidean Jordan algebras.

Theorem 2.3 (Jordan algebraic characterization of symmetric cones).
A cone is symmetric iff it is the cone of squares of some Euclidean Jordan algebra.

Example Let $\mathcal{J} = \mathcal{S}^n$, the space of real symmetric matrices with the operation $X \bullet Y := \frac{XY+YX}{2}$ for $X, Y \in \mathcal{S}^n$. We can choose $\mathcal{Q}(X,Y) := \text{Trace}\,(XY)$ and e to be the identity matrix. Then $(\mathcal{J}, \bullet)$ is a Euclidean Jordan algebra with identity. We obtain the cone of symmetric positive semidefinite matrices as the squares of real symmetric matrices.

Since $\bullet$ is a bilinear map, for every $x \in \mathcal{J}$ a linear operator $L(x)$ can be defined such that $L(x)y = x \bullet y$ for all $y \in \mathcal{J}$. For $x, y \in \mathcal{J}$, let

$$Q_{x,y} := L(x)L(y) + L(y)L(x) - L(x \bullet y) \text{ and } Q_x := Q_{x,x} = 2L^2(x) - L(x^2),$$

where Q_x is called the quadratic representation of x. Clearly $Q_{x,y}z$ and $Q_x z$ are in $\mathcal{J}$ for all $x, y, z \in \mathcal{J}$.

Example For $X \in \mathcal{S}^n$ $L(X)$ is the operator from $\mathcal{S}^n$ to itself such that $L(X)[Y] = \frac{XY+YX}{2}$. A further computation shows that $Q_{X,Y}[Z] = \frac{XZY+YZX}{2}$ and $Q_X[Z] = XZX$. Q_X plays an important role in the analysis of interior-point methods for semidefinite programming. The operator Q_x in Jordan algebras plays a similar role in our analysis.

An element $x \in \mathcal{J}$ is called *invertible* if there exists a $y = \sum_{i=0}^{k} \gamma_i x^i$ for some finite $k < \infty$ and real numbers γ_i such that $y \bullet x = e$, and is written x^{-1}. The following are some of the basic properties of Q_x (see Propositions II.3.1 and II.3.3 in [1]).

Lemma 2.4 *Let $x, y \in \mathcal{J}$. Then*

1. $Q_x x^{-1} = x$ *(or equivalently $Q_x L(x^{-1}) = L(x)$), $Q_x^{-1} = Q_{x^{-1}}$ and $Q_x e = x^2$.*
2. $Q_{Q_y x} = Q_y Q_x Q_y$.

Using the Jordan identity, the notions of rank, the minimum and the characteristic polynomial, the trace and the determinant can be defined in the following way.

Definition 2.5 *a. For $x \in \mathcal{J}$, let r be the smallest integer such that the set $\{e, x, x^2, \ldots, x^r\}$ is linearly dependent. Then r is called the degree of x and is denoted by* $\deg(x)$.

b. The rank of $\mathcal{J}$, denoted by $\text{rank}\,(\mathcal{J})$*, is defined as the maximum of* $\deg(x)$ *over all $x \in \mathcal{J}$. An element is called regular if its degree equals the rank of the Jordan algebra.*

For an element x of degree d, there exist real numbers $a_1(x), \ldots, a_d(x)$ such that

$$x^d - a_1(x)x^{d-1} + \ldots + (-1)^d a_d(x)e = 0, \text{ where } 0 \text{ is the zero vector.}$$

Then the polynomial $\lambda^d - a_1(x)\lambda^{d-1} + \ldots + (-1)^d a_d(x) = 0$ is called the *minimum polynomial* of x. The *characteristic polynomial* is defined to be the minimum polynomial for a regular element. Using the fact that the regular elements are dense in $\mathcal{J}$, the characteristic polynomial can be continuously extended to all of $\mathcal{J}$ (see [1]). Therefore the characteristic polynomial is a degree r polynomial in λ, where r is the rank of $\mathcal{J}$.

The roots $\lambda_1, \ldots, \lambda_r$ of the characteristic polynomial of x are called the *eigenvalues* of x. The roots of the minimum and the characteristic polynomial are the same except for their multiplicity and the minimum polynomial always divides the characteristic polynomial.

Definition 2.6 *Let $x \in \mathcal{J}$ and $\lambda_1, \ldots, \lambda_r$ be its eigenvalues. Then,*

1. $\text{Trace}\,(x) := \lambda_1 + \ldots + \lambda_r$ *is called the trace of x;*
2. $\text{Det}\,(x) := \lambda_1 \cdots \lambda_r$ *is called the determinant of x.*

Trace can be shown to be a linear function of x. For the identity element, $\text{Trace}\,(e) = r$ and $\text{Det}\,(e) = 1$ as all its eigenvalues are unity.

Example The above definitions correspond to the usual notions of characteristic polynomials, eigenvalues, trace and determinant of matrices. For matrices, $\deg(X)$ corresponds to the degree of the minimum polynomial of X, which is the same as the number of distinct eigenvalues of X.

Next, the concept of Jordan frames is introduced and a spectral decomposition result is presented. An *idempotent* c is a nonzero element of $\mathcal{J}$ such that $c^2 = c$. A complete system of orthogonal idempotents is a set $\{c_1, \ldots, c_k\}$ of idempotents, where $c_i \bullet c_j = 0$ for all $i \neq j$, and $c_1 + \ldots + c_k = e$. An idempotent is *primitive* if it is not the sum of two other idempotents. A complete system of orthogonal primitive idempotents is called a *Jordan frame*. Note that in Jordan frames $k = r$, that is Jordan frames always contain r primitive idempotents.

Theorem 2.7 (Spectral decomposition, Theorem III.1.2, [1]). *Let $\mathcal{J}$ be a Euclidean Jordan algebra. For $x \in \mathcal{J}$ there exist a Jordan frame $c_1, \ldots, c_r$ and real numbers $\lambda_1, \ldots, \lambda_r$ such that $x = \lambda_1 c_1 + \cdots + \lambda_r c_r$, where the λ_i's are called the eigenvalues of x.*

Using this we can define the following: (analogous to functions on the real line)

1. The square root: $x^{1/2} := \lambda_1^{1/2} c_1 + \cdots + \lambda_r^{1/2} c_r$ whenever all $\lambda_i \geq 0$, and undefined otherwise.
2. The inverse: $x^{-1} := \lambda_1^{-1} c_1 + \cdots + \lambda_r^{-1} c_r$ whenever all $\lambda_i \neq 0$, and undefined otherwise. (This is consistent with our earlier definition by Proposition II.2.4 in [1].)
3. The square: $x^2 := \lambda_1^2 c_1 + \cdots + \lambda_r^2 c_r$.

These definitions can be shown to be well-defined. Note that x^2 can be viewed as either $x \bullet x$ or as the extension of the "square" function on the reals. Also note that $(x^{1/2})^2 = x$. It can be shown that an element is in (the interior of) the cone of squares iff all its eigenvalues are non-negative (positive).

Next, norms and inner products are defined on $\mathcal{J}$. Since $\text{Trace}\,(x \bullet y)$ is a bilinear function, the inner product can be defined as $\langle x, y \rangle := \text{Trace}\,(x \bullet y)$. For $x \in \mathcal{J}$, with eigenvalues λ_i, $1 \leq i \leq r$, the Frobenius norm and the spectral norm (or the 2-norm) can be defined as (see Proposition III.1.5 in [1])

$$\|x\|_F := \sqrt{\sum_{i=1}^{r} \lambda_i^2} = \sqrt{\text{Trace}\,(x^2)} \quad \text{and} \quad \|x\|_2 := \max_i |\lambda_i|.$$

Then the Cauchy-Schwarz inequality holds;

$$|\langle x, y \rangle| \leq \|x\|_F \|y\|_F.$$

As all the eigenvalues of e are unity, $\|e\|_F = \sqrt{r}$ and $\|e\|_2 = 1$.

Example For a matrix $X \in \mathcal{S}^n$, we have the spectral decomposition that there exists a set of orthonormal vectors $\{q_i, 1 \leq i \leq n\} \subset \Re^n$ and real numbers $\lambda_1, \ldots, \lambda_n$ such that $X = \sum_i \lambda_i q_i q_i^T$. It can be checked that the matrices $q_i q_i^T$ form a primitive system of orthogonal idempotents. The inner product is the usual trace inner product of matrices and the spectral and Frobenius norms have their usual definitions.

Since $\text{Trace}\,(\cdot,\cdot)$ is associative (see Proposition II.4.1 in [1]), i.e., $\text{Trace}\,(x \bullet (y \bullet z)) = \text{Trace}\,((x \bullet y) \bullet z)$,

$$\langle L(x)p, q \rangle = \text{Trace}\,((x \bullet p) \bullet q) = \text{Trace}\,((p \bullet x) \bullet q) = \text{Trace}\,(p \bullet (x \bullet q)) = \langle p, L(x)q \rangle$$

shows that $L(x)$ is a self-adjoint operator. As the definition of Q_x depends only on $L(x)$ and $L(x^2)$, both of which are self-adjoint, Q_x is also self-adjoint.

We recall parts of Lemma 12, 13, and 14 in [9] in the next two lemmas.

