Math. Program., Ser. A 96: 409–438 (2003)

Digital Object Identifier (DOI) 10.1007/s10107-003-0380-z

S.H. Schmieta · F. Alizadeh

Extension of primal-dual interior point algorithms to symmetric cones

Received: April 2000 / Accepted: May 2002
Published online: March 28, 2003 –

Abstract. In this paper we show that the so-called commutative class of primal-dual interior point algorithms which were designed by Monteiro and Zhang for semidefinite programming extends word-for-word to optimization problems over all symmetric cones. The machinery of Euclidean Jordan algebras is used to carry out this extension. Unlike some non-commutative algorithms such as the $XS+SX$ method, this class of extensions does not use concepts outside of the Euclidean Jordan algebras. In particular no assumption is made about representability of the underlying Jordan algebra. As a special case, we prove polynomial iteration complexities for variants of the short-, semi-long-, and long-step path-following algorithms using the Nesterov-Todd, XS, or SX directions.

1. Introduction

Our purpose in this article is to study the relationship between polynomiality proofs of primal-dual interior point algorithms for cone-LP problems over *symmetric cones*. This paper is companion to our earlier work [SA01] and complements the results presented there. Below we provide some background information and then explain the contrast between the present and earlier work.

A *cone-LP* problem over $\mathcal{K}$, or simply a $\mathcal{K}$-LP problem is defined as the pair of primal-dual optimization problems

$$\begin{array}{llll} \min & \langle \mathbf{c}, \mathbf{x} \rangle & \max & \langle \mathbf{b}, \mathbf{y} \rangle \\ \text{s.t.} & A\mathbf{x} = \mathbf{b} & \text{s.t.} & A^T\mathbf{y} + \mathbf{s} = \mathbf{c} \\ & \mathbf{x} \in \mathcal{K} & & \mathbf{s} \in \mathcal{K}^* \end{array} \tag{1}$$

Here $\mathcal{K} \subset \Re^n$ is a closed, pointed, convex cone with a nonempty interior in $\Re^n$, and $\mathcal{K}^*$ is its dual cone:

$$\mathcal{K}^* = \{\mathbf{z} : \langle \mathbf{x}, \mathbf{z} \rangle \geq 0 \text{ for all } \mathbf{x} \in \mathcal{K}\}.$$

S.H. Schmieta*: Axioma Inc., Marietta, GA USA 30068, e-mail: sschmieta@axiomainc.com

F. Alizadeh**: RUTCOR and School of Business, Rutgers University, Piscataway, NJ USA 08854-8003, e-mail: alizadeh@rutcor.rutgers.edu

* Part of this research was conducted when the first author was a postdoctoral associate at Center for Computational Optimization at Columbia University.

** Research supported in part by the U.S. National Science Foundation grant CCR-9901991 and Office of Naval Research contract number N00014-96-1-0704.

We remind the reader of a few essential properties of cone-LP's. The first one is that all convex programming problems can be formulated in the cone-LP format. Second under the assumption that there are primal feasible points residing in the interior of $\mathcal{K}$ (or dual feasible points $(\mathbf{y}, \mathbf{s})$ with $\mathbf{s}$ in the interior of $\mathcal{K}^*$) the optimal values of the objective functions in the two optimization problems coincide. Third it can be shown that a generalization of the complementary slackness theorem holds, that is the set

$$C = \{(\mathbf{x}, \mathbf{s}) : \mathbf{x} \in \mathcal{K}, \mathbf{s} \in \mathcal{K}^*, \text{ and } \langle \mathbf{x}, \mathbf{s} \rangle = 0\}$$

has dimension at most n ([AS97, Gül97]). Thus there is a set of generically independent equations $f_i(\mathbf{x}, \mathbf{s}) = 0$ for $i = 1, \ldots, n$ that define C. The implication is that primal and dual feasibility and the complementary slackness relations $f_i(\mathbf{x}, \mathbf{s}) = 0$ – in the absence of various degeneracies – determine the optimal solutions. This observation is the starting point of primal-dual algorithms.

In this paper we are interested in interior point primal-dual algorithms in the special case where $\mathcal{K}$ is a *symmetric* cone, that is $\mathcal{K}$ is self-dual and its automorphism group acts transitively on its interior. Thus for each $\mathbf{x}, \mathbf{s} \in \operatorname{Int} \mathcal{K}$, the interior of $\mathcal{K}$, there is a linear transformation $\mathcal{A}$ such that $\mathcal{A}(\mathbf{x}) = \mathbf{s}$ and $\mathcal{A}(\mathcal{K}) = \mathcal{K}$. Symmetric cones are intimately related to Euclidean Jordan algebras [FK94]; these algebras provide us with an essential toolbox for the analysis presented below.

We are in particular concerned with the extension of primal-dual methods from semidefinite programming (SDP) – that is the cone-LP where $\mathcal{K}$ is the cone of positive semidefinite real symmetric matrices – to all symmetric cone-LP's.

The kind of primal-dual methods we will deal with in this paper are direct extensions of algorithms by Monteiro and Adler [MA88], and Kojima, Mizuno, and Yoshise [KMY89] for linear programming. The essence of these methods in linear programming involves applying Newton's method to primal feasibility, $A\mathbf{x} = \mathbf{b}$, dual feasibility, $A^T\mathbf{y} + \mathbf{s} = \mathbf{c}$, and a relaxed form of complementarity conditions, that is $x_i s_i = \mu$. It is important that the complementarity conditions be in this *symmetric* format; thus for instance formulations such as $x_i - \mu/s_i = 0$ or $s_i - \mu/x_i = 0$, while mathematically equivalent to $x_i s_i = \mu$ will result in different directions upon applying Newton's method and will not be considered here. We only mention that for primal-only or dual-only methods, the situation is somewhat simpler in that polynomiality proofs usually extend from linear programming – in a sense word-for-word – to semidefinite programming and in general to optimization over symmetric cones, see [Ali95, NS96, AS00].

The first extension of primal-dual interior point methods to a more general setting than linear programming was achieved by Nesterov and Todd [NT97, NT95]. These authors developed the concept of *self-scaled barriers* and self-scaled cones (those cones that are endowed with a self-scaled barrier) and showed that a particular primal-dual algorithm has an iteration complexity proportional to $\sqrt{r}$, where r is the *self-concordance parameter* of the cone (see [NN94]). It turns out that self-scaled cones are precisely the symmetric cones, and thus the Nesterov-Todd (NT) algorithm was the first primal-dual method for optimization over symmetric cones.

Later several authors derived new algorithms, or classes of algorithms, for semidefinite programming. The main starting point was the way we write the relaxed complementarity conditions for SDP. At first sight the natural relation seems to be $XS = \mu I$

where X and S are symmetric matrices and I is the identity matrix. The difficulty is that applying Newton's method to the system of equations consisting of primal and dual feasibility and $XS = \mu I$ yields a set of directions that are nonsymmetric matrices. Helmberg et al [HRVW96], Monteiro [Mon96] and Kojima et al [KSH97] derived a class of algorithms based on using $XS = \mu I$ (or $SX = \mu I$) known as the XS or SX methods. Essentially they apply Newton's method to $XS = \mu I$ or $SX = \mu I$ and subsequently symmetrize the resulting ΔX. Kojima et al and Monteiro also provided variants of this class of algorithms with polynomial-time iteration complexity. Alizadeh et al [AHO98] presented the $XS + SX$ method which uses $XS + SX = 2\mu I$ as the relaxed complementarity condition. Then Zhang [Zha98] using concepts developed in Monteiro [Mon96] presented a general framework within which one could express all of the algorithms mentioned. Zhang proposed the relaxed complementarity condition to be written in the form $PXSP^{-1} + P^{-1}SXP = 2\mu I$ where P is a nonsingular symmetric positive definite matrix. The class of Newton directions derived in this manner is called the *Monteiro-Zhang family of similarly scaled directions*. It turns out that with appropriate choice of P one can derive the $XS + SX(P = I)$, $XS(P = S^{1/2})$, $SX(P = X^{-1/2})$ and the Nesterov-Todd methods (see [TTT98] and below). Next, in two other works the groundwork was laid for unified polynomial-time complexity proofs for the similarly scaled family. In [MZ98] Monteiro and Zhang proved that the long-step primal-dual algorithm has polynomial-time iteration complexity if the direction belongs to a subclass of the Monteiro-Zhang family known as the *commutative class*. The commutative class includes the Nesterov-Todd, XS, and SX methods, but not the $XS + SX$ method. Next Monteiro in [Mon96] showed that the short-step method can be applied to all of the algorithms in the Monteiro-Zhang family, including noncommutative methods like the $XS + SX$ method.

In a more general context Faybusovich [Fay97b, Fay98] extended the polynomiality proofs of Nesterov and Todd and Kojima et al to optimization problems over symmetric cones. He used the properties of Euclidean Jordan algebras as the basic toolbox to carry out his analysis. Tsuchiya [Tsu97, Tsu98] also used Jordan algebraic techniques to extend the methods of [MZ98] to optimization over the Lorentz cone, a special symmetric cone. Such optimization problems include convex quadratically constrained quadratic programming problems (QCQP). He obtained polynomial-time iteration complexity for short-step, semi-long-step, and long-step algorithms, but restricted his analysis to the NT and XS methods. In addition he posed the question whether his analysis can be generalized to all symmetric cones. We answer this question in the affirmative. Finally, we showed in [SA01] that Monteiro's short-step analysis applies to all symmetric cone optimization problems whose underlying Euclidean Jordan algebra is derived from an associative algebra. Simultaneously, Monteiro and Tsuchiya in [MT98] showed that the analog of the $XS + SX$ method can be extended to Lorentz cone optimization problems.

As was mentioned, the present paper is a companion of our earlier aforementioned work [SA01]. There is an intimate relationship between symmetric cones and Euclidean Jordan algebras, and properties of these algebras serve as a fundamental toolbox for complexity analysis of optimization over symmetric cones. In [SA01] we extended the notion of similarly scaled primal-dual algorithms to all symmetric cone optimization problems. We then showed that the short-step method applied to *any* direction in this class results in polynomial-time iteration complexity, but we could only extend the analysis to those symmetric cones whose Jordan algebras were *representable*, that is they were derived from some associative algebra. This excluded the exceptional *Albert algebra* and its associated 27-dimensional symmetric cone (see the remainder of the paper for definitions). In the present paper we show that the short-, semi-long- and long-step methods for the *commutative class* of directions, a special subclass of the similarly scaled class of directions, have polynomial-time iteration complexity. Furthermore our analysis is entirely based on Jordan algebraic tools and does not require the existence of any associative algebra inducing our Jordan algebra. As a result, our analysis applies to *all* symmetric cones. Thus, in contrast to our earlier work in [SA01], we have a more restrictive class of algorithms, but can apply it to a wider class of optimization problems.