Lemma 2.8 *Let $x = \lambda_1 c_1 + \cdots + \lambda_r c_r$, using the spectral decomposition. Then the following statements hold.*

1. *The matrices $L(x)$ and Q_x commute and thus share a common system of eigenvectors.*
2. *The eigenvalues of $L(x)$ have the form $\frac{\lambda_i+\lambda_j}{2},\ 1 \le i \le j \le r$. In particular, $x \in \mathcal{K}$ (int $\mathcal{K}$) iff $L(x)$ is positive semidefinite (definite). However, not every $\frac{\lambda_i+\lambda_j}{2}$ is an eigenvalue of $L(x)$.*
3. *The eigenvalues of Q_x have the form $\lambda_i\lambda_j,\ 1 \le i \le j \le r$. However, not every $\lambda_i\lambda_j$ is an eigenvalue of Q_x.*

Henceforth the minimum (maximum) eigenvalue of x will be denoted by $\lambda_{\min}(x)$ ($\lambda_{\max}(x)$).

Lemma 2.9 *Let $x \in \mathcal{J}$, then we have*

$$\lambda_{min}(x) = \min_u \frac{\langle u, u \bullet x\rangle}{\langle u, u\rangle}.$$

For $x, y \in \mathcal{J}$, we have

$$\begin{aligned} \lambda_{min}(x+y) &\ \ge\ \lambda_{min}(x) - \|y\|_F \\ \|x \bullet y\|_F &\ \le\ \|x\|_F\|y\|_F. \end{aligned}$$

Proof : For proofs of all but the last part, see Lemma 13 and Lemma 14 in [9]. The last part follows from

$$\|x \bullet y\|_F = \|L(x)y\|_F \le \|L(x)\|\|y\|_F = \|x\|_2\|y\|_F \le \|x\|_F\|y\|_F.$$

The first equality follows from the definition of $L(x)$, and $\|L(x)\|$ refers to the operator norm induced by $\|\cdot\|_F$. For the second equality note that the spectral norm of a self-adjoint linear operator is $\|L(x)\| = \max_i |\lambda_i(L(x))|$. By Lemma 2.8 $\max_i |\lambda_i(L(x))| = \max_i |\lambda_i(x)| = \|x\|_2$. Lastly, note that 2-norm is bounded by the Frobenius norm. □

We state two useful propositions about the operator Q_x.

Proposition 2.10 (Proposition III.2.2, Faraut and Koranyi [1].) *If $x, y \in$ int $\mathcal{K}$, then $Q_x y \in$ int $\mathcal{K}$.*

By noting that $x^{-1} \in \mathcal{K}$ and $Q_{x^{-1}} = Q_x^{-1}$ (from Lemma 2.4) it follows that Q_x is also onto and hence an automorphism of $\mathcal{K}$.

Proposition 2.11 *Let $x, y \in$ int $\mathcal{K}$, then*

1. *$Q_{x^{1/2}}s$ and $Q_{s^{1/2}}x$ have the same spectrum.*
2. *If $p \in$ int $\mathcal{K}$ define $\tilde{x} := Q_p x$ and $\tilde{s} := Q_{p^{-1}}s$, then $Q_{x^{1/2}}s$ and $Q_{\tilde{x}^{1/2}}\tilde{s}$ have the same spectrum.*

Furthermore, $\operatorname{Trace}(Q_{x^{1/2}}s) = \langle s, x\rangle$.

Proof : See Proposition 21 in [9] for proofs of 1 and 2. To complete the proof of the proposition, note that if $\{\lambda_i\}$ are the eigenvalues of $Q_{x^{1/2}}s$, then using the self-adjointness of $Q_{x^{1/2}}$ we have

$$\operatorname{Trace}(Q_{x^{1/2}}s) = \operatorname{Trace}((Q_{x^{1/2}}s) \bullet e) = \langle Q_{x^{1/2}}s, e\rangle = \langle s, Q_{x^{1/2}}e\rangle = \langle s, x\rangle.$$

□

Now we are ready to state and prove the Lyapunov Lemma for Euclidean Jordan algebras.

Lemma 2.12 (Lyapunov Lemma for Euclidean Jordan Algebras) *Suppose that $\mathcal{J}$ is a Euclidean Jordan algebra. If $x \in$ int $\mathcal{K}$, $w \in \mathcal{K}$ then there exists $s \in \mathcal{K}$ such that $x \bullet s = w$.*

Proof : Let us set $s = 2 \int_0^\infty Q_{v(t)} w \, dt$, where $x = \sum_{i=1}^r \lambda_i c_i$, is the spectral decomposition of x and $v(t) = \sum_{i=1}^r e^{-\lambda_i t} c_i$. Clearly $v(t) \in \mathcal{J}$ as $c_i \in \mathcal{J}$. By expanding using the spectral decomposition and integrating we obtain $s = 2 \sum_{i,j} \frac{1}{\lambda_i + \lambda_j} Q_{c_i, c_j} w$ and hence, s is well-defined and $s \in \mathcal{J}$. Observe that $v(t) \in$ int $\mathcal{K}$ as $e^{-\lambda_i t} > 0$ for all t and hence $Q_{v(t)}$ is an automorphism of $\mathcal{K}$. It follows that $Q_{v(t)} w \in \mathcal{K}$. For $u \in \mathcal{K}$, we have

$$\langle s, u \rangle = 2 \left\langle \int_0^\infty Q_{v(t)} w \, dt, u \right\rangle = 2 \int_0^\infty \left\langle Q_{v(t)} w, u \right\rangle \, dt \geq 0.$$

Consequently $s \in \mathcal{K}$. By Proposition II.3.4 in [1] $Q_{v(t)} = e^{-2tL(x)}$. Therefore,

$$\frac{d}{dt} Q_{v(t)} w = \frac{d}{dt} e^{-2tL(x)} w = -2L(x) e^{-2tL(x)} w = -2L(x) Q_{v(t)} w = -2x \bullet Q_{v(t)} w.$$

We can substitute for s in the equation and see that

$$x \bullet s = 2 \int_0^\infty x \bullet Q_{v(t)} w \, dt = \int_0^\infty -\frac{d}{dt} \left(Q_{v(t)} w \right) \, dt = w.$$

□

The operator commutativity for a Jordan algebra is defined and an important related result is stated. The notion of operator commutativity is not to be confused with the commutativity of elements of the Jordan algebra.

Definition 2.13 *We say two elements x, y of a Jordan algebra $\mathcal{J}$ operator commute if $L(x)L(y) = L(y)L(x)$. In other words, x and y operator commute if for all z, $x \bullet (y \bullet z) = y \bullet (x \bullet z)$.*

Theorem 2.14 (Theorem 27, [9]) *Let x and y be two elements of Euclidean Jordan algebra $\mathcal{J}$. Then x and y operator commute if and only if there is a Jordan frame $c_1, \ldots, c_r$ such that $x = \sum_{i=1}^r \lambda_i c_i$ and $s = \sum_{i=1}^r \mu_i c_i$ for some λ_i, μ_i.*

A Jordan algebra is called *simple* if it cannot be represented as the sum of two Jordan algebras. Simple Jordan algebras have been classified into the following five cases and consequently we have a classification for symmetric cones (see Chapter V in [1]). This classification is due to Jordan, Von Neumann and Wigner [4].

Theorem 2.15 (Chapter V, Faraut and Koranyi [1].) *Let $\mathcal{J}$ be a simple Euclidean Jordan algebra. Then $\mathcal{J}$ is isomorphic to one of the following algebras, where for the matrix algebras, the operation is defined by $X \bullet Y = \frac{1}{2}(XY + YX)$:*

1. *the algebra $\mathcal{E}_{n+1}$, the algebra of quadratic forms in $\Re^{n+1}$ under the operation $x \bullet y = (x^T y; x_0 \bar{y} + y_0 \bar{x})$, where $x = (x_0; \bar{x}), y = (y_0; \bar{y}) \in \Re \times \Re^n$.*

2. *the algebra $(\mathcal{S}^n, \bullet)$ of $n \times n$ symmetric matrices.*

3. *the algebra $(\mathcal{H}_n, \bullet)$ of $n \times n$ complex Hermitian matrices.*

4. *the algebra $(\mathcal{Q}_n, \bullet)$ of $n \times n$ quaternion Hermitian matrices.*

5. *the exceptional Albert algebra, i.e., the algebra $(\mathcal{O}_3, \bullet)$ of 3×3 octonian Hermitian matrices.*

3 Algorithm and Analysis

3.1 Problem background

We begin with the problem statement and discuss some of the theory relevant to developing interior-point algorithms: the perturbed optimality conditions, central path and the Newton systems that give rise to the commutative class of search directions. In the following subsection, we present the algorithm and analyze its convergence.

Let $\mathcal{J}$ be a Euclidean Jordan algebra of dimension n and rank r, and $\mathcal{K}$ be its cone of squares. Consider the following primal and the associated dual problem.

Primal and Dual

$$(P) \qquad \min\{\langle c, x\rangle : \ Ax = b, \ x \in \mathcal{K}\} \tag{3.1}$$

and

$$(D) \qquad \max\{\langle b, y\rangle_Y : \ A^*y + s = c, \ s \in \mathcal{K}, \ y \in Y\}, \tag{3.2}$$

where $c \in \mathcal{J}$ and $b \in Y$, a finite dimensional real vector space with an inner product $\langle \cdot, \cdot\rangle_Y$. Here A is a linear operator that maps $\mathcal{J}$ into Y. A^* is defined to be the linear operator that maps Y to $\mathcal{J}$ such that $\langle A^*y, x\rangle = \langle Ax, y\rangle_Y$ for all $x \in \mathcal{J}, y \in Y$.

We call $x \in \mathcal{K}$ primal feasible if $Ax = b$. Similarly $(s, y) \in \mathcal{K} \times Y$ is called dual feasible if $A^*y + s = c$. Let

$$\begin{aligned} \mathcal{F}^0(P) &:= \{x \in \mathcal{J} : \ Ax = b, \ x \in \text{int}\,\mathcal{K}\} \text{ and} \\ \mathcal{F}^0(D) &:= \{(s, y) \in \mathcal{J} \times Y : \ A^*y + s = c, \ s \in \text{int}\,\mathcal{K}\} \end{aligned}$$

represent the interior feasible solutions of the primal and the dual. For the rest of the paper, we will assume that A is surjective, $\mathcal{F}^0(P) \neq \emptyset$, and $\mathcal{F}^0(D) \neq \emptyset$. For a given primal feasible x and dual feasible (s, y), $\langle s, x\rangle$ is called the duality gap due to the well-known relation

$$\langle b, y\rangle_Y - \langle c, x\rangle = \langle Ax, y\rangle_Y - \langle A^*y + s, x\rangle = \langle s, x\rangle \geq 0.$$

Since the iterates in our algorithm may not satisfy the linear constraints, $\langle s, x\rangle$ will be referred to as the *complementarity*. Let us note that $\langle s, x\rangle = 0$ for feasible (x, s, y) is sufficient for optimality. By Lemma 2.2 in [2], for $x, s \in \mathcal{K}$ $\langle s, x\rangle = 0$ is equivalent to $s \bullet x = 0$. Using our assumptions above, the optimality conditions for the primal and dual problems can be written as

$$\begin{aligned} Ax &= b \\ A^*y + s &= c \\ s \bullet x &= 0 \\ x, s \in \mathcal{K}, y \in Y, & \end{aligned} \tag{3.3}$$

where $s \bullet x = 0$ is usually referred to as the complementary slackness condition.

The perturbed optimality conditions (PC_μ) are obtained by replacing $s \bullet x = 0$ in (3.3) with the "perturbed" complementary slackness condition, $s \bullet x = \mu e$ for $\mu > 0$. Interior-point algorithms follow the solutions to (PC_μ) as μ goes to zero. The perturbed optimality conditions have unique solutions for all positive μ, and these solutions form the so-called central trajectory (see [2]). Note that the duality gap of the solutions is proportional to μ, i.e., $\langle s, x\rangle = \text{Trace}\,(s \bullet x) = \mu \text{Trace}\,(e) = \mu r$. IPMs employ Newton's method to target the solution of $(PC_{\sigma\mu})$, where $\sigma \in (0,1)$, (x,s,y) is the current iterate and $\mu = \frac{\langle s,x\rangle}{r}$. Such algorithms are called primal-dual path-following algorithms; primal-dual, because the primal and the dual are treated symmetrically in the optimality conditions.

The following lemma motivates different, but equivalent, ways of forming the perturbed optimality conditions, thus leading to different Newton systems.

Lemma 3.1 (Lemma 28 in [9]) *Let x, s and p be in some Euclidean Jordan algebra $\mathcal{J}$, $x, s \in$ int $\mathcal{K}$ and p invertible. Then $s \bullet x = \mu\, e$ iff $Q_{p^{-1}}(s) \bullet Q_p(x) = \mu\, e$.*

Therefore for a scaling $p \in \text{int}\,\mathcal{K}$, (PC_μ) can be equivalently written as

$$\begin{aligned} \tilde{A}\tilde{x} &= b \\ \tilde{A}^* y + \tilde{s} &= \tilde{c} \\ \tilde{s} \bullet \tilde{x} &= \mu e \\ \tilde{x}, \tilde{s} \in \mathcal{K}, y &\in Y, \end{aligned}$$

where $\tilde{x} = Q_p x$, $\tilde{s} = Q_{p^{-1}} s$, $\tilde{A} = A Q_{p^{-1}}$, and $\tilde{c} = Q_{p^{-1}} c$. We restrict our attention to the following set of scalings

$$\mathcal{C}(x,s) := \{p: \ p \in \text{int}\,\mathcal{K} \text{ such that } Q_p(x) \text{ and } Q_{p^{-1}}(s) \text{ operator commute}\}.$$

Note that $p = e$ need not be in $\mathcal{C}(x,s)$. For $p = x^{-1/2}$ we get the xs-method, for $p = s^{1/2}$ we get the sx-method and for the choice of $p = \left[Q_{x^{1/2}}(Q_{x^{1/2}} s)^{-1/2}\right]^{-1/2} = \left[Q_{s^{-1/2}}(Q_{s^{1/2}} x)^{1/2}\right]^{-1/2}$, we get the Nesterov-Todd (NT) method. The Newton equations corresponding to a scaling in $\mathcal{C}(x,s)$ are

Scaled Newton Equations

$$\begin{array}{rcccl} & \tilde{A}^* \triangle y & + \ \triangle \tilde{s} & = & \tilde{c} - \tilde{A}^* y - \tilde{s}, \\ \tilde{A} \triangle \tilde{x} & & & = & b - \tilde{A}\tilde{x}, \\ \tilde{s} \bullet \triangle \tilde{x} & & + \ \triangle \tilde{s} \bullet \tilde{x} & = & \sigma\mu e - \tilde{s} \bullet \tilde{x}. \end{array} \tag{3.4}$$

Though $\mathcal{C}(x,s)$ seems to be a restrictive class, it does include some of the most interesting choices of scalings.

Our algorithm will restrict the iterates to the following neighborhood, called the minus-infinity neighborhood, of the central path. For a given constant $\gamma \in [0,1]$

$$\mathcal{N}_{-\infty}(\gamma) := \{(x,s,y) \in \mathcal{K} \times \mathcal{K} \times Y: \ d_{-\infty}(x,s) \le \gamma\mu\}, \tag{3.5}$$

where

$$d_{-\infty}(x,s) := \mu - \lambda_{min}(z), \ \mu = \frac{\langle s,x\rangle}{r} \text{ and } z = Q_{x^{1/2}} s.$$

A few observations about z are in order. As $x^{1/2} \in \mathcal{K}$ and $Q_{x^{1/2}}$ is an automorphism of $\mathcal{K}$, $z \in \mathcal{K}$ and hence $\lambda_i(z)$ are nonnegative. By Proposition 2.11 $\langle s,x\rangle = \text{Trace}\,(z) = \sum_i \lambda_i(z)$. The neighborhood contains the central path and γ represents the size of the neighborhood as it can be shown that the set $\mathcal{N}_{-\infty}(0) \cap \left[\mathcal{F}^0(P) \times \mathcal{F}^0(D)\right]$ is exactly the central path and $\mathcal{N}_{-\infty}(1) \cap \left[\mathcal{F}^0(P) \times \mathcal{F}^0(D)\right] = \mathcal{F}^0(P) \times \mathcal{F}^0(D)$.

Now we discuss the symmetry and scale-invariance of the neighborhoods. By part (i) of Proposition 2.11, $Q_{x^{1/2}} s$ and $Q_{s^{1/2}} x$ have the same spectrum. Hence the centrality measure $d_{-\infty}(x,s)$ and the neighborhood $\mathcal{N}_{-\infty}$ are symmetric with respect to x and s.

Proposition 3.2 *The neighborhood is scaling invariant, that is (x,s) is in the neighborhood iff $(\tilde{x}, \tilde{s})$ is.*

Proof : Let $\tilde{z} := Q_{\tilde{x}^{1/2}}\tilde{s}$. By part (ii) of Proposition 2.11 $\lambda(\tilde{z})$ is the same as $\lambda(z)$. Since $\langle \tilde{s}, \tilde{x} \rangle = \langle Q_{p^{-1}}s, Q_p x \rangle = \langle s, x \rangle$, the result follows by substituting the expressions in the definition of $\mathcal{N}_{-\infty}(\gamma)$. $\square$

Hence the scaling transformations are not just automorphisms of the cone but they also map the neighborhood to itself. As the definition of $\mathcal{N}_{-\infty}$ is independent of y, sometimes y in (x, s, y) is suppressed for convenience and we write (x, s) instead, but y should be clear from the context.

3.2 Algorithm and Analysis of Convergence

Having discussed the key elements needed for the algorithm, we describe the infeasible-interior-point-method in detail.

Algorithm IIPM :

1 Let $1 > \beta > \sigma > 0$, $\epsilon^* > 0$, $\gamma \in (0, 1)$, $x_0 \in \text{int}\,\mathcal{K}$, $y_0 \in Y$ and $s_0 \in \text{int}\,\mathcal{K}$ be given such that $(x_0, s_0, y_0) \in \mathcal{N}_{-\infty}(\gamma)$. Set $k = 0$, $\phi_p^0 = 1$ and $\phi_d^0 = 1$.

2 Choose a $p \in \mathcal{C}(x_k, s_k)$ and form the corresponding scaled iterate. Solve for $(\triangle\tilde{x}_k, \triangle\tilde{s}_k, \triangle y_k)$ from the scaled Newton equations in (3.4) at $(\tilde{x}_k, \tilde{s}_k, y_k)$. Let $(\triangle x_k, \triangle s_k, \triangle y_k)$ = $(Q_{p^{-1}}\triangle\tilde{x}_k, Q_p\triangle\tilde{s}_k, \triangle y_k)$.