In §2 we provide background information on Euclidean Jordan algebras. In §3 we take the analyses by Tsuchiya and Monteiro-Zhang [Tsu98, MZ98] for the commutative class of directions and prove that they extend verbatim to all symmetric cone optimization problems.

2. Euclidean Jordan algebras

In this section we outline a minimal foundation of the theory of Euclidean Jordan algebras. This theory serves as our basic toolbox for the analysis of primal-dual interior point methods. Our presentation mostly follows Faraut and Korányi [FK94].

Let $\mathcal{J}$ be an n-dimensional vector space over the field of real numbers with a multiplication "$\circ$" where the map $(\mathbf{x}, \mathbf{y}) \to \mathbf{x} \circ \mathbf{y}$ is bilinear. Then $(\mathcal{J}, \circ)$ is a Jordan algebra if for all $\mathbf{x}, \mathbf{y} \in \mathcal{J}$

1. $\mathbf{x} \circ \mathbf{y} = \mathbf{y} \circ \mathbf{x}$,
2. $\mathbf{x} \circ (\mathbf{x}^2 \circ \mathbf{y}) = \mathbf{x}^2 \circ (\mathbf{x} \circ \mathbf{y})$ where $\mathbf{x}^2 = \mathbf{x} \circ \mathbf{x}$.

A Jordan algebra $\mathcal{J}$ is called *Euclidean* if there exists a symmetric, positive definite quadratic form Q on $\mathcal{J}$ which is also associative that is

$$Q(\mathbf{x} \circ \mathbf{y}, \mathbf{z}) = Q(\mathbf{x}, \mathbf{y} \circ \mathbf{z}).$$

To make our presentation more concrete, throughout this section we examine the following two examples of Euclidean Jordan algebras.

Example 1 (The Jordan algebras of matrices $\mathcal{M}_n^+$ and $\mathcal{S}_n^+$). The set $\mathcal{M}_n$ of $n \times n$ real matrices with the multiplication $X \circ Y \stackrel{\text{def}}{=} (XY + YX)/2$ forms a Jordan algebra which will be denoted by $\mathcal{M}_n^+$. It is not a Euclidean Jordan algebra, though. The subspace $\mathcal{S}_n$ of real symmetric matrices also forms a Jordan algebra under the "$\circ$" operation; in fact it is a Jordan subalgebra of M_n^+. $(\mathcal{S}_n, \circ)$ is Euclidean since if we define $Q(X, Y) = \operatorname{Trace}(X \circ Y) = \operatorname{Trace}(XY)$, then clearly Trace is positive definite, since $\operatorname{Trace}(X \circ X) > 0$ for $X \neq 0$. Its associativity is easy to prove by using the fact that $\operatorname{Trace}(XY) = \operatorname{Trace}(YX)$. We write $\mathcal{S}_n^+$ for this algebra. □

Example 2 (The quadratic forms algebra $\mathcal{E}_{n+1}^+$). Let $\mathcal{E}_{n+1}$ be the $(n + 1)$-dimensional real vector space whose elements $\mathbf{x}$ are indexed from zero. Define the product

$$\mathbf{x} \circ \mathbf{y} \stackrel{\text{def}}{=} \begin{pmatrix} x^T y \\ x_0 y_1 + x_1 y_0 \\ \vdots \\ x_0 y_n + x_n y_0 \end{pmatrix}.$$

Then it is easily verified that $(\mathcal{E}_{n+1}, \circ)$ is a Jordan algebra. Furthermore, $Q(\mathbf{x}, \mathbf{y}) \stackrel{\text{def}}{=} \mathbf{x}^T \mathbf{y}$ is both associative and positive definite. Thus, $(\mathcal{E}_{n+1}, \circ)$ is Euclidean. We will use the notation $\mathcal{E}_{n+1}^+$ for this algebra. □

Definition 1. *If $\mathcal{J}$ is a Euclidean Jordan Algebra then its cone of squares is the set*

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

Recall that symmetric cones are closed, pointed, convex cones that are self-dual and their automorphism group acts transitively on their interior. The relevance of the theory of Euclidean Jordan algebras for $\mathcal{K}$-LP optimization problem stems from the following theorem, which can be found in [FK94].

Theorem 2 (Jordan algebraic characterization of symmetric cones). *A cone is symmetric iff it is the cone of squares of some Euclidean Jordan algebra.*

Example 3 (Cone of Squares of $\mathcal{S}_n^+$). A symmetric matrix is square of of another symmetric matrix iff it is positive semidefinite. Thus the cone of squares of $\mathcal{S}_n^+$ is the cone of positive semidefinite matrices. We write $X \succcurlyeq 0$ if X is positive semidefinite. □

Example 4 (Cone of squares of $\mathcal{E}_{n+1}^+$). It is straightforward to show that the cone of squares of $\mathcal{E}_{n+1}^+$ is $\mathcal{Q} \stackrel{\text{def}}{=} \{x \in \Re^{n+1} : x_0 \geq \|\bar{x}\|\}$ where $\bar{x} = (x_1, \ldots, x_n)^T$, and $\|\cdot\|$ indicates the Euclidean norm, see for example [AS00, AG02]. $\mathcal{Q}$ is called the Lorentz cone or the second order cone. □

A Jordan algebra $\mathcal{J}$ has an identity element, if there exists a (necessarily unique) element $\mathbf{e}$ such that $\mathbf{x} \circ \mathbf{e} = \mathbf{e} \circ \mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in \mathcal{J}$. Jordan algebras are not necessarily associative, but they are power associative, i.e. the algebra generated by a single element $\mathbf{x} \in \mathcal{J}$ is associative.

In the subsequent development we deal exclusively with Euclidean Jordan algebras with identity. Many of the results hold true for the more general case of Jordan algebras, but this generality is not needed here.

Since "$\circ$" is bilinear for every $\mathbf{x} \in \mathcal{J}$, there exists a matrix $L(\mathbf{x})$ such that for every $\mathbf{y}$, $\mathbf{x} \circ \mathbf{y} = L(\mathbf{x})\mathbf{y}$. In particular, $L(\mathbf{x})\mathbf{e} = \mathbf{x}$ and $L(\mathbf{x})\mathbf{x} = \mathbf{x}^2$. For each $\mathbf{x}, \mathbf{y} \in \mathcal{J}$ define

$$\mathbf{Q}_{\mathbf{x},\mathbf{y}} \stackrel{\text{def}}{=} L(\mathbf{x})L(\mathbf{y}) + L(\mathbf{y})L(\mathbf{x}) - L(\mathbf{x} \circ \mathbf{y}) \qquad \mathbf{Q}_{\mathbf{x}} \stackrel{\text{def}}{=} 2L^2(\mathbf{x}) - L(\mathbf{x}^2).$$

$\mathbf{Q}_{\mathbf{x}}$ is *the quadratic representation* of $\mathbf{x}$. Clearly $\mathbf{Q}_{\mathbf{x},\mathbf{z}}\mathbf{y}$ and $\mathbf{Q}_{\mathbf{x}}\mathbf{y}$ are in $\mathcal{J}$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{J}$. The quadratic representation is an essential concept in the theory of Jordan algebras and will play an important role in our subsequent development. Indeed it is possible to start from a proper axiomatization of the quadratic representation and define "$\circ$" from it, see for example [Jac68].

Example 5 (Identity, and L and $\mathbf{Q}_x$ operators for $\mathcal{S}_n^+$). Clearly the identity element for $\mathcal{S}_n^+$ is the usual identity I for square matrices. Applying the vec operator to a $n \times n$ matrix to turn it into a vector, we get,

$$\operatorname{vec}\big(X \circ Y\big) = \operatorname{vec}\left(\frac{XY + YX}{2}\right) = \frac{1}{2}\Big(I \otimes X + X \otimes I\Big)\operatorname{vec}(Y).$$

Thus, for the $\mathcal{S}_n^+$ algebra, $L(X) = \frac{1}{2}\big(X \otimes I + I \otimes X\big)$, which is also known as Kronecker sum of X and X, see, for example [HJ90] for properties of this operator. A further calculation shows that $\mathbf{Q}_{X,Z}\operatorname{vec}(Y) = \frac{1}{2}\operatorname{vec}\big(XYZ + ZYX\big)$ and thus, as an operator, $\mathbf{Q}_{X,Z} = \frac{1}{2}\big(X \otimes Z + Z \otimes X\big)$, in particular, $\mathbf{Q}_X = X \otimes X$. Thus the quadratic representation in Jordan algebras are generalization of the operator that sends Y to XYX in symmetric matrices. As the latter is used extensively in both analysis and development of interior point methods for semidefinite programming, the former will similarly play a prominent role in extension of such methods to cone-LP's over symmetric cones. □

Example 6 (Identity, and L and $\mathbf{Q}_\mathbf{x}$ operators for $\mathcal{E}_{n+1}^+$). The identity element is the vector $e = (1, 0, \ldots, 0)^T$. From the definition of Jordan multiplication for $\mathcal{E}_{n+1}^+$ it is seen that

$$L(x) = \operatorname{Arw}(\mathbf{x}) \stackrel{\text{def}}{=} \begin{pmatrix} x_0 & \bar{x}^T \\ \bar{x} & x_0 I \end{pmatrix}$$

This matrix is an arrow-shaped matrix which is related to the Lorentz transformation. From here it is easily verified that

$$Q_\mathbf{x} \stackrel{\text{def}}{=} 2\operatorname{Arw}^2(\mathbf{x}) - \operatorname{Arw}(\mathbf{x}^2) = \begin{pmatrix} \|\mathbf{x}\|^2 & 2x_0\bar{\mathbf{x}}^\top \\ 2x_0\bar{\mathbf{x}} & \gamma(\mathbf{x})I + 2\bar{\mathbf{x}}\bar{\mathbf{x}}^\top \end{pmatrix} = \big(2\mathbf{x}\mathbf{x}^\top - \gamma(\mathbf{x})R\big) \qquad (2)$$

where $\gamma(x) = x_0^2 - \|\bar{x}\|^2$ and R is the diagonal reflection matrix with $R_{00} = 1$ and $R_{ii} = -1$ for $i = 1, \ldots, n$. □

Since a Jordan algebra $\mathcal{J}$ is power associative we can define the concepts of rank, the minimum and the characteristic polynomials, eigenvalues, trace, and determinant for it in the following way.

For each $\mathbf{x} \in \mathcal{J}$ let r be the smallest integer such that the set $\{e, \mathbf{x}, \mathbf{x}^2, \ldots, \mathbf{x}^r\}$ is linearly dependent. Then r is the *degree* of $\mathbf{x}$ which we denote as $\deg(\mathbf{x})$. The *rank* of $\mathcal{J}$, $\operatorname{rk}(\mathcal{J})$, is the largest $\deg(\mathbf{x})$ of any member $\mathbf{x} \in \mathcal{J}$. An element $\mathbf{x}$ is called *regular* if its degree equals the rank of the Jordan algebra.