3 Let $(x(\alpha), s(\alpha), y(\alpha)) := (x_k, s_k, y_k) + \alpha(\triangle x_k, \triangle s_k, \triangle y_k)$. Compute the largest step length $\bar{\alpha}_k \in (0, 1]$ such that for all $\alpha \in [0, \bar{\alpha}_k]$, $(x(\alpha), s(\alpha), y(\alpha)) \in \mathcal{N}_{-\infty}(\gamma)$, $\langle s(\alpha), x(\alpha) \rangle \geq \max(\phi_p^k, \phi_d^k)(1 - \alpha) \langle s_0, x_0 \rangle$, and $\langle s(\alpha), x(\alpha) \rangle \leq (1 - (1 - \beta)\alpha) \langle s_k, x_k \rangle$.

4 Choose a primal step length $\alpha_p^k > 0$ and a dual step length $\alpha_d^k > 0$ such that

$$\begin{aligned}
(x_{k+1}, s_{k+1}, y_{k+1}) &:= (x_k + \alpha_p^k \triangle x_k, s_k + \alpha_d^k \triangle s_k, y_k + \alpha_d^k \triangle y_k) \in \mathcal{N}_{-\infty}(\gamma), \\
\langle s_{k+1}, x_{k+1} \rangle &\geq \max(\phi_p^k(1 - \alpha_p^k), \phi_d^k(1 - \alpha_d^k)) \langle s_0, x_0 \rangle \text{ and} \\
\langle s_{k+1}, x_{k+1} \rangle &\leq (1 - (1 - \beta)\bar{\alpha}_k) \langle s_k, x_k \rangle .
\end{aligned}$$

Set $\phi_p^{k+1} = \phi_p^k(1 - \alpha_p^k)$ and $\phi_d^{k+1} = \phi_d^k(1 - \alpha_d^k)$.

5 Increase k by 1. If $\langle s_k, x_k \rangle < \epsilon^* \langle s_0, x_0 \rangle$, then STOP. Otherwise, repeat step 2.

On the choice of step lengths: if we choose $\alpha_p^k = \alpha_d^k = \bar{\alpha}_k$, all the conditions in Step 4 are satisfied. However, we are free to choose different step lengths as long we can make a comparable progress in the feasibility and complementarity while remaining inside the neighborhood.

Using the Newton equations we can show that ϕ_p^k and ϕ_d^k satisfy the relations

$$Ax_k - b = \phi_p^k(Ax_0 - b) \text{ and } \quad A^*y_k + s_k - c = \phi_d^k(A^*y_0 + s_0 - c), \tag{3.6}$$

and hence they represent the relative infeasibilities at (x_k, s_k, y_k). At every iterate we maintain the feasibility condition,

$$\langle s_k, x_k \rangle \geq \max(\phi_p^k, \phi_d^k) \langle s_0, x_0 \rangle , \tag{3.7}$$

which ensures that the infeasibilities approach zero as the complementarity, $\langle s, x \rangle$, approaches zero. The following theorem forms the skeleton of the convergence argument and sets the agenda for the rest of the paper.

Theorem 3.3 *If $\bar{\alpha}_k \geq \alpha^*$ for all k for some $\alpha^* > 0$, then the* **IIPM** *will terminate with (x_k, s_k, y_k) such that $\|Ax_k - b\| \leq \epsilon^* \|Ax_0 - b\|$, $\|A^*y_k + s_k - c\| \leq \epsilon^* \|A^*y_0 + s_0 - c\|$ and $\langle s_k, x_k \rangle \leq \epsilon^* \langle s_0, x_0 \rangle$ in $\mathcal{O}(\frac{1}{\alpha^*} \ln \left(\frac{1}{\epsilon^*}\right))$ iterations.*

Proof : All the conditions in Step 3 of **IIPM** are satisfied for α^*. Since for each k, $\bar{\alpha}_k \geq \alpha^*$, if we choose $k = \left\lceil \frac{1}{(1-\beta)\alpha^*} \right\rceil \ln\left(\frac{1}{\epsilon^*}\right)$, then we have

$$\begin{aligned}
\ln(\langle s_k, x_k\rangle) &\leq \ln\left(\langle s_{k-1}, x_{k-1}\rangle (1-\alpha^*(1-\beta))\right) \\
&\leq \ln\left(\langle s_0, x_0\rangle (1-\alpha^*(1-\beta))^k\right) \\
&\leq \ln(\langle s_0, x_0\rangle) - k\alpha^*(1-\beta) \\
&\leq \ln(\langle s_0, x_0\rangle) + \ln(\epsilon^*) = \ln(\epsilon^* \langle s_0, x_0\rangle).
\end{aligned}$$

The first inequality follows from the decrease in complementarity condition, the second from the same applied inductively, and the third inequality from the identity $1+\xi \leq e^{\xi}$ for all $\xi > -1$. The fourth inequality follows from our assumption on k.

From condition (3.7), it follows that $\max(\phi_p^k, \phi_d^k) \leq \frac{\langle s_k, x_k\rangle}{\langle s_0, x_0\rangle} \leq \epsilon^*$. Then (3.6) implies that

$$\|Ax_k - b\| \leq \epsilon^* \|Ax_0 - b\|, \text{ and } \|A^*y_k + s_k - c\| \leq \epsilon^* \|A^*y_0 + s_0 - c\|.$$

□

In the rest of the paper, we prove that such a lower bound on α^* exists and establish an estimate of the lower bound that leads to the polynomial convergence result for the **IIPM**. For simplicity, we will often write x, y, s and $\bar{\phi}$ for x_k, y_k, s_k and $\max(\phi_p^k, \phi_d^k)$ respectively. The indices should be clear from the context.

Let $(x, s, y) \in \mathcal{N}_{-\infty}(\gamma)$ and satisfy the feasibility condition (3.7). For a fixed $p \in \mathcal{C}(x, s)$, let $(\triangle\tilde{x}, \triangle\tilde{s}, \triangle y)$ be the direction computed in Step 2 of the algorithm. We will use the following notation:

$$\begin{aligned}
&\tilde{x}(\alpha) = \tilde{x} + \alpha\triangle\tilde{x}, \quad \tilde{s}(\alpha) = \tilde{s} + \alpha\triangle\tilde{s}, \\
&x(\alpha) = x + \alpha\triangle x, \quad s(\alpha) = s + \alpha\triangle s, \\
&\tilde{\mu}(\alpha) = \mu(\tilde{x}(\alpha), \tilde{s}(\alpha)) = \frac{\langle \tilde{s}(\alpha), \tilde{x}(\alpha)\rangle}{r}, \text{ and } \tilde{z}(\alpha) = Q_{\tilde{x}(\alpha)^{1/2}}\tilde{s}(\alpha).
\end{aligned}$$

As a word of caution, since p need not lie in $\mathcal{C}(x(\alpha), s(\alpha))$, $\tilde{x}(\alpha)$ and $\tilde{s}(\alpha)$ do not necessarily operator commute. We collect some basic properties of the scaled directions and the Newton system.

Lemma 3.4 *Given the Newton equations, the following identities hold:*

$$\begin{aligned}
\tilde{s}(\alpha) \bullet \tilde{x}(\alpha) &= (1-\alpha)\,\tilde{s} \bullet \tilde{x} + \alpha\,\sigma\mu e + \alpha^2\, \triangle\tilde{s} \bullet \triangle\tilde{x}, \\
\langle \tilde{s}, \tilde{x}\rangle &= \langle s, x\rangle\,, \textit{ and} \\
\tilde{\mu}(\alpha) &= \mu(1-\alpha+\sigma\alpha) + \alpha^2 \frac{\langle \triangle s, \triangle x\rangle}{r}.
\end{aligned}$$

Proof : The first equality follows by direct expanding the third equation of the scaled Newton system. The second follows because

$$\langle \tilde{s}, \tilde{x}\rangle = \left\langle Q_{p^{-1}}s, Q_p x\right\rangle = \langle s, x\rangle\,.$$

For the last equation, we use the third Newton equation in (3.4) to get

$$\tilde{\mu}(\alpha) = \frac{\langle \tilde{s}(\alpha), \tilde{x}(\alpha)\rangle}{r} = \frac{\langle \tilde{s}, \tilde{x}\rangle}{r} + \alpha\frac{\langle \tilde{s}, \triangle\tilde{x}\rangle + \langle \triangle\tilde{s}, \tilde{x}\rangle}{r} + \alpha^2\frac{\langle \triangle\tilde{s}, \triangle\tilde{x}\rangle}{r} = \mu(1-\alpha+\sigma\alpha) + \alpha^2\frac{\langle \triangle\tilde{s}, \triangle\tilde{x}\rangle}{r}.$$

□

The following result is very essential in obtaining the bounds on the step lengths.

Lemma 3.5 *Let* $(x, s) \in$ *int* $\mathcal{K} \times$ *int* $\mathcal{K}$*. Then* $\lambda_{min}(s \bullet x) \leq \lambda_{min}(z)$ *and equality holds if* x *and* s *operator commute.*

Proof : The proof outline follows Lemma 30 in [9]. First observe that $Q_{x^{1/2},x^{-1/2}} Q_{x^{1/2}} = L(x)$, because

$$\begin{aligned} Q_{x^{1/2},x^{-1/2}} Q_{x^{1/2}} &= Q_{x^{1/2}}(2L(x^{-1/2})L(x^{1/2}) - I) \\ &= 2(Q_{x^{1/2}} L(x^{-1/2}))L(x^{1/2}) - Q_{x^{1/2}} \\ &= 2L^2(x^{1/2}) - Q_{x^{1/2}} = L(x). \end{aligned}$$

Here, we used part (a) of Lemma 2.4. As a result we have $Q_{x^{1/2},x^{-1/2}} z = Q_{x^{1/2},x^{-1/2}} Q_{x^{1/2}} s = x \bullet s$.

In Lemma 30 in [9], it is shown that $\text{Trace}\,(Q_{x^{1/2},x^{-1/2}} u) = \text{Trace}\,(u)$. Note that by Lemma 2.12 we know that $\mathcal{K} \subset L(x)(\mathcal{K}) = Q_{x^{1/2},x^{-1/2}} Q_{x^{1/2}}(\mathcal{K}) = Q_{x^{1/2},x^{-1/2}}(\mathcal{K})$, as $Q_{x^{1/2}}$ is an automorphism of $\mathcal{K}$. The result follows from the following two chains of relations.

$$\lambda_{\min}(s \bullet x) = \min_u \frac{\langle u, (s \bullet x) \bullet u \rangle}{\langle u, u \rangle} = \min_{\text{Trace}\,(u^2)=1} \left\langle u^2, s \bullet x \right\rangle = \min_{\text{Trace}\,(u^2)=1} \left\langle u^2, Q_{x^{1/2},x^{-1/2}} z \right\rangle$$

$$\begin{aligned} \min_{\text{Trace}\,(u^2)=1} \left\langle u^2, Q_{x^{1/2},x^{-1/2}} z \right\rangle &= \min_{\text{Trace}\,(u^2)=1} \left\langle z, Q_{x^{1/2},x^{-1/2}} u^2 \right\rangle \\ &\leq \min_{\text{Trace}\,(Q_{x^{1/2},x^{-1/2}} u^2)=1} \left\{ \left\langle z, Q_{x^{1/2},x^{-1/2}} u^2 \right\rangle : Q_{x^{1/2},x^{-1/2}} u^2 \in \mathcal{K} \right\} \\ &= \min \left\{ \langle z, t \rangle : \ \text{Trace}\,(t) = 1, t \in Q_{x^{1/2},x^{-1/2}}(\mathcal{K}) \right\} \\ &\leq \min \left\{ \langle z, t \rangle : \ \text{Trace}\,(t) = 1, t \in \mathcal{K} \right\} \\ &= \min_{\text{Trace}\,(v^2)=1} \left\langle z, v^2 \right\rangle = \lambda_{\min}(z) \end{aligned}$$

The equality when $\tilde{x}$ and $\tilde{s}$ operator commute is established in Lemma 30 in [9]. Hence the proof of the lemma is complete. $\square$

As a consequence, using Proposition 2.11 and the definition of $\mathcal{N}_{-\infty}(\gamma)$, let us note that

$$\lambda_{\min}(\tilde{s} \bullet \tilde{x}) = \lambda_{\min}(\tilde{z}) = \lambda_{\min}(z) \geq (1 - \gamma)\mu.$$

We find an interval for which $(x(\alpha), s(\alpha))$ lies in the neighborhood.

Lemma 3.6 *Let* $\delta_x = \|\triangle \tilde{x}\|_F$ *and* $\delta_s = \|\triangle \tilde{s}\|_F$*. If* $(x, s) \in \mathcal{N}_{-\infty}(\gamma)$*, then* $(x(\alpha), s(\alpha)) \in \mathcal{N}_{-\infty}(\gamma)$ *for all* $0 \leq \alpha \leq \hat{\alpha}_1$*, where*

$$\hat{\alpha}_1 := \frac{\gamma\sigma \langle s, x \rangle}{2(r + 1 - \gamma)\delta_x \delta_s}\,. \tag{3.8}$$

Proof : We first bound the left and right hand side of the inequality defining the neighborhood $\mathcal{N}_{-\infty}(\gamma)$. To begin with a bound on the eigenvalue of $z(\alpha)$, we have

$$\begin{aligned} \lambda_{\min}(z(\alpha)) = \lambda_{\min}(\tilde{z}(\alpha)) &\geq \lambda_{\min}(\tilde{s}(\alpha) \bullet \tilde{x}(\alpha)) \\ &= \lambda_{\min}((1 - \alpha)\tilde{s} \bullet \tilde{x} + \alpha\sigma\mu e + \alpha^2 \triangle \tilde{s} \bullet \triangle \tilde{x}) \\ &\geq (1 - \alpha)\lambda_{\min}(\tilde{s} \bullet \tilde{x}) + \alpha\sigma\mu - \alpha^2 \delta_x \delta_s \\ &\geq (1 - \alpha)(1 - \gamma)\mu + \alpha\sigma\mu - \alpha^2 \delta_x \delta_s. \end{aligned}$$

The first equality follows from part (ii) of Proposition 2.11, the first inequality follows from Lemma 3.5, the second inequality follows from Lemma 2.9 and the last inequality follows because $(\tilde{x}, \tilde{s}) \in \mathcal{N}_{-\infty}(\gamma)$. Using Lemma 3.4 and Cauchy-Schwarz we can see that

$$\begin{aligned}(1-\gamma)\mu(\alpha) &= (1-\gamma)(\mu(1-\alpha+\sigma\alpha)+\alpha^2\frac{\langle\triangle s,\triangle x\rangle}{r}) \\ &\leq (1-\gamma)\left[\mu(1-\alpha+\sigma\alpha)+\alpha^2\frac{\delta_x\delta_s}{r}\right].\end{aligned}$$

Using $\langle s, x\rangle = \mu r$, we can see that

$$(1-\alpha)(1-\gamma)\mu+\alpha\sigma\mu-\alpha^2\delta_x\delta_s \geq (1-\gamma)\left[\mu(1-\alpha+\sigma\alpha)+\alpha^2\frac{\delta_x\delta_s}{r}\right]$$

holds for all $\alpha \in [0, 2\hat{\alpha}_1]$. Since the right hand side of the inequality is positive for all $\alpha \in [0, \hat{\alpha}_1]$, $\lambda_{\min}(z(\alpha)) > 0$ for all $\alpha \in [0, \hat{\alpha}_1]$. Let α_0 be the least $\alpha \leq \hat{\alpha}_1$ such that $x(\alpha), s(\alpha) \in \mathcal{K}$ for all $\alpha \leq \alpha_0$ and $x(\alpha_0) \in \partial\mathcal{K}$ (or $s(\alpha_0) \in \partial\mathcal{K}$). Then $\lambda_{\min}(z(\alpha_0)) = 0$, which is a contradiction. Hence $x(\alpha), s(\alpha) \in \operatorname{int}\mathcal{K}$. Hence $(x(\alpha), s(\alpha), y(\alpha)) \in \mathcal{N}_{-\infty}(\gamma)$ for all $\alpha \in [0, \hat{\alpha}_1]$. □

Note that the length of the interval obtained depends on the size of the scaled Newton directions.

For the feasibility condition in Step 3 we want an $\hat{\alpha}_2$ such that (3.7) holds for all $(x(\alpha), s(\alpha))$, $\alpha \in [0, \hat{\alpha}_2]$. Using Lemma 3.4, the feasibility condition on (x, s) and Cauchy-Schwarz, we get

$$\begin{aligned}\frac{\langle s(\alpha), x(\alpha)\rangle}{\langle s_0, x_0\rangle}-\bar{\phi}(1-\alpha) &= \frac{\langle s, x\rangle}{\langle s_0, x_0\rangle}(1+\alpha(\sigma-1))+\alpha^2\frac{\langle\triangle s,\triangle x\rangle}{\langle s_0, x_0\rangle}-\bar{\phi}(1-\alpha) \\ &= \left(\frac{\langle s, x\rangle}{\langle s_0, x_0\rangle}-\bar{\phi}\right)(1-\alpha)+\alpha\sigma\frac{\langle s, x\rangle}{\langle s_0, x_0\rangle}+\alpha^2\frac{\langle\triangle s,\triangle x\rangle}{\langle s_0, x_0\rangle} \\ &\geq \frac{\alpha}{\langle s_0, x_0\rangle}\left(\sigma\langle s, x\rangle-\alpha\delta_x\delta_s\right).\end{aligned}$$

Therefore the condition $\langle s(\alpha), x(\alpha)\rangle - \bar{\phi}(1-\alpha)\langle s_0, x_0\rangle \geq 0$ holds for all $\alpha \in [0, \hat{\alpha}_2]$, where

$$\hat{\alpha}_2 := \frac{\sigma\langle s, x\rangle}{\delta_x\delta_s}. \tag{3.9}$$

For the last condition in Step 3, Cauchy-Schwarz yields

$$\begin{aligned}\langle s(\alpha), x(\alpha)\rangle &= \langle s, x\rangle(1-\alpha(1-\sigma))+\alpha^2\langle\triangle s,\triangle x\rangle \\ &\leq \langle s, x\rangle\left(1-\alpha(1-\sigma)+\alpha^2\frac{\delta_x\delta_s}{\langle s, x\rangle}\right).\end{aligned}$$