For an element $\mathbf{x}$ of degree d in a rank-r algebra $\mathcal{J}$, since $\{\mathbf{e}, \mathbf{x}, \mathbf{x}^2, \ldots, \mathbf{x}^d\}$ is linearly dependent, there are real numbers $a_1(\mathbf{x}), \ldots, a_d(\mathbf{x})$ such that

$$\mathbf{x}^d - a_1(\mathbf{x})\mathbf{x}^{d-1} + \cdots + (-1)^d a_d(\mathbf{x})\mathbf{e} = \mathbf{0}. \quad (\mathbf{0} \text{ is the zero vector})$$

The polynomial $\lambda^d - a_1(\mathbf{x})\lambda^{d-1} + \cdots + (-1)^d a_d(\mathbf{x})$ is the *minimum polynomial* of $\mathbf{x}$.

Now, as shown in Faraut and Koranyi, [FK94], each coefficient $a_i(\mathbf{x})$ of the minimum polynomial is a homogeneous polynomial of degree i, thus in particular it is a continuous function of $\mathbf{x}$. We can now define the notion of characteristic polynomials as follows: If $\mathbf{x}$ is a regular element of the algebra, then we define its characteristic polynomial to be equal to its minimum polynomial. Next, since the set of regular elements are dense in $\mathcal{J}$ (see [FK94]) we can continuously extend characteristic polynomials to all elements $\mathbf{x}$ of $\mathcal{J}$. Therefore, the characteristic polynomial is a degree r polynomial in λ.

Definition 3. *The roots,* $\lambda_1, \cdots, \lambda_r$ *of the characteristic polynomial of* $\mathbf{x}$ *are the* eigenvalues *of* $\mathbf{x}$.

It is possible, in fact certain in the case of nonregular elements, that the characteristic polynomial have multiple roots. However, the minimum polynomial has only simple roots. Indeed, the characteristic and minimum polynomials have the same set of roots, except for their multiplicities. Thus the minimum polynomial of $\mathbf{x}$ always divides its characteristic polynomial.

Definition 4. *Let* $\mathbf{x} \in \mathcal{J}$ *and* $\lambda_1, \ldots, \lambda_r$ *be the roots of its characteristic polynomial* $p(\lambda) = \lambda^r - p_1(\mathbf{x})\lambda^{r-1} + \cdots + (-1)^r p_r(\mathbf{x})$. *Then*

1. $\operatorname{tr}(\mathbf{x}) \stackrel{\text{def}}{=} \lambda_1 + \cdots + \lambda_r = p_1(\mathbf{x})$ *is the trace of* $\mathbf{x}$ *in* $\mathcal{J}$;
2. $\det(\mathbf{x}) \stackrel{\text{def}}{=} \lambda_1 \cdots \lambda_r = p_r(\mathbf{x})$ *is the determinant of* $\mathbf{x}$ *in* $\mathcal{J}$.

Note that trace is a linear function of $\mathbf{x}$.

Example 7 (Characteristic polynomials, eigenvalues, trace and determinant in $\mathcal{S}_n^+$*).* These notions coincide with the familiar ones in symmetric matrices. Note that $\deg(X)$ is the number of distinct eigenvalues of X and thus is at most n for an $n \times n$ symmetric matrix, in other words the $\operatorname{rank}(\mathcal{S}_n^+) = n$. □

Example 8 (Characteristic polynomials, eigenvalues, trace and determinant in $\mathcal{E}_{n+1}^+$*).* Every vector $\mathbf{x} \in \mathcal{E}_{n+1}^+$ satisfies the quadratic equation

$$\mathbf{x}^2 - 2x_0\mathbf{x} + (x_0^2 - \|\overline{\mathbf{x}}\|^2)\mathbf{e} = \mathbf{0} \tag{3}$$

Thus $\operatorname{rank}(\mathcal{E}_{n+1}^+) = 2$ independent of the dimension of its underlying vector space. Furthermore, each element $\mathbf{x}$ has two eigenvalues, $x_0 \pm \|\overline{\mathbf{x}}\|$; $\operatorname{tr}(\mathbf{x}) = 2x_0$ and $\det(\mathbf{x}) = x_0^2 - \|\overline{\mathbf{x}}\|^2$. Except for multiples of identity, every element has degree 2. □

Proposition 5 ([FK94]). *If* $\mathbf{x}, \mathbf{y} \in \mathcal{J}$ *then*

$$\det(\mathbf{Q_y x}) = \det(\mathbf{y})^2 \det(\mathbf{x}) = \det(\mathbf{y}^2) \det(\mathbf{x}). \tag{4}$$

Together with the eigenvalues comes a decomposition of $\mathbf{x}$ into idempotents, its *spectral decomposition*. Recall that an *idempotent* $\mathbf{c}$ is a nonzero element of $\mathcal{J}$ where $\mathbf{c}^2 = \mathbf{c}$.

1. A *complete system of orthogonal idempotents* is a set $\{\mathbf{c}_1, \ldots, \mathbf{c}_k\}$ of idempotents where $\mathbf{c}_i \circ \mathbf{c}_j = \mathbf{0}$ for all $i \neq j$, and $\mathbf{c}_1 + \cdots + \mathbf{c}_k = \mathbf{e}$.
2. An idempotent is *primitive* if it is not sum of two other idempotents.
3. A complete system of orthogonal primitive idempotents is called a *Jordan frame*.

Note that in Jordan frames $k = r$, that is Jordan frames always have r primitive idempotents in them.

Theorem 6 (Spectral decomposition, type I). *Let $\mathcal{J}$ be a Euclidean Jordan algebra. Then for $\mathbf{x} \in \mathcal{J}$ there exist unique real numbers $\lambda_1, \dots, \lambda_k$, all distinct, and a unique complete system of orthogonal idempotents $\mathbf{c}_1, \dots, \mathbf{c}_k$ such that*

$$\mathbf{x} = \lambda_1 \mathbf{c}_1 + \cdots + \lambda_k \mathbf{c}_k. \tag{5}$$

See [FK94].

Theorem 7 (Spectral decomposition, type II). *Let $\mathcal{J}$ be a Euclidean Jordan algebra with rank r. Then for $\mathbf{x} \in \mathcal{J}$ there exists a Jordan frame $\mathbf{c}_1, \dots, \mathbf{c}_r$ and real numbers $\lambda_1, \dots, \lambda_r$ such that*

$$\mathbf{x} = \lambda_1 \mathbf{c}_1 + \cdots + \lambda_r \mathbf{c}_r \tag{6}$$

and the λ_i are the eigenvalues of $\mathbf{x}$.

A direct consequence of these facts is that eigenvalues of elements of Euclidean Jordan algebras are always real; this is not the case for arbitrary power associative algebras or even non-Euclidean Jordan algebras.

Example 9 (Spectral decomposition in $\mathcal{S}_n^+$). Every symmetric matrix can be diagonalized by an orthogonal matrix: $X = Q \Lambda Q^T$. This relation may be written as $X = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T$, where the λ_i are the eigenvalues of X and $\mathbf{q}_i$, columns of Q, their corresponding eigenvectors. Since the $\mathbf{q}_i$ form an orthonormal set, it follows that the set of rank one matrices $\mathbf{q}_i \mathbf{q}_i^T$ form a Jordan frame: $\left(\mathbf{q}_i \mathbf{q}_i^T\right)^2 = \mathbf{q}_i \mathbf{q}_i^T$ and $\left(\mathbf{q}_i \mathbf{q}_i^T\right)\left(\mathbf{q}_j \mathbf{q}_j^T\right) = 0$ for $i \neq j$; finally $\sum_i \mathbf{q}_i \mathbf{q}_i^T = I$. This gives the type II spectral decomposition of X. For type I, let $\lambda_1 > \cdots > \lambda_k$ be distinct eigenvalues of X, where each λ_i has multiplicity m_i. Suppose that $\mathbf{q}_{i_1}, \dots, \mathbf{q}_{i_{m_i}}$ are a set of orthogonal eigenvectors of λ_i. Define $P_i = \mathbf{q}_{i_1} \mathbf{q}_{i_1}^T + \cdots + \mathbf{q}_{i_{m_i}} \mathbf{q}_{i_{m_i}}^T$. Then the P_i form an orthogonal system of idempotents, which add up to I. Also, note that even though for a given eigenvalue λ_i, the corresponding eigenvectors $\mathbf{q}_{i_r}$ may not be unique, the P_i *are* unique for each λ_i. Thus, the identity $X = \lambda_1 P_1 + \cdots + \lambda_k P_k$ is the type I spectral decomposition of X. □

Example 10 (Spectral decomposition in $\mathcal{E}_{n+1}^+$). Consider the following identity

$$\mathbf{x} = \frac{1}{2}(x_0 + \|\overline{\mathbf{x}}\|) \begin{pmatrix} 1 \\ \frac{\overline{\mathbf{x}}}{\|\overline{\mathbf{x}}\|} \end{pmatrix} + \frac{1}{2}(x_0 - \|\overline{\mathbf{x}}\|) \begin{pmatrix} 1 \\ -\frac{\overline{\mathbf{x}}}{\|\overline{\mathbf{x}}\|} \end{pmatrix} \tag{7}$$

We have already mentioned that $\lambda_{1,2} = x_0 \pm \|\overline{\mathbf{x}}\|$. Let us define $\mathbf{c}_1 = \frac{1}{2}\left(1, \frac{\overline{\mathbf{x}}}{\|\overline{\mathbf{x}}\|}\right)^T$, and $\mathbf{c}_2 = \frac{1}{2}\left(1, -\frac{\overline{\mathbf{x}}}{\|\overline{\mathbf{x}}\|}\right)^T$, and observe that $\mathbf{c}_2 = R\mathbf{c}_1$, with R as defined earlier. Also, $\mathbf{c}_i^2 = \mathbf{c}_i$ for $i = 1, 2$, and $\mathbf{c}_1 \circ \mathbf{c}_2 = \mathbf{0}$. Thus, (7) is the type II spectral decomposition of $\mathbf{x}$ which can be alternatively written as $\mathbf{x} = \lambda_1 \mathbf{c}_1 + \lambda_2 \mathbf{c}_2$. Since only multiples of identity $\alpha\mathbf{e}$ have multiple eigenvalues, their type I spectral decomposition is simply $\alpha\mathbf{e}$, with $\mathbf{e}$ the singleton system of orthonormal idempotents. □

Now it is possible to extend the definition of any real valued continuous function $f(\cdot)$ to elements of Jordan algebras using their eigenvalues:

$$f(\mathbf{x}) \stackrel{\text{def}}{=} f(\lambda_1)\mathbf{c}_1 + \cdots + f(\lambda_k)\mathbf{c}_k.$$

We are in particular interested in the following functions:

1. The square root: $\mathbf{x}^{1/2} \stackrel{\text{def}}{=} \lambda_1^{1/2}\mathbf{c}_1 + \cdots + \lambda_k^{1/2}\mathbf{c}_k$, whenever all $\lambda_i \geq 0$, and undefined otherwise.
2. The inverse: $\mathbf{x}^{-1} \stackrel{\text{def}}{=} \lambda_1^{-1}\mathbf{c}_1 + \cdots + \lambda_k^{-1}\mathbf{c}_k$ whenever all $\lambda_i \neq 0$ and undefined otherwise.