It suffices to have

$$\left[1-\alpha(1-\sigma)+\alpha^2\frac{\delta_x\delta_s}{\langle s, x\rangle}\right]-(1-\alpha(1-\beta))=\alpha\left(\alpha\frac{\delta_x\delta_s}{\langle s, x\rangle}-(\beta-\sigma)\right)\leq 0.$$

Solving for α from the above inequality, we can see that the last condition holds for all $\alpha \in [0, \hat{\alpha}_3]$, where

$$\hat{\alpha}_3 := \frac{(\beta-\sigma)\langle s, x\rangle}{\delta_x\delta_s}. \tag{3.10}$$

So far, we have obtained a lower bound on the step sizes in terms of δ_x, δ_s and $\langle s, x\rangle$. Now, we will obtain a bound on $\frac{\delta_x \delta_s}{\langle s,x\rangle}$, which appears in (3.8), (3.9, and (3.10). Let us introduce the operator, $G := L(\tilde{s})^{-1}L(\tilde{x})$, which is useful in bounding $\delta_x\delta_s$. Recall the third scaled Newton equation:

$$L(\tilde{s})\triangle\tilde{x} + L(\tilde{x})\triangle\tilde{s} = \sigma\mu e - L(\tilde{s})L(\tilde{x})e.$$

Since $\tilde{x}$ and $\tilde{s}$ operator commute, and G is a symmetric matrix, by multiplying this equation by $(L(\tilde{x})L(\tilde{s}))^{-1/2}$, we get

$$G^{-1/2}\triangle\tilde{x} + G^{1/2}\triangle\tilde{s} = \sigma\mu(L(\tilde{x})L(\tilde{s}))^{-1/2}e - G^{1/2}\tilde{s} =: h.$$

The analysis of **IIPM** is intricate because $\left\langle G^{1/2}\triangle\tilde{s}, G^{-1/2}\triangle\tilde{x}\right\rangle = \langle\triangle s, \triangle x\rangle \neq 0$. Now let us define

$$t^2 := \|G^{1/2}\triangle\tilde{s}\|_F^2 + \|G^{-1/2}\triangle\tilde{x}\|_F^2.$$

The following proposition will lead to a bound on the size of $\frac{\delta_x\delta_s}{\langle s,x\rangle}$.

Proposition 3.7 *$t_k^2 \leq \omega\langle s_k, x_k\rangle$, where ω is a constant independent of k.*

Before we prove the proposition, let us pause here to see its relevance in bounding $\delta_x\delta_s$. We state the following technical, but useful result (Lemma 33 in [9]).

Lemma 3.8 *Let $u, v \in \mathcal{J}$ and G be a positive definite self-adjoint operator. Then*

$$\|u\|_F\|v\|_F \leq \frac{1}{2}\sqrt{\kappa_G}\left(\|G^{1/2}u\|_F^2 + \|G^{-1/2}v\|_F^2\right),$$

where $\kappa_G = \frac{\lambda_{max}(G)}{\lambda_{min}(G)}$ is the condition number of G.

Note that in our application, κ_G may depend on the iteration number k, but the following lemma provides a bound on the condition number of G for the methods we are interested in (see Lemma 36 in [9]).

Lemma 3.9 *For the NT method $\kappa_G = 1 =: \kappa$. For the xs and the sx methods,*

$$\text{if } (x, s) \in \mathcal{N}_{-\infty}(\gamma), \text{then } \kappa_G \leq \frac{r}{1-\gamma} =: \kappa.$$

Using the above lemmas, we have the following bound on $\delta_x\delta_s$:

$$\delta_x\delta_s \leq \frac{t^2}{2}\sqrt{\kappa} \leq \frac{\omega}{2}\sqrt{\kappa}\langle s, x\rangle. \tag{3.11}$$

Now we prove the proposition.

Proof of Proposition 3.7 : We first note the following identity:

$$\begin{aligned}\|G^{1/2}\triangle\tilde{s} + G^{-1/2}\triangle\tilde{x}\|_F^2 &= \|G^{1/2}\triangle\tilde{s}\|_F^2 + \|G^{-1/2}\triangle\tilde{x}\|_F^2 + 2\left\langle G^{1/2}\triangle\tilde{s}, G^{-1/2}\triangle\tilde{x}\right\rangle \\ &= \|G^{1/2}\triangle\tilde{s}\|_F^2 + \|G^{-1/2}\triangle\tilde{x}\|_F^2 + 2\langle\triangle\tilde{s}, \triangle\tilde{x}\rangle.\end{aligned}$$

Using what we just derived and Lemmas 34 and 35 of [9], we can see that for $h = \sigma\mu(L(\tilde{x})L(\tilde{s}))^{-1/2}e - G^{1/2}\tilde{s}$,

$$\|h\|_F^2 = t^2 + 2\langle\triangle\tilde{s}, \triangle\tilde{x}\rangle = \sum_i^r \frac{(\sigma\mu - \lambda_i(\tilde{z}))^2}{\lambda_i(\tilde{z})} \leq \left(1 - 2\sigma + \frac{\sigma^2}{1-\gamma}\right)\langle s, x\rangle. \tag{3.12}$$

We take a small detour to introduce some convenient notation which helps us in stating a key claim in the proof of this proposition, and is also used in the arguments for polynomiality of convergence. Let us assume a reference point (u_0, v_0, r_0) feasible to the equality constraints (and not necessarily in the cone) such that $x_0 - u_0, s_0 - v_0 \in \text{int}\,\mathcal{K}$, where (x_0, s_0, y_0) is the initial iterate in **IIPM**. This condition is easily satisfied by scaling the initial point for any given (u_0, v_0, r_0). For a given sequence of iterates $\{(x_k, s_k, y_k)\}$ we define:

$$\begin{aligned} u_{k+1} &= (1-\alpha_p^k)(u_k - x_k) + x_{k+1}; \\ r_{k+1} &= (1-\alpha_d^k)(r_k - y_k) + y_{k+1}; \\ v_{k+1} &= (1-\alpha_d^k)(v_k - s_k) + s_{k+1}. \end{aligned}$$

From the above definitions, we can observe the following properties:

$$\begin{gathered} x_{k+1} - u_{k+1} = \phi_p^{k+1}(x_0 - u_0) \in \text{int}\,\mathcal{K}; \\ s_{k+1} - v_{k+1} = \phi_d^{k+1}(s_0 - v_0) \in \text{int}\,\mathcal{K}; \\ Au_k = b \text{ and } A^* r_k + v_k = c \text{ for all } k; \\ A(x_k + \triangle x_k - u_k) = A(x + \triangle x_k) - Au_k = b - b = 0; \\ A^*(y_k + \triangle y_k - r_k) + s_k + \triangle s_k - v_k = 0. \end{gathered} \tag{3.13}$$

(The third line holds for $k = 0$ by assumption, and then holds for all k by induction using the last two lines.) The following result is the key to proving the proposition:

Claim 3.10

$$\langle s, x\rangle \frac{\langle s_0 - v_0, x_0 - u_0\rangle}{\langle s_0, x_0\rangle} + \langle \triangle s, \triangle x\rangle + \xi t\sqrt{\langle s, x\rangle} \geq 0,$$

where

$$\xi = \xi_k := \sqrt{\frac{r}{1-\gamma}}\left[\frac{\langle s, x-u\rangle + \langle s-v, x\rangle}{\langle s, x\rangle}\right]. \tag{3.14}$$

The claim is proved in the appendix. For now, we substitute $\langle \triangle s, \triangle x\rangle$ from the inequality in (3.12), to get

$$t^2 \quad \leq \quad \langle s, x\rangle\, \bar{\chi} + 2\sqrt{\langle s, x\rangle}\, \xi t,$$

where

$$\bar{\chi} := 1 - 2\sigma + \frac{\sigma^2}{1-\gamma} + 2\left\{\frac{\langle s_0 - v_0, x_0 - u_0\rangle}{\langle s_0, x_0\rangle}\right\} \text{ is independent of } k. \tag{3.15}$$

Therefore,

$$t_k^2 \leq \langle s_k, x_k\rangle \left(\xi_k + \sqrt{\xi_k^2 + \bar{\chi}}\right)^2.$$

From Lemma 4.1 in [8], we have the following useful bound: Let (x, s, y) be any iterate generated by **IIPM** and (x^*, s^*, y^*) be an optimal solution to (P) and (D). Then

$$\frac{\langle s, x-u\rangle + \langle s-v, x\rangle}{\langle s, x\rangle} \leq 1 + \frac{\langle s^*, x_0 - u_0\rangle + \langle s_0 - v_0, x^*\rangle + \langle s_0 - v_0, x_0 - u_0\rangle}{\langle s_0, x_0\rangle}.$$

Therefore ξ_k is uniformly bounded by $\bar{\xi}$ where

$$\bar{\xi} = \sqrt{\frac{r}{1-\gamma}}\left\{1 + \frac{\langle s^*, x_0 - u_0\rangle + \langle s_0 - v_0, x^*\rangle + \langle s_0 - v_0, x_0 - u_0\rangle}{\langle s_0, x_0\rangle}\right\}. \tag{3.16}$$

Hence we can choose ω to be

$$\omega = \left(\bar{\xi} + \sqrt{\bar{\xi}^2 + \bar{\chi}}\right)^2. \tag{3.17}$$

□

Recall that the conclusion of Proposition 3.7 led to a bound on $\delta_x\delta_s$ in (3.11). Hence we can bound from below the $\hat{\alpha}$'s in (3.8), (3.9), and (3.10) in the following way:

$$\hat{\alpha}_1 = \frac{\gamma\sigma\langle s,x\rangle}{2(r+1-\gamma)\delta_x\delta_s} \geq \frac{\gamma\sigma\langle s,x\rangle}{2(r+1-\gamma)\frac{\omega}{2}\sqrt{\kappa}\langle s,x\rangle} = \frac{\gamma\sigma}{(r+1-\gamma)\omega\sqrt{\kappa}} \quad =: \quad \bar{\alpha}_1, \tag{3.18}$$

$$\hat{\alpha}_2 = \frac{\sigma\langle s,x\rangle}{\delta_x\delta_s} \geq \frac{2\sigma}{\omega\sqrt{\kappa}} \quad =: \quad \bar{\alpha}_2, \text{ and} \tag{3.19}$$

$$\hat{\alpha}_3 = \frac{(\beta-\sigma)\langle s,x\rangle}{\delta_x\delta_s} \geq \frac{2(\beta-\sigma)}{\omega\sqrt{\kappa}} \quad =: \quad \bar{\alpha}_3. \tag{3.20}$$

Taking into account the above bounds, we define

$$\alpha^* := \min\left(1, \frac{\gamma\sigma}{(r+1-\gamma)\omega\sqrt{\kappa}}, \frac{2\sigma}{\omega\sqrt{\kappa}}, \frac{2(\beta-\sigma)}{\omega\sqrt{\kappa}}\right) = \Omega\left(\frac{1}{r\omega\sqrt{\kappa}}\right). \tag{3.21}$$

For this choice of α^*, for $\alpha \in [0,\alpha^*]$ all the conditions in Step 3 (and hence Step 4 by the remarks following the algorithm) of **IIPM** are satisfied. This bound implies the global convergence of **IIPM** by Theorem 3.3. Also, note that since $\langle\triangle\tilde{s},\triangle\tilde{x}\rangle = 0$ for feasible-IPMs, (3.12) implies that

$$t^2 \leq \left(1 - 2\sigma + \frac{\sigma^2}{1-\gamma}\right)\langle s,x\rangle.$$

Hence ω in the case of feasible-IPMs is replaced by a constant independent of the data and we obtain $\mathcal{O}(r\sqrt{\kappa}\ln(1/\epsilon))$ iteration complexity for feasible-IPMs by Theorem 3.3. This is the bound obtained by Schmieta and Alizadeh in [9].

With some restrictions on the size of initial points, we can show that ω is polynomially bounded and consequently obtain the polynomial convergence of **IIPM**. Let (u_0, r_0, v_0) be the solution to

$$\min\{\|u\|_F \;:\; Au = b\} \text{ and } \min\{\|v\|_F \;:\; A^*r + v = c\}, \text{ and}$$

$$x_0 = s_0 = \rho_0 e \in \text{int}\,\mathcal{K},$$

where e is the identity element of the Euclidean Jordan algebra and $\rho_0 > \max(\|u_0\|_2,\ \|v_0\|_2)$. This implies that $x_0 - u_0 \in$ int $\mathcal{K}$ and $s_0 - v_0 \in$ int $\mathcal{K}$. Let us assume that for some constant $\Psi > 0$,

$$\rho_0 \geq \frac{1}{\Psi}\rho^* := \frac{1}{\Psi}\min\{\max(\|x^*\|_2, \|s^*\|_2) : (x^*, s^*) \text{ solves } (P) \text{ and } (D)\}. \tag{3.22}$$

(Note that we can always increase ρ_0.) Now we can obtain a bound for ω. First let us note two useful facts: $\|\cdot\|_F \leq \sqrt{r}\|\cdot\|_2$ and $\langle s_0, x_0\rangle = \rho_0^2 r$. Therefore, using Cauchy-Schwarz, we can see that $\langle p,q\rangle \leq \|p\|_F\|q\|_F \leq r\|p\|_2\|q\|_2$. Now we can bound $\bar{\xi}$ in (3.16) as follows:

$$\begin{aligned}
\bar{\xi} &= \sqrt{\frac{r}{1-\gamma}}\left\{1 + \frac{\langle s^*, x_0 - u_0\rangle + \langle s_0 - v_0, x^*\rangle + \langle s_0 - v_0, x_0 - u_0\rangle}{\langle s_0, x_0\rangle}\right\} \\
&\leq \sqrt{\frac{r}{1-\gamma}}\left\{1 + \frac{2\rho^*\rho_0 r + 2\rho^*\rho_0 r + 4\rho_0^2 r}{\rho_0^2 r}\right\} \\
&= \sqrt{\frac{r}{1-\gamma}}\left\{5 + 4\frac{\rho^*}{\rho_0}\right\} \leq \sqrt{\frac{r}{1-\gamma}}(5 + 4\Psi) \text{ (using (3.22)).}
\end{aligned}$$

For a bound on $\bar{\chi}$ in (3.15), we have

$$\bar{\chi} = 1 - 2\sigma + \frac{\sigma^2}{1-\gamma} + 2\left\{\frac{\langle s_0 - v_0, x_0 - u_0\rangle}{\langle s_0, x_0\rangle}\right\} \leq 1 + \frac{1}{1-\gamma} + 2\cdot\frac{4\rho_0^2 r}{\rho_0^2 r} = 9 + \frac{1}{1-\gamma}.$$

Therefore,

$$\omega = \left(\bar{\xi} + \sqrt{\bar{\xi}^2 + \bar{\chi}}\right)^2 = O(r). \tag{3.23}$$

Having obtained bounds on the key quantities defining α^* in (3.21), we state our main theorem.

Theorem 3.11 *Suppose that* $\kappa_G \leq \kappa < \infty$ *for all iterations of* **IIPM**. *Then* **IIPM** *will terminate in* $\mathcal{O}(\sqrt{\kappa}r^2\ln(1/\epsilon^*))$ *iterations. Hence the* NT *method takes* $\mathcal{O}(r^2\ln(1/\epsilon^*))$ *iterations, and the* xs *and the* sx *methods take* $\mathcal{O}(r^{2.5}\ln(1/\epsilon^*))$ *iterations.*

Proof : For any $\alpha \in [0, \alpha^*]$, α^* as defined in (3.21), all the conditions in Step 3 of **IIPM** are satisfied. Thus by Theorem 3.3, **IIPM** will terminate in $k = \lceil\frac{1}{\alpha^*}\rceil\ln\left(\frac{1}{\epsilon^*}\right) = O\left(\sqrt{\kappa}r^2\ln(1/\epsilon^*)\right)$ iterations.

The second part of the theorem follows from the bound on κ in Lemma 3.11 for the xs, the xs and the NT method. ☐

4 Conclusion

We have established polynomial convergence of infeasible-interior-point methods for three important methods: the xs, sx and the Nesterov-Todd (NT) method. To our knowledge this is the first time an infeasible-interior-point method has been analysed for the NT-method using the $\mathcal{N}_{-\infty}$ neighborhood for both semidefinite programming and conic programs over symmetric cones. The algorithm presented here is closely related to the algorithms used in practice to solve large-scale linear programs. The complexity obtained for the NT-method (in this general setting) coincides with the bound obtained for linear programming by Zhang. The work by Rangarajan and Todd shows convergence of the NT-method using another neighborhood defined globally over the cone.

Acknowledgements The author would like to thank Mike Todd for carefully reading through this manuscipt and providing valuable feedback, and for insightful discussions. This work would not have been possible without his encouragement and support.

References

[1] J. Faraut, A. Koranyi, Analysis on Symmetric Cones. *Oxford University Press*, 1994.

[2] L. Faybusovich, Linear systems in Jordan algebras and primal-dual interior-point algorithms, *Journal of Computational and Applied Mathematics*, 86:149-175, 1997.

[3] O. Güler, Barrier functions in interior-point methods. *Mathematics of Operations Research*, 21:860-885, 1996.

[4] Jordan, P. and von Neumann, J. and Wigner, E., On algebraic generalization of the quantum mechanical formalism, *Annals of Mathematics*, 35: 29-64, 1934.

[5] R. D. C. Monteiro, Y. Zhang, A unified analysis for a class of path-following primal-dual interior-point algorithms for semidefinite programming. *Mathematical Programming*, 81:281-299, 1998.

[6] Yu. E. Nesterov, A. S. Nemirovskii, Interior-Point Polynomial Algorithms in Convex Programming. *SIAM, Philadelphia*, 1994.

[7] Yu. E. Nesterov, M. J. Todd, Self-scaled barriers and interior point methods for convex programming. *Mathematics of Operations Research*, 22(1):1-42, 1997.

[8] B. Rangarajan, M. J. Todd, Convergence of infeasible-interior-point methods for self-scaled conic programming. Technical Report 1388, School of OR&IE, Cornell University.

[9] S. H. Schmieta, F. Alizadeh, Extension of primal–dual interior–point algorithm to symmetric cones. *Mathematical Programming*, 96(3):409-438, 2003.

[10] Y. Zhang, On extending some primal-dual interior-point algorithms from linear programming to semidefinite programming. *SIAM Journal on Optimization*, 8(1):365-386, 1998.