Note that $\left(\mathbf{x}^{1/2}\right)^2 = \mathbf{x}$, and $\mathbf{x}^{-1} \circ \mathbf{x} = \mathbf{e}$. If $\mathbf{x}^{-1}$ is defined, we call $\mathbf{x}$ *invertible* We call $\mathbf{x} \in \mathcal{J}$ *positive semidefinite* if all its eigenvalues are nonnegative, and *positive definite* if all its eigenvalues are positive. We write $\mathbf{x} \succcurlyeq 0$ (respectively $\mathbf{x} \succ 0$) if $\mathbf{x}$ is positive semidefinite (respectively positive definite.) It is clear that an element is positive semidefinite if, and only if belongs to the cone of squares; it is positive definite if, and only if it belongs to the interior of the cone of squares.

We may also define various norms on $\mathcal{J}$ as functions of eigenvalues much the same way that unitarily invariant norms are defined on square matrices:

$$\|\mathbf{x}\|_{\mathrm{F}} \stackrel{\text{def}}{=} \left(\sum \lambda_i^2\right)^{1/2} = \sqrt{\mathrm{tr}(\mathbf{x}^2)}, \qquad \|\mathbf{x}\|_2 = \max_i |\lambda_i| \tag{8}$$

Observe that $\|e\|_{\mathrm{F}} = \sqrt{r}$. Finally since "$\circ$" is bilinear and $\mathrm{tr}(\mathbf{x} \circ \mathbf{y})$ is a symmetric positive definite quadratic form which is associative, $\mathrm{tr}\big(\mathbf{x} \circ (\mathbf{y} \circ \mathbf{z})\big) = \mathrm{tr}\big((\mathbf{x} \circ \mathbf{y}) \circ \mathbf{z}\big)$, we may define the inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle \stackrel{\text{def}}{=} \mathrm{tr}(\mathbf{x} \circ \mathbf{y})$$

Example 11 (Inverse, square root, and norms in $\mathcal{S}_n^+$). Again, here these notions coincide with the familiar ones. $\|X\|_F = \left(\sum_{ij} X_{ij}^2\right)^{1/2} = \left(\sum_i \lambda_i^2\right)^{1/2}$ is the Frobenius norm of X and $\|X\|_2 = \max_i |\lambda_i(X)|$ is the familiar spectral norm. The inner product $\mathrm{Trace}(X \circ Y) = \mathrm{Trace}(XY)$, is denoted by $X \bullet Y$. □

Example 12 (Inverse, square root, and norms in $\mathcal{E}_{n+1}^+$). Let $\mathbf{x} = \lambda_1\mathbf{c}_1 + \lambda_2\mathbf{c}_2$ with $\{\mathbf{c}_1, \mathbf{c}_2\}$ its Jordan frame. Then,

$$\begin{aligned}
\mathbf{x}^{-1} &= \tfrac{1}{\lambda_1}\mathbf{c}_1 + \tfrac{1}{\lambda_2}\mathbf{c}_2 = \tfrac{R\mathbf{x}}{\det(x)} \qquad \text{when } \det \mathbf{x} \neq 0,\\
\mathbf{x}^{1/2} &= \sqrt{\lambda_1}\mathbf{c}_1 + \sqrt{\lambda_2}\mathbf{c}_2,\\
\|\mathbf{x}\|_F^2 &= \lambda_1^2 + \lambda_2^2 = 2x_0^2 + 2\|\overline{\mathbf{x}}\|^2 = 2\|\mathbf{x}\|^2,\\
\|\mathbf{x}\|_2 &= \max\{\lambda_1, \lambda_2\} = |x_0| + \|\overline{\mathbf{x}}\|,\\
\langle \mathbf{x}, \mathbf{y} \rangle &= \mathrm{tr}(x \circ y) = 2\mathbf{x}^T\mathbf{y}.
\end{aligned}$$

□

Note that since the inner product $\langle \mathbf{x}, \mathbf{y}\rangle$ is associative, it follows that $L(\mathbf{x})$ is symmetric with respect to $\langle\cdot,\cdot\rangle$, because $\langle L(\mathbf{x})\mathbf{y}, \mathbf{z}\rangle = \langle \mathbf{x}\circ\mathbf{y}, \mathbf{z}\rangle = \langle \mathbf{y}, \mathbf{x}\circ\mathbf{z}\rangle = \langle \mathbf{y}, L(x)\mathbf{z}\rangle$. From definition of $\mathbf{Q_x}$, it follows that it too is symmetric with respect to $\langle\cdot,\cdot\rangle$.

We are now ready to state some further fundamental properties of $\mathbf{Q_x}$ operator.

Lemma 8 (Properties of Q.). *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be elements of a Jordan algebra and* $\mathbf{x}$ *invertible. Then*

1. $\mathbf{Q_{x,y}} = \frac{1}{2}\left(\mathbf{Q_{x+y}} - \mathbf{Q_x} - \mathbf{Q_y}\right)$.
2. $\mathbf{Q_x x^{-1}} = \mathbf{x}$, $\mathbf{Q_x^{-1}} = \mathbf{Q_{x^{-1}}}$, $\mathbf{Q_x e} = \mathbf{x}^2$.
3. $\mathbf{Q_{x,x^{-1}} Q_x} = \mathbf{Q_x Q_{x,x^{-1}}} = 2L(\mathbf{x})\mathbf{Q_{e,x}} - \mathbf{Q_x} = L(\mathbf{x}^2)$.
4. $\mathbf{Q_{Q_y x}} = \mathbf{Q_y Q_x Q_y}$.
5. $\mathbf{Q_{x^k}} = \left(\mathbf{Q_x}\right)^k$.
6. $\mathbf{Q_x} L(\mathbf{x}^{-1}) = L(\mathbf{x})$.

All the statements are taken from [FK94] and [Jac68], except item 5, which can be proved by observing that

$$\mathbf{Q_{x^{2k}}} = \mathbf{Q_{Q_{x^k} e}} = \mathbf{Q_{x^k} Q_e Q_{x^k}} = \left(\mathbf{Q_{x^k}}\right)^2$$

where the first equality is by item 2, the second one by item 4 and the last one by noting that $\mathbf{Q_e} = I$, the identity matrix. Similarly,

$$\mathbf{Q_{x^{2k+1}}} = \mathbf{Q_{Q_{x^k} x}} = \mathbf{Q_{x^k} Q_x Q_{x^k}}$$

Now a simple induction proves the claim for all integer k.

2.1. Peirce decomposition

An important concept in the theory of Jordan algebras is the *Peirce decomposition*. Notice that for an idempotent $\mathbf{c}$, since $\mathbf{c}^2 = \mathbf{c}$, one can show that $2L^3(\mathbf{c}) - 3L^2(\mathbf{c}) + L(\mathbf{c}) = \mathbf{0}$, see [FK94], Proposition III.1.3. Therefore, eigenvalues of $L(\mathbf{c})$ are 0, $\frac{1}{2}$ and 1. Furthermore, the eigenspace corresponding to each eigenvalue of $L(\mathbf{c})$ is the set of $\mathbf{x}$ such that $L(c)\mathbf{x} = i\mathbf{x}$ or equivalently $\mathbf{c}\circ\mathbf{x} = i\mathbf{x}$, for $i = 0, \frac{1}{2}, 1$. Therefore, [Jac68].

Theorem 9 (Peirce decomposition, type I). *Let* $\mathcal{J}$ *be a Jordan algebra and* $\mathbf{c}$ *an idempotent. Then* $\mathcal{J}$*, as a vector space, can be decomposed as*

$$\mathcal{J} = \mathcal{J}_1(\mathbf{c}) \oplus \mathcal{J}_0(\mathbf{c}) \oplus \mathcal{J}_{\frac{1}{2}}(\mathbf{c}) \tag{9}$$

where

$$\mathcal{J}_i(\mathbf{c}) = \{\mathbf{x} \mid \mathbf{x}\circ\mathbf{c} = i\mathbf{x}\}\,. \tag{10}$$

The three eigenspaces $\mathcal{J}_i(\mathbf{c})$, are called *Peirce spaces* with respect to $\mathbf{c}$.

Now let $\{\mathbf{c}_1, \ldots, \mathbf{c}_k\}$ be an orthonormal system of idempotents. Each $\mathbf{c}_i$ has it own three Peirce spaces $\mathcal{J}_0(\mathbf{c}_i)$, $\mathcal{J}_{\frac{1}{2}}(\mathbf{c}_i)$, and $\mathcal{J}_1(\mathbf{c}_i)$. It can be shown that $L(\mathbf{c}_i)$ all commute and thus share a common system of eigenvectors, [FK94] Lemma IV.1.3. In fact, the common eigenspaces are of two types ([FK94] Theorem IV.2.1):

i. $\mathcal{J}_{ii} \stackrel{\text{def}}{=} \mathcal{J}_1(\mathbf{c}_i) = \mathcal{J}_0(\mathbf{c}_j)$ for all $j \neq i$, and
ii. $\mathcal{J}_{ij} \stackrel{\text{def}}{=} \mathcal{J}_{\frac{1}{2}}(\mathbf{c}_i) \cap \mathcal{J}_{\frac{1}{2}}(\mathbf{c}_j)$.

Thus, with respect to an orthonormal system of idempotents $\{\mathbf{c}_1, \ldots, \mathbf{c}_k\}$, one can give a finer decomposition:

Theorem 10 (Peirce decomposition, type II ([FK94] Theorem IV.2.1)). *Let $\mathcal{J}$ be a Jordan algebra with identity and $\mathbf{c}_i$ a system of orthogonal idempotents such that $\mathbf{e} = \sum_i \mathbf{c}_i$. Then we have the Peirce decomposition $\mathcal{J} = \bigoplus_{i \leq j} \mathcal{J}_{ij}$ where*

$$\mathcal{J}_{ii} = \mathcal{J}_1(\mathbf{c}_i) = \{\mathbf{x} \mid \mathbf{x} \circ \mathbf{c}_i = \mathbf{x}\} \tag{11}$$

$$\mathcal{J}_{ij} = \mathcal{J}_{\frac{1}{2}}(\mathbf{c}_i) \cap \mathcal{J}_{\frac{1}{2}}(\mathbf{c}_j) = \left\{\mathbf{x} \mid \mathbf{x} \circ \mathbf{c}_i = \frac{1}{2}\mathbf{x} = \mathbf{x} \circ \mathbf{c}_j\right\}. \tag{12}$$

The Peirce spaces $\mathcal{J}_{ij}$ are orthogonal with respect to any symmetric bilinear form.

Lemma 11 (Properties of Peirce spaces). *Let $\mathcal{J}$ be a Jordan algebra with $\mathbf{e} = \sum_i \mathbf{c}_i$ where the $\mathbf{c}_i$ are orthogonal idempotents and let $\mathcal{J}_{ij}$ be the Peirce decomposition relative to the $\mathbf{c}_i$. Then*

1. $\mathcal{J}_{ii} \circ \mathcal{J}_{ii} \subseteq \mathcal{J}_{ii}$, $\mathcal{J}_{ii} \circ \mathcal{J}_{ij} \subseteq \mathcal{J}_{ij}$, $\mathcal{J}_{ij} \circ \mathcal{J}_{ij} \subseteq \mathcal{J}_{ii} + \mathcal{J}_{jj}$
2. $\mathcal{J}_{ii} \circ \mathcal{J}_{jj} = \{\mathbf{0}\}$, *if* $i \neq j$
3. $\mathcal{J}_{ij} \circ \mathcal{J}_{jk} \subseteq \mathcal{J}_{ik}$, $\mathcal{J}_{ij} \circ \mathcal{J}_{kk} = \{\mathbf{0}\}$
4. $\mathcal{J}_{ij} \circ \mathcal{J}_{kl} = \{\mathbf{0}\}$ *if* $\{i, j\} \cap \{k, l\} = \emptyset$
5. $\mathcal{J}_0(\mathbf{c}_i) = \bigoplus_{j,k \neq i} \mathcal{J}_{jk}$.

In a Euclidean Jordan algebra, the Peirce decomposition is closely related to the orthogonal decomposition of the vector space with respect to $L(\mathbf{x})$. If $\mathbf{x} = \sum_{i=1}^{k} \lambda_i \mathbf{c}_i$ then the Peirce spaces $\mathcal{J}_{ij}$ corresponding to the system of orthogonal idempotents $\{\mathbf{c}_1, \ldots \mathbf{c}_k\}$ are eigenspaces of $L(\mathbf{x})$. An immediate consequence of this observation is the following:

Lemma 12. *Let $\mathbf{x} \in \mathcal{J}$ with spectral decomposition type I: $\mathbf{x} = \lambda_1 \mathbf{c}_1 + \cdots + \lambda_k \mathbf{c}_k$. Then the following statements hold.*

i. *The matrices, $L(\mathbf{x})$ and $\mathbf{Q}_{\mathbf{x}}$ commute and thus share a common system of eigenvectors; in fact the $\mathbf{c}_i$ are among their common eigenvectors.*
ii. *The eigenvalues of $L(\mathbf{x})$ have the form*

$$\frac{\lambda_i + \lambda_j}{2} \qquad 1 \leq i \leq j \leq k,$$

in particular, all λ_i are eigenvalues of $L(\mathbf{x})$, and $\mathbf{x}$ is positive (semi-definite) definite iff $L(\mathbf{x})$ is positive (semi-definite) definite, moreover, for $\mathbf{x} \succcurlyeq \mathbf{0}$, $L(\mathbf{x})$ is invertible if $\mathbf{x}$ is positive definite.
iii. *The eigenvalues of $\mathbf{Q}_{\mathbf{x}}$ have the form*

$$\lambda_i \lambda_j \qquad 1 \leq i \leq j \leq k,$$

Lemma 13. *Let* $\mathbf{x} \in \mathcal{J}$*, then we obtain the smallest and the largest eigenvalue as*

$$\lambda_{\min}(\mathbf{x}) = \min_{\mathbf{u}} \frac{\langle \mathbf{u}, \mathbf{x} \circ \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}, \qquad \lambda_{\max}(\mathbf{x}) = \max_{\mathbf{u}} \frac{\langle \mathbf{u}, \mathbf{x} \circ \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

Proof. Replace $\mathbf{x} \circ \mathbf{u}$ by $L(\mathbf{x})\mathbf{u}$ and note that $\lambda_{\min}(L(\mathbf{x})) = \lambda_{\min}(\mathbf{x})$ and $\lambda_{\max}(L(\mathbf{x})) = \lambda_{\max}(\mathbf{x})$. □

Lemma 14. *Let* $\mathbf{x}, \mathbf{y} \in \mathcal{J}$*, then we can bound the eigenvalues of* $\mathbf{x} + \mathbf{y}$ *as follows*

$$\lambda_{\min}(\mathbf{x} + \mathbf{y}) \geq \lambda_{\min}(\mathbf{x}) - \|\mathbf{y}\|_{\mathrm{F}} \tag{13}$$

$$\lambda_{\max}(\mathbf{x} + \mathbf{y}) \leq \lambda_{\max}(\mathbf{x}) - \|\mathbf{y}\|_{\mathrm{F}}\,. \tag{14}$$

Proof. For (13) we have

$$\begin{aligned}
\lambda_{\min}(\mathbf{x} + \mathbf{y}) &= \min_{\mathbf{u}} \frac{\langle \mathbf{u}, (\mathbf{x} + \mathbf{y}) \circ \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \\
&= \min_{\mathbf{u}} \frac{\langle \mathbf{u}, \mathbf{x} \circ \mathbf{u} \rangle + \langle \mathbf{u}, \mathbf{y} \circ \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \\
&\geq \min_{\mathbf{u}} \frac{\langle \mathbf{u}, \mathbf{x} \circ \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} + \min_{\mathbf{u}} \frac{\langle \mathbf{u}, \mathbf{y} \circ \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \\
&= \lambda_{\min}(\mathbf{x}) + \lambda_{\min}(\mathbf{y}) \\
&\geq \lambda_{\min} - \|\mathbf{y}\|_{\mathrm{F}}\,.
\end{aligned}$$

The proof for (14) is similar. □

Example 13 (Peirce decomposition in $\mathcal{S}_n^+$*).* Let $E = \{\mathbf{q}_1\mathbf{q}^T, \ldots, \mathbf{q}_n\mathbf{q}_n^T\}$ be a Jordan frame, where the $\mathbf{q}_i$ are orthonormal set of vectors in $\Re^n$. To see the Peirce spaces associated with the idempotent $C = \mathbf{q}_1\mathbf{q}_1^T + \cdots + \mathbf{q}_m\mathbf{q}_m^T$, first observe that C has eigenvalues 1 (with multiplicity m) and 0 (with multiplicity $n - m$). Next recall that $L(C) = \frac{1}{2}\big(C \otimes I + I \otimes C\big)$. In general, if A and B are square matrices with eigenvalues λ_i and ω_j, respectively, and with corresponding eigenvectors $\mathbf{u}_i$ and $\mathbf{v}_j$ then $A \otimes I + I \otimes B$ has eigenvalues $\lambda_i + \omega_j$, and $A \otimes B$ has eigenvalues $\lambda_i\omega_j$, and corresponding eigenvectors $\mathbf{u}_i \otimes \mathbf{v}_j = \mathrm{vec}(\mathbf{v}_j\mathbf{u}_i^T)$, for $i, j = 1, \ldots, n$; in particular, $A \otimes B$ and $A \otimes I + I \otimes B$ commute. Thus, the eigenvalues of $L(C)$ are $1, \frac{1}{2}, 0$. Therefore, the Peirce space $\mathcal{J}_i(C)$ consists of those matrices $A \in \mathcal{S}_n$ where $iA = A \circ (\sum_{i=1}^m \mathbf{q}_i\mathbf{q}_i^T) = \frac{1}{2}\sum_{i=1}^m (A\mathbf{q}_i\mathbf{q}_i^T + \mathbf{q}_i\mathbf{q}_i^T A)$. It follows that,

$$\begin{aligned}
\mathcal{J}_1(C) &= \big\{A \in \mathcal{S}_n \mid \mathbf{q}_i^T A\mathbf{q}_j = 0 \quad \text{if } m+1 \leq i \leq n \quad \text{or} \quad m+1 \leq j \leq n\big\}, \\
\mathcal{J}_0(C) &= \big\{A \in \mathcal{S}_n \mid \mathbf{q}_i^T A\mathbf{q}_j = 0 \quad \text{if } 1 \leq i \leq m \quad \text{or} \quad 1 \leq i \leq m\big\}, \\
\mathcal{J}_{\frac{1}{2}}(C) &= \big\{A \in \mathcal{S}_n \mid \mathbf{q}_i^T A\mathbf{q}_j = 0 \quad \text{if } 1 \leq i, j \leq m \quad \text{or} \quad m+1 \leq i, j \leq n\big\}.
\end{aligned}$$

□

Example 14 (Peirce decomposition in $\mathcal{E}_{n+1}^+$*).* Let $\mathbf{x} = \lambda_1\mathbf{c}_1 + \lambda_2\mathbf{c}_2$ be the spectral decomposition of $\mathbf{x}$ with $\lambda_1 \neq \lambda_2$. First, it can be verified by inspection that the matrix

$\operatorname{Arw}(\mathbf{x})$ has eigenvalues $\lambda_{1,2} = x_0 \pm \|\bar{\mathbf{x}}\|$, each with multiplicity 1 and corresponding eigenvectors $\mathbf{c}_{1,2}$; and $\lambda_3 = x_0$ with multiplicity $n-1$. Also, $\mathbf{Q_x}$ has eigenvalues $\left(x_0 \pm \|\bar{\mathbf{x}}\|\right)^2$ each with multiplicity one and corresponding eigenvectors $\mathbf{c}_{1,2}$. The remaining eigenvalue of $\mathbf{Q_x}$ is $\lambda_1\lambda_2 = \det \mathbf{x}$ with multiplicity $n-1$. An idempotent $\mathbf{c}$ which is not equal to identity element $\mathbf{e}$ is of the form $\mathbf{c} = \frac{1}{2}(1, \mathbf{q})$ where $\mathbf{q}$ is a unit length vector. Thus $\operatorname{Arw}(\mathbf{c})$ has one eigenvalue equal to 1, another equal to 0 and $n-1$ equal to $\frac{1}{2}$. It is easy to verify that

$$\begin{aligned} \mathcal{J}_1(\mathbf{c}) &= \{\alpha \mathbf{c} \mid \alpha \in \Re\}, \\ \mathcal{J}_0(\mathbf{c}) &= \{\alpha R\mathbf{c} \mid \alpha \in \Re\}, \\ J_{\frac{1}{2}}(\mathbf{c}) &= \{(0, \mathbf{p})^T \mid \mathbf{p}^T\mathbf{q} = 0\}. \end{aligned}$$

□

For the following statements we need the notion of a simple algebra.

Definition 15. *An (Euclidean Jordan) algebra $\mathcal{A}$ is called* simple *iff is not the direct sum of two (Euclidean Jordan) subalgebras.*

Proposition 16 ([FK94]). *If $\mathcal{J}$ is a Euclidean Jordan algebra, then it is, in a unique way, a direct sum of simple Euclidean Jordan algebras.*

Lemma 17 ([FK94]). *Let $\mathcal{J}$ be an n-dimensional simple Euclidean Jordan algebra and $(\mathbf{a}, \mathbf{b})$, $(\mathbf{a}_1, \mathbf{b}_1)$ two pairs of orthogonal primitive idempotents. Then*

$$\dim\left(\mathcal{J}_{\frac{1}{2}}(\mathbf{a}) \cap \mathcal{J}_{\frac{1}{2}}(\mathbf{b})\right) = \dim\left(\mathcal{J}_{\frac{1}{2}}(\mathbf{a}_1) \cap \mathcal{J}_{\frac{1}{2}}(\mathbf{b}_1)\right) = d, \tag{15}$$

and, if $\operatorname{rk}(\mathcal{J}) = r$,

$$n = r + \frac{d}{2} r(r-1).$$

Example 15 (The parameter d for $\mathcal{S}_n^+$ and $\mathcal{E}_{n+1}^+$). For $\mathcal{S}_n^+$, $d = 1$. For $\mathcal{E}_{n+1}^+$, $d = n-1$. □

We also will need the fact that the quadratic representation of a positive definite element of $\mathcal{J}$ is an automorphism of the cone of squares of $\mathcal{J}$.

Proposition 18 (Proposition III.2.2 in [FK94]). *Let $\mathbf{x}, \mathbf{y} \in \mathcal{J}$ be two positive definite elements of a Euclidean Jordan algebra. Then $\mathbf{Q_x y}$ is positive definite.*

Proposition 19. *Two elements $\mathbf{x}$ and $\mathbf{y}$ of a simple Euclidean Jordan algebra have the same spectrum if, and only if $L(\mathbf{x})$ and $L(\mathbf{y})$ have the same spectrum.*

Proof. As the spectrum of $L(\mathbf{x})$ is determined by the spectrum of $\mathbf{x}$, the "only if" part is true.

Now assume $\mathbf{x}$ and $\mathbf{y}$ have different spectra. Denote by $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_r$ the eigenvalues of $\mathbf{x}$ and by μ_i the eigenvalues of $\mathbf{y}$ ordered the same way. Let k be the largest index such that $\lambda_k \ne \mu_k$. W.l.o.g. we can assume $\lambda_k > \mu_k$. If $k = r$, then

λ_k is larger than all eigenvalues of $L(\mathbf{y})$ but it is an eigenvalue of $L(\mathbf{x})$ so the spectra differ. In the other case we have $k < r$ and $\lambda_i = \mu_i$ for $i > k$. So the multiplicities for eigenvalues larger than λ_k are the same in the spectra of $\mathbf{x}$ and $\mathbf{y}$. But the multiplicity of λ_k in the spectrum of $\mathbf{x}$ is strictly larger than its multiplicity in the spectrum of $\mathbf{y}$. Hence the multiplicity of $(\lambda_k + \lambda_r)/2$ is larger in the spectrum of $L(\mathbf{x})$ than in that of $L(\mathbf{y})$. That completes the proof. □

Corollary 20. *Two positive definite elements of a simple Euclidean Jordan algebra have the same spectrum if, and only if their quadratic representations $\mathbf{Q_x}$ and $\mathbf{Q_y}$ have the same spectrum.*

Proof. Since the eigenvalues of $\mathbf{Q_x}$ are λ_i^2 and $\lambda_i\lambda_j$ the "only if" direction is true.

Since all the $\lambda_i > 0$ we can apply the same reasoning as in the proof of the previous proposition. □

Proposition 21. *Let $\mathcal{J}$ be the direct sum of simple Euclidean Jordan algebras $\mathcal{J}_i$. Let $\mathbf{x}, \mathbf{y}, \mathbf{p} \in \mathcal{J}$ be positive definite elements. Define $\widetilde{\mathbf{x}} = \mathbf{Q}_p\mathbf{x}$ and $\underset{\sim}{\mathbf{s}} = \mathbf{Q}_{p^{-1}}\mathbf{s}$. Then*

i. *$\mathbf{Q}_{\mathbf{x}^{1/2}}\mathbf{s}$ and $\mathbf{Q}_{\mathbf{s}^{1/2}}\mathbf{x}$ have the same spectrum.*
ii. *$\mathbf{a} = \mathbf{Q}_{\mathbf{x}^{1/2}}\mathbf{s}$ and $\mathbf{b} = \mathbf{Q}_{\widetilde{\mathbf{x}}^{1/2}}\underset{\sim}{\mathbf{s}}$ have the same spectrum.*

Proof. We can, in a unique way, write $\mathbf{x}$, $\mathbf{s}$, and $\mathbf{p}$ in terms of elements in the simple algebras $\mathcal{J}_i$.

$$\mathbf{x} = \bigoplus_i \mathbf{x}_i, \quad \mathbf{s} = \bigoplus_i \mathbf{s}_i, \quad \mathbf{p} = \bigoplus_i \mathbf{p}_i, \qquad \mathbf{x}_i, \mathbf{s}_i, \mathbf{p}_i \in \mathcal{J}_i.$$

By orthogonality of the $\mathcal{J}_i$ we have

$$\mathbf{Q_p x} = \bigoplus_i \mathbf{Q}_{\mathbf{p}_i}\mathbf{x}_i \qquad \mathbf{Q}_{\mathbf{p}^{-1}}\mathbf{s} = \bigoplus_i \mathbf{Q}_{\mathbf{p}_i^{-1}}\mathbf{s}_i,$$

so it suffices to show the claim for simple Euclidean algebras. Now,

i. By corollary 20 and proposition 18, it suffices to show that $\mathbf{Q}_{\mathbf{Q}_{\mathbf{x}^{1/2}}\mathbf{s}}$ and $\mathbf{Q}_{\mathbf{Q}_{\mathbf{s}^{1/2}}\mathbf{x}}$ are similar. But

$$\mathbf{Q}_{\mathbf{Q}_{\mathbf{x}^{1/2}}\mathbf{s}} = \mathbf{Q}_{\mathbf{x}^{1/2}}\mathbf{Q_s}\mathbf{Q}_{\mathbf{x}^{1/2}}$$
$$\mathbf{Q}_{\mathbf{Q}_{\mathbf{s}^{1/2}}\mathbf{x}} = \mathbf{Q}_{\mathbf{s}^{1/2}}\mathbf{Q_x}\mathbf{Q}_{\mathbf{s}^{1/2}}$$

Since in general the matrices $AB^2A = (AB)(BA)$ and $BA^2B = (BA)(AB)$ have the same spectrum, inserting $A = \mathbf{Q}_{\mathbf{x}^{1/2}}$ and $B = \mathbf{Q}_{\mathbf{s}^{1/2}}$ proves the assertion.

ii. Again, $\mathbf{a}$ and $\mathbf{b}$ have the same eigenvalues iff their quadratic representations are similar matrices. Now

$$\mathbf{Q_a} = \mathbf{Q}_{\mathbf{Q}_{\mathbf{x}^{1/2}}\mathbf{s}} = \mathbf{Q}_{\mathbf{x}^{1/2}}\mathbf{Q_s}\mathbf{Q}_{\mathbf{x}^{1/2}}$$

and

$$\mathbf{Q_b} = \mathbf{Q}_{\mathbf{Q}_{\widetilde{\mathbf{x}}^{1/2}}\underset{\sim}{\mathbf{s}}} = \mathbf{Q}_{\widetilde{\mathbf{x}}^{1/2}}\mathbf{Q}_{\underset{\sim}{\mathbf{s}}}\mathbf{Q}_{\widetilde{\mathbf{x}}^{1/2}}.$$

Using the property of the quadratic representation that $\mathbf{Q}_{\mathbf{x}}^k = \mathbf{Q}_{\mathbf{x}^k}$, $\mathbf{Q}_{\mathbf{a}}$ is similar to $\mathbf{Q}_{\mathbf{x}}\mathbf{Q}_{\mathbf{s}}$ and $\mathbf{Q}_{\mathbf{b}}$ is similar to $\mathbf{Q}_{\widetilde{\mathbf{x}}}\mathbf{Q}_{\underset{\sim}{\mathbf{s}}}$. From the definition of $\widetilde{\mathbf{x}}$ and $\underset{\sim}{\mathbf{s}}$ we obtain

$$\mathbf{Q}_{\widetilde{\mathbf{x}}}\mathbf{Q}_{\underset{\sim}{\mathbf{s}}} = \mathbf{Q}_{\mathbf{p}}\mathbf{Q}_{\mathbf{x}}\mathbf{Q}_{\mathbf{p}}\mathbf{Q}_{\mathbf{p}^{-1}}\mathbf{Q}_{\mathbf{s}}\mathbf{Q}_{p^{-1}} = \mathbf{Q}_{\mathbf{p}}\mathbf{Q}_{\mathbf{x}}\mathbf{Q}_{\mathbf{s}}\mathbf{Q}_{\mathbf{p}}^{-1}.$$

So $\mathbf{Q}_{\mathbf{a}}$ and $\mathbf{Q}_{\mathbf{b}}$ are similar, which proves the claim. □

Finally we state the classification theorem of Euclidean Jordan algebras:

Theorem 22 ([FK94] Chapter V.). *Let $\mathcal{J}$ be a simple Euclidean Jordan algebra. Then $\mathcal{J}$ is isomorphic to one of the following algebras*

1. *the algebra $\mathcal{E}_{n+1}^+$,*
2. *the algebra $\mathcal{S}_n^+$ of $n \times n$ symmetric matrices,*
3. *the algebra $(\mathcal{H}_n, \circ)$ of $n \times n$ complex Hermitian matrices under the operation $X \circ Y = \frac{1}{2}\big(XY + YX\big)$,*
4. *the algebra $(\mathcal{Q}_n, \circ)$ of $n \times n$ quaternion Hermitian matrices under the operation, $X \circ Y = \frac{1}{2}\big(XY + YX\big)$,*
5. *the exceptional Albert algebra, that is the algebra $(\mathcal{O}_3, \circ)$ of 3×3 octonion Hermitian matrices under the operation, $X \circ Y = \frac{1}{2}\big(XY + YX\big)$.*

Since octonion multiplication is not associative, the 27-dimensional Albert algebra is not induced by an associative operation, as the other four are. That is why it is called *exceptional*.

2.2. *Operator commutativity*

We say two elements $\mathbf{x}, \mathbf{y}$ of a Jordan algebra $\mathcal{J}$ *operator commute* if $L(\mathbf{x})L(\mathbf{y}) = L(\mathbf{y})L(\mathbf{x})$. In other words, $\mathbf{x}$ and $\mathbf{y}$ operator commute if for all $\mathbf{z}$, $\mathbf{x} \circ (\mathbf{y} \circ \mathbf{z}) = \mathbf{y} \circ (\mathbf{x} \circ \mathbf{z})$. If $A \subseteq \mathcal{J}$ we denote the set of elements in $\mathcal{J}$ that operator commute with all $\mathbf{a} \in A$ by $C_{\mathcal{J}}(A)$.

Lemma 23 ([Jac68]). *If $\mathbf{c}$ is an idempotent in $\mathcal{J}$ then $C_{\mathcal{J}}(\{\mathbf{c}\}) = \mathcal{J}_0(\mathbf{c}) \oplus \mathcal{J}_1(\mathbf{c})$. So $C_{\mathcal{J}}(\mathbf{c})$ is a subalgebra.*

Theorem 24 (Operator commutativity [Jac68]). *Let $\mathcal{J}$ be an arbitrary finite-dimensional Jordan algebra, B a subalgebra. Then $C_{\mathcal{J}}(B)$ is a subalgebra and*

$$C_{\mathcal{J}}(B) = \bigcap_{i=1}^{k} C_{\mathcal{J}}(\{\mathbf{c}_i\}) = \bigcap_{i=1}^{k} \left(\mathcal{J}_0(\mathbf{c}_i) \oplus \mathcal{J}_1(\mathbf{c}_i)\right), \tag{16}$$

where $\{\mathbf{c}_1, \ldots, \mathbf{c}_k\}$ are idempotents that form a basis of B.

Lemma 25. *If $\{\mathbf{c}_1, \ldots, \mathbf{c}_k\}$ is a complete system of orthogonal idempotents in $\mathcal{J}$ and B is the subalgebra generated by them, then*

$$C_{\mathcal{J}}(B) = \bigoplus_{i=1}^{k} \mathcal{J}_1(\mathbf{c}_i). \tag{17}$$

Proof. Using the properties of Peirce spaces and their orthogonality, we obtain

$$\begin{aligned}
C_{\mathcal{J}}(B) &= \bigcap_{i=1}^{k} (\mathcal{J}_0(\mathbf{c}_i) \oplus \mathcal{J}_1(\mathbf{c}_i)) = \bigcap_{i=1}^{k} \left(\mathcal{J}_1(\mathbf{c}_i) \oplus \bigoplus_{s,t \neq i} \mathcal{J}_{st} \right) \\
&= \bigcap_{i=1}^{k} \left(\left(\bigoplus_{s=1}^{k} \mathcal{J}_1(\mathbf{c}_s) \right) \oplus \left(\bigoplus_{s \neq t \neq i} \mathcal{J}_{st} \right) \right) \\
&= \bigoplus_{i=1}^{k} \mathcal{J}_1(\mathbf{c}_i)
\end{aligned}$$

□

Lemma 26. *Let $\mathcal{J}$ be a Euclidean Jordan algebra and $\mathbf{x} \in \mathcal{J}$. Let also that $\mathbf{x}$ have a type I spectral decomposition with idempotents $\mathbf{c}_1, \ldots, \mathbf{c}_k$. Denote by B the subalgebra of $\mathcal{J}$ generated by the $\mathbf{c}_i$. Then $C_{\mathcal{J}}(\{\mathbf{x}\}) = C_{\mathcal{J}}(B)$.*

Proof. Obviously $C_{\mathcal{J}}(B) \subseteq C_{\mathcal{J}}(\{\mathbf{x}\})$. To prove the other direction assume that $\mathbf{y}$ operator commutes with $\mathbf{x}$. So $L(\mathbf{x})$ and $L(\mathbf{y})$ commute. Then if $p()$ is an arbitrary polynomial, $p(L(\mathbf{x}))$ commutes with $L(\mathbf{y})$. Since the λ_i are distinct, there are polynomials $p_i()$ such that $p_i(\lambda_i) = 1$ and $p_i(\lambda_j) = 0$ if $j \neq i$. Hence $L(\mathbf{y})$ commutes with all the $L(\mathbf{c}_i)$ and $\mathbf{y} \in C_{\mathcal{J}}(B)$. □

Theorem 27. *Let $\mathbf{x}$ and $\mathbf{y}$ be two elements of a Euclidean Jordan algebra $\mathcal{J}$. Then $\mathbf{x}$ and $\mathbf{y}$ operator commute if, and only if there is a Jordan frame $\mathbf{c}_1, \ldots, \mathbf{c}_r$ such that $\mathbf{x} = \sum_{i=1}^{r} \lambda_i \mathbf{c}_i$ and $\mathbf{y} = \sum_{i=1}^{r} \mu_i \mathbf{c}_i$.*

Proof. Let B be the subalgebra generated by the Jordan frame. Then $\mathbf{x} \in B$ and $\mathbf{y} \in C_{\mathcal{J}}(B)$ so they operator commute. That proves the "if" direction.

So now assume that $\mathbf{x}$ and $\mathbf{y}$ operator commute. Denote by $\mathbf{f}_1, \ldots, \mathbf{f}_k$ the complete systems of idempotent from the type I spectral decomposition of $\mathbf{x}$. Then

$$C_{\mathcal{J}}(\{\mathbf{x}\}) = \bigoplus_{i=1}^{k} \mathcal{J}_1(\mathbf{f}_i).$$

Since $\mathbf{y} \in C_{\mathcal{J}}(\{\mathbf{x}\})$, there is a unique decomposition of

$$\mathbf{y} = \mathbf{g}_1 + \cdots + \mathbf{g}_k, \qquad \mathbf{g}_i \in \mathcal{J}_1(\mathbf{f}_i).$$

The $\mathcal{J}_1(\mathbf{f}_i)$ are Euclidean subalgebras with $\mathbf{f}_i$ as identity. So for each $\mathbf{g}_i$ there exists a type II spectral decomposition in $\mathcal{J}_1(\mathbf{f}_i)$. Denote by $\mathbf{c}_{i1}, \ldots, \mathbf{c}_{ik_i}$ the Jordan frame of $\mathbf{g}_i$ and by $\mu_{i1}, \ldots, \mu_{ik_i}$ the corresponding eigenvalues. Then we can write $\mathbf{y}$ as

$$\mathbf{y} = \sum_{i=1}^{k} \sum_{j=1}^{k_i} \mu_{ij} \mathbf{c}_{ij}.$$

Obviously the $\mathbf{c}_{ij}$ are mutually orthogonal and sum up to $\mathbf{e}$. Hence they form a Jordan frame. $\mathbf{x}$ can be written using this Jordan frame in the following way

$$\mathbf{x} = \sum_{i=1}^{k} \lambda_i f_i = \sum_{i=1}^{k} \sum_{j=1}^{k_i} \lambda_i \mathbf{c}_{ij}.$$

□

Example 16 (Operator commutativity in $\mathcal{S}_n^+$). The operator commutativity is a generalization of the notion of commutativity of the associative matrix product. If $X, Y \in \mathcal{S}_n$ commute, they share a common system of orthonormal eigenvectors, which means they have a common Jordan frame. From the eigenstructure of Kronecker sums and Kronecker products it is clear that X and Y commute iff $\mathbf{Q}_X = X \otimes X$ and $\mathbf{Q}_Y = Y \otimes Y$ commute, and equivalently that $X \otimes I + I \otimes X$ and $Y \otimes I + I \otimes Y$ commute. □

Example 17 (Operator commutativity in $\mathcal{E}_{n+1}^+$). Here from the eigenstructure of $\mathrm{Arw}\,(\cdot)$ described in Example 14 it is easily verified that if $\mathrm{Arw}\,(\mathbf{x})$ and $\mathrm{Arw}\,(\mathbf{y})$ commute then there is a Jordan frame $\{\mathbf{c}_1, \mathbf{c}_2\}$ such that $\mathbf{x} = \lambda_1 \mathbf{c}_1 + \lambda_2 \mathbf{c}_2$, and $\mathbf{y} = \omega_1 \mathbf{c}_1 + \omega_2 \mathbf{c}_2$. □

3. Newton's method and commutative directions

Let $\mathcal{J}$ be a Euclidean Jordan algebra with dimension n, rank r, and cone of squares $\mathcal{K}$. Consider the primal-dual pair of $\mathcal{K}$-LP problems

$$(P) \qquad \min\{\langle \mathbf{c}, \mathbf{x}\rangle \mid \langle \mathbf{a}_i, \mathbf{x}\rangle = \mathbf{b}_i, i = 1, \ldots, m, \mathbf{x} \in \mathcal{K}, \} \tag{18}$$

and

$$(D) \qquad \max\{\mathbf{b}^T \mathbf{y} \mid \sum_{i=1}^{m} \mathbf{y}_i \mathbf{a}_i + \mathbf{s} = \mathbf{c}, \mathbf{s} \in \mathcal{K}, \mathbf{y} \in \Re^m\}, \tag{19}$$

where $\mathbf{c}, \mathbf{a}_i \in \mathcal{J}$, for $i = 1, \ldots, m$ and $\mathbf{b} \in \Re^m$. We call $\mathbf{x} \in \mathcal{K}$ primal feasible if $\langle \mathbf{a}_i, \mathbf{x}\rangle = \mathbf{b}_i$ for $i = 1, \ldots, m$. Similarly, $(\mathbf{s}, \mathbf{y}) \in \mathcal{K} \times \Re^m$ is called dual feasible if $\sum_{i=1}^m \mathbf{y}_i \mathbf{a}_i + \mathbf{s} = \mathbf{c}$. Let $A \in \Re^{m \times n}$ be the matrix corresponding to the linear transformation that maps $\mathbf{x}$ to the m-vector whose i^{th} component is $\langle \mathbf{a}_i, \mathbf{x}\rangle$. Throughout we assume that $\operatorname{rank}(A) = m$ and that both primal and dual problems contain positive definite feasible points.

The sets of interior feasible solutions are

$$\mathcal{F}^0(P) \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathcal{J} \mid A\mathbf{x} = \mathbf{b}, \mathbf{x} \in \operatorname{Int} \mathcal{K}\},$$
$$\mathcal{F}^0(D) \stackrel{\text{def}}{=} \{(\mathbf{s}, \mathbf{y}) \in \mathcal{J} \times \Re^m \mid A^T \mathbf{y} + \mathbf{s} = \mathbf{c}, \mathbf{s} \in \operatorname{Int} \mathcal{K}\}.$$

The complementary slackness theorem for this pair of $\mathcal{K}$-LP's is given by $\mathbf{x} \circ \mathbf{s} = \mathbf{0}$ and thus the system

$$\begin{array}{l} A\mathbf{x} = \mathbf{b} \\ A^T \mathbf{y} + \mathbf{s} = \mathbf{c} \\ \mathbf{x} \circ \mathbf{s} = \mathbf{0} \\ \mathbf{x}, \mathbf{s} \in \mathcal{K}, \quad \mathbf{y} \in \Re^m \end{array} \tag{20}$$

completely determines the primal and dual optimal solutions in the absence of various degeneracies [AS97, Fay97a]. The purpose of this section is to show that we can solve a relaxation of this system by Newton's method and obtain polynomial convergence for various primal-dual interior-point methods based on it.

If we drop the condition that $\mathbf{x} \in \mathcal{K}$ in the primal problem (P) and add the barrier $-\sigma\mu \ln\det \mathbf{x}$ to the primal objective, and then derive the optimality conditions of the resulting optimization problem, we obtain a system with primal feasibility, dual feasibility, and $\mathbf{x} \circ \mathbf{s} = \sigma\mu\mathbf{e}$ which is a relaxation of the complementarity conditions $\mathbf{x} \circ \mathbf{s} = \mathbf{0}$. Similarly, if the condition $\mathbf{s} \in \mathcal{K}$ is dropped from the dual problem (D) and the term $\sigma\mu \ln\det \mathbf{s}$ is added to the dual objective, deriving the optimality condition results in the same relaxed system, see [AS97]. Applying Newton's method to this system leads us to the linear system

$$\begin{aligned} &A\Delta\mathbf{x} = \mathbf{b} - A\mathbf{x} \\ &A^T\Delta\mathbf{y} + \Delta\mathbf{s} = \mathbf{c} - \mathbf{s} - A^T\mathbf{y} \\ &\Delta\mathbf{x} \circ \mathbf{s} + \mathbf{x} \circ \Delta\mathbf{s} = \sigma\mu\mathbf{e} - \mathbf{x} \circ \mathbf{s}, \end{aligned} \tag{21}$$

where $(\Delta\mathbf{x}, \Delta\mathbf{s}, \Delta\mathbf{y}) \in \mathcal{J} \times \mathcal{J} \times \Re^m$, $\sigma \in [0, 1]$ is a centering parameter and $\mu = \langle \mathbf{x}, \mathbf{s}\rangle/r$ (remember $r = \mathrm{rk}(\mathcal{J}) = \langle \mathbf{e}, \mathbf{e}\rangle$) is the normalized duality gap. The direction obtained this way is known as the $\mathbf{xs} + \mathbf{sx}$ direction [AHO98]. We now show an equivalent way of writing the complementarity conditions:

Lemma 28. *Let* $\mathbf{x}$, $\mathbf{s}$, *and* $\mathbf{p}$ *be in some Euclidean Jordan algebra* $\mathcal{J}$, $\mathbf{x}, \mathbf{s} \succ \mathbf{0}$, *and* $\mathbf{p}$ *invertible. Then* $\mathbf{x} \circ \mathbf{s} = \alpha\mathbf{e}$ *iff* $\mathbf{Q_p x} \circ \mathbf{Q_{p^{-1}} s} = \alpha\mathbf{e}$.

Proof. We know that $\mathbf{Q_p} L(\mathbf{p}^{-1}) = L(\mathbf{p})$ (see lemma 8) and that $\mathbf{Q_p}$ and $L(\mathbf{p}^{-1})$ commute. Therefore,

$$\mathbf{p}^{-1} \circ \mathbf{Q_p x} = L(\mathbf{p}^{-1})\mathbf{Q_p x} = L(\mathbf{p})\mathbf{x} = \mathbf{p} \circ \mathbf{x}.$$

Regarding the left- and right-hand sides of this equation as functions of $\mathbf{p}$ and taking directional derivatives of both sides in the $\mathbf{u}$ direction, we obtain

$$(-\mathbf{Q_{p^{-1}} u}) \circ \mathbf{Q_p x} + 2\mathbf{p}^{-1} \circ \mathbf{Q_{p,u} x} = \mathbf{u} \circ \mathbf{x}.$$

Setting $\mathbf{u} = \mathbf{s}$ and noting that $L(\mathbf{x})$ and $L(\mathbf{s})$ commute because $\mathbf{s} = \mu\mathbf{x}^{-1}$, it follows that

$$\begin{aligned} &(-\mathbf{Q_{p^{-1}} s}) \circ \mathbf{Q_p x} + 2\mathbf{p}^{-1} \circ \mathbf{Q_{p,s} x} = \mathbf{s} \circ \mathbf{x} \\ \iff &(-\mathbf{Q_{p^{-1}} s}) \circ \mathbf{Q_p x} + 2L(\mathbf{p}^{-1})(L(\mathbf{p})L(\mathbf{s})\mathbf{x} + L(\mathbf{s})L(\mathbf{x})\mathbf{p} - L(\mathbf{x})L(\mathbf{s})\mathbf{p})) = \mathbf{s} \circ \mathbf{x} \\ \iff &(-\mathbf{Q_{p^{-1}} s}) \circ \mathbf{Q_p x} + 2\mu L(\mathbf{p}^{-1})\mathbf{p} = \mathbf{s} \circ \mathbf{x} \\ \iff &\mathbf{Q_{p^{-1}} s} \circ \mathbf{Q_p x} - 2\mu\mathbf{e} = -\mu\mathbf{e} \\ \iff &\mathbf{Q_{p^{-1}} s} \circ \mathbf{Q_p x} = \mu\mathbf{e} \qquad \square \end{aligned}$$

Therefore, we may write the relaxed complementarity condition $\mathbf{x} \circ \mathbf{s} = \sigma\mu\mathbf{e}$ as $(\mathbf{Q_p x}) \circ (\mathbf{Q_{p^{-1}} s}) = \sigma\mu\mathbf{e}$, where $\mathbf{p}$ is invertible. Then the Newton system becomes

$$\begin{aligned} A\Delta\mathbf{x} &= \mathbf{b} - Ax \\ A^{\mathrm{T}}\Delta\mathbf{y} + \Delta\mathbf{s} &= \mathbf{c} - \mathbf{s} - A^{\mathrm{T}}y \\ (\mathbf{Q_p}\Delta\mathbf{x}) \circ (\mathbf{Q_{p^{-1}} s}) + (\mathbf{Q_p x}) \circ (\mathbf{Q_{p^{-1}}}\Delta\mathbf{s}) &= \sigma\mu\mathbf{e} - (\mathbf{Q_p x}) \circ (\mathbf{Q_{p^{-1}} s}), \end{aligned} \tag{22}$$

Denote by $\mathfrak{C}(\mathbf{x}, \mathbf{s})$ the set of all elements so that the scaled elements operator commute, i.e.

$$\mathfrak{C}(\mathbf{x}, \mathbf{s}) = \left\{\mathbf{p} \mid \mathbf{p} \text{ nonsingular, } \mathbf{Q_p x} \text{ and } \mathbf{Q_{p^{-1}} s} \text{ operator commute}\right\}.$$

This is a subclass of the Monteiro-Zhang family of search directions called the commutative class.

For now we only assume that $\mathbf{p}$ is invertible. In the following we denote by $\mathbf{w} = \mathbf{Q_{x^{1/2}} s}$, $\widetilde{\mathbf{x}} = \mathbf{Q_p x}$, $\underset{\sim}{\mathbf{s}} = \mathbf{Q_{p^{-1}} s}$, and $\widetilde{\mathbf{w}} = \mathbf{Q}_{\widetilde{\mathbf{x}}^{1/2}} \underset{\sim}{\mathbf{s}}$. With this notation, the Newton system becomes

$$\begin{aligned} \underset{\sim}{A} \widetilde{\Delta \mathbf{x}} &= \mathbf{b} - \underset{\sim}{A} \widetilde{\mathbf{x}} \\ \underset{\sim}{A}^{\mathrm{T}} \Delta \mathbf{y} + \underset{\sim}{\Delta}\, \mathbf{s} &= \underset{\sim}{\mathbf{c}} - \underset{\sim}{\mathbf{s}} - \underset{\sim}{A}^{\mathrm{T}} \mathbf{y} \\ \widetilde{\Delta \mathbf{x}} \circ \underset{\sim}{\mathbf{s}} + \widetilde{\mathbf{x}} \circ \underset{\sim}{\Delta}\, \mathbf{s} &= \sigma \mu \mathbf{e} - \widetilde{\mathbf{x}} \circ \underset{\sim}{\mathbf{s}}, \end{aligned} \tag{23}$$

where $\underset{\sim}{A} = A\mathbf{Q_{p^{-1}}}$ and $\underset{\sim}{\mathbf{c}} = \mathbf{Q_{p^{-1}} c}$. The system (23) is equivalent to (22), because, $\mathbf{Q_p}(\mathcal{K}) = \mathcal{K}$ for symmetric cones $\mathcal{K}$.

We will study short-, semi-long-, and long-step algorithms associated with the following centrality measures defined for $(\mathbf{x}, \mathbf{s}) \in \operatorname{Int} \mathcal{K} \times \operatorname{Int} \mathcal{K}$:

$$d_F(\mathbf{x}, \mathbf{s}) \overset{\text{def}}{=} \left\|\mathbf{Q_{x^{1/2}} s} - \mu \mathbf{e}\right\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{r} (\lambda_i(\mathbf{w}) - \mu)^2} \tag{24}$$

$$\begin{aligned} d_2(\mathbf{x}, \mathbf{s}) &\overset{\text{def}}{=} \left\|\mathbf{Q_{x^{1/2}} s} - \mu \mathbf{e}\right\|_2 = \max_{i=1,\dots,r} |\lambda_i(\mathbf{w}) - \mu| \\ &= \max\{\lambda_{\max}(\mathbf{w}) - \mu, \mu - \lambda_{\min}(\mathbf{w})\} \end{aligned} \tag{25}$$

$$d_{-\infty}(\mathbf{x}, \mathbf{s}) \overset{\text{def}}{=} \mu - \lambda_{\min}(\mathbf{w}) \tag{26}$$

Given a constant $\gamma \in (0, 1)$, we define the following neighborhoods of the central path

$$\mathcal{N}_F(\gamma) \overset{\text{def}}{=} \{(\mathbf{x}, \mathbf{s}, \mathbf{y}) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D) \mid d_F(\mathbf{x}, \mathbf{s}) \le \gamma\mu\} \tag{27}$$

$$\mathcal{N}_2(\gamma) \overset{\text{def}}{=} \{(\mathbf{x}, \mathbf{s}, \mathbf{y}) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D) \mid d_2(\mathbf{x}, \mathbf{s}) \le \gamma\mu\} \tag{28}$$

$$\mathcal{N}_{-\infty}(\gamma) \overset{\text{def}}{=} \{(\mathbf{x}, \mathbf{s}, \mathbf{y}) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D) \mid d_{-\infty}(\mathbf{x}, \mathbf{s}) \le \gamma\mu\} \tag{29}$$

Note that by part i. of Proposition 21 $\mathbf{Q_{x^{1/2}} s}$ and $\mathbf{Q_{s^{1/2}} x}$ operator commute, and thus the centrality measures $d_\bullet(\mathbf{x}, \mathbf{s})$ and their corresponding neighborhoods $\mathcal{N}_\bullet(\gamma)$ are symmetric with respect to $\mathbf{x}$ and $\mathbf{s}$.

Proposition 29. *The three neighborhoods defined in* (27)–(29) *are scaling invariant, i.e.* $(\mathbf{x}, \mathbf{s})$ *is in the neighborhood iff* $(\widetilde{\mathbf{x}}, \underset{\sim}{\mathbf{s}})$ *is. Furthermore we have*

$$\mathcal{N}_F(\gamma) \subseteq \mathcal{N}_2(\gamma) \subseteq \mathcal{N}_{-\infty