5 Appendix

Claim 3.10

$$\langle s,x\rangle \frac{\langle s_0-v_0,x_0-u_0\rangle}{\langle s_0,x_0\rangle}+\langle \triangle s,\triangle x\rangle+\xi t\sqrt{\langle s,x\rangle}\geq 0,$$

where

$$\xi=\xi_k:=\sqrt{\frac{r}{1-\gamma}}\left[\frac{\langle s,x-u\rangle+\langle s-v,x\rangle}{\langle s,x\rangle}\right].$$

Proof : By expanding $\langle \triangle s+s-v,\triangle x+x-u\rangle$ and using (3.13), we find that

$$\langle \triangle s,\triangle x\rangle+\langle s-v,x-u\rangle+\langle \triangle s,x-u\rangle+\langle s-v,\triangle x\rangle=0. \tag{5.1}$$

We will now bound the last three terms in the expansion. First, using Cauchy-Schwarz, we see that

$$\langle s-v,\triangle x\rangle=\langle \tilde{s}-\tilde{v},\triangle\tilde{x}\rangle=\left\langle G^{1/2}(\tilde{s}-\tilde{v}),G^{-1/2}\triangle\tilde{x}\right\rangle\leq\|G^{1/2}(\tilde{s}-\tilde{v})\|_F\|G^{-1/2}\triangle\tilde{x}\|_F\leq\|G^{1/2}(\tilde{s}-\tilde{v})\|_F t. \tag{5.2}$$

Next, note that

$$\|G^{1/2}(\tilde{s}-\tilde{v})\|_F^2=\left\langle G^{1/2}(\tilde{s}-\tilde{v}),G^{1/2}(\tilde{s}-\tilde{v})\right\rangle=\langle \tilde{s}-\tilde{v},G(\tilde{s}-\tilde{v})\rangle. \tag{5.3}$$

Since $\tilde{x}$ and $\tilde{s}$ operator commute, operators G and $Q_{\tilde{x}}$ commute. Hence we have

$$\langle \tilde{s}-\tilde{v},G(\tilde{s}-\tilde{v})\rangle=\left\langle Q_{\tilde{x}}^{1/2}(\tilde{s}-\tilde{v}),Q_{\tilde{x}}^{-1}GQ_{\tilde{x}}^{1/2}(\tilde{s}-\tilde{v})\right\rangle\leq\lambda_{\max}(Q_{\tilde{x}}^{-1}G)\|Q_{\tilde{x}}^{1/2}(\tilde{s}-\tilde{v})\|_F^2. \tag{5.4}$$

We state the following lemma and prove it later (the second part is analogous to Lemma 2.2 in [8]).

Lemma 5.1 *If $G=L(\tilde{s})^{-1}L(\tilde{x})$, then $\lambda_{max}(Q_{\tilde{x}}^{-1}G)=\frac{1}{\lambda_{min}(\tilde{z})}$. If $q\in\mathcal{K}$ and $\tilde{q}=Q_{p^{-1}}q$, then*

$$\|Q_{\tilde{x}^{1/2}}\tilde{q}\|_F\leq\langle\tilde{q},\tilde{x}\rangle=\langle q,x\rangle.$$

By substituting $q = s - v$ in the second part of Lemma 5.1, we get $\|Q_{\tilde{x}}^{1/2}(\tilde{s} - \tilde{v})\|_F \leq \langle s - v, x\rangle$. Using (5.3), and (5.4), we see that

$$\|G^{1/2}(\tilde{s} - \tilde{v})\|_F^2 \leq \lambda_{\max}(Q_{\tilde{x}}^{-1}G)\|Q_{\tilde{x}}^{1/2}(\tilde{s} - \tilde{v})\|_F^2 \leq \frac{1}{\lambda_{\min}(z)} \langle s - v, x\rangle^2 .$$

As $(x, s) \in \mathcal{N}_{-\infty}(\gamma)$, $\lambda_{\min}(z) \geq (1-\gamma)\mu$ and from (5.2) we have

$$\langle s - v, \triangle x\rangle \leq \|G^{1/2}(\tilde{s} - \tilde{v})\|_F \|G^{-1/2}\triangle\tilde{x}\|_F \leq \sqrt{\frac{1}{(1-\gamma)\mu}} \langle s - v, x\rangle\, t.$$

Similarly it can be shown that

$$\langle \triangle s, x - u\rangle \leq \sqrt{\frac{1}{(1-\gamma)\mu}} \langle s, x - u\rangle\, t.$$

Also using the feasibility condition (3.7), (3.13), and $\bar{\phi} \leq 1$, we get

$$\langle s - v, x - u\rangle \leq \bar{\phi}^2 \langle s_0 - v_0, x_0 - u_0\rangle \leq \frac{\langle s, x\rangle}{\langle s_0, x_0\rangle} \langle s_0 - v_0, x_0 - u_0\rangle .$$

Substituting the above bounds into (5.1) and using (3.14), we get

$$\begin{aligned} 0 &\leq \langle \triangle s, \triangle x\rangle + \frac{\langle s, x\rangle}{\langle s_0, x_0\rangle} \langle s_0 - v_0, x_0 - u_0\rangle + \sqrt{\frac{1}{(1-\gamma)\mu}} \langle s, x - u\rangle\, t + \sqrt{\frac{1}{(1-\gamma)\mu}} \langle s - v, x\rangle\, t \\ &= \langle \triangle s, \triangle x\rangle + \langle s, x\rangle \frac{\langle s_0 - v_0, x_0 - u_0\rangle}{\langle s_0, x_0\rangle} + \xi t\sqrt{\langle s, x\rangle}. \end{aligned}$$

□

Proof of Lemma 5.1 : Suppose $\{\lambda_i : 1 \leq i \leq r\}$ are the eigenvalues of $\tilde{x}$ with eigenvectors $\{c_i : 1 \leq i \leq r\}$ from the spectral decomposition of type II. Since $\tilde{x}$ and $\tilde{s}$ operator commute, they share the same Jordan frame. So, let the corresponding eigenvalues of $\tilde{s}$ be $\{\mu_i : 1 \leq i \leq r\}$. Then using Lemma 2.4 and Theorem 2.14, we have the following two results:

$$\begin{aligned} \lambda_{\max}\left(Q_{\tilde{x}}^{-1}L(\tilde{s})^{-1}L(\tilde{x})\right) &= \lambda_{\max}\left(Q_{\tilde{x}^{-1}}L(\tilde{x})L(\tilde{s})^{-1}\right) \\ &= \lambda_{\max}\left(L(\tilde{x}^{-1})L(\tilde{s})^{-1}\right) \\ &= \max_{1\leq i\leq j\leq r}\left[\left(\frac{1}{\lambda_i} + \frac{1}{\lambda_j}\right)\frac{1}{\mu_i + \mu_j}\right], \text{ and} \end{aligned}$$

$$\lambda_{\min}(\tilde{z})^2 = \lambda_{\min}\left(Q_{\tilde{x}^{1/2}}\tilde{s}\right)^2 = \lambda_{\min}\left(Q_{\tilde{x}^{1/2}}Q_{\tilde{s}}Q_{\tilde{x}^{1/2}}\right) = \lambda_{\min}\left(Q_{\tilde{s}}Q_{\tilde{x}}\right) = \min_{1\leq i\leq j\leq r} \lambda_i\lambda_j\mu_i\mu_j.$$

It is staighforward to verify that

$$\left[\left(\frac{1}{\lambda_i} + \frac{1}{\lambda_j}\right)\frac{1}{\mu_i + \mu_j}\right] \leq \max\left(\frac{1}{\lambda_i\mu_i}, \frac{1}{\lambda_j\mu_j}\right), \text{ and } \lambda_i\lambda_j\mu_i\mu_j \geq \min\left((\lambda_i\mu_i)^2, (\lambda_j\mu_j)^2\right).$$

This proves the first part of the lemma.

For the second part, the equality is easy to see. To show the inequality, note that

$$\lambda_{\max}(Q_{\tilde{x}^{1/2}}\tilde{q}) \le \|Q_{\tilde{x}^{1/2}}\tilde{q}\|_F.$$

For $p := \frac{Q_{\tilde{x}^{1/2}}\tilde{q}}{\|Q_{\tilde{x}^{1/2}}\tilde{q}\|_F}$, $\lambda_{\max}(p) \le 1$ and hence $e - p \in \mathcal{K}$. Since

$$\langle \tilde{q}, \tilde{x} \rangle = \langle \tilde{q}, Q_{\tilde{x}^{1/2}} e \rangle = \langle Q_{\tilde{x}^{1/2}}\tilde{q}, e \rangle = \langle Q_{\tilde{x}^{1/2}}\tilde{q}, e - p \rangle + \langle Q_{\tilde{x}^{1/2}}\tilde{q}, p \rangle ,$$

we have

$$\langle \tilde{q}, \tilde{x} \rangle \;\; = \;\; \langle Q_{\tilde{x}^{1/2}}\tilde{q}, e - p \rangle + \langle Q_{\tilde{x}^{1/2}}\tilde{q}, p \rangle \ge \langle Q_{\tilde{x}^{1/2}}\tilde{q}, p \rangle = \|Q_{\tilde{x}^{1/2}}\tilde{q}\|_F.$$

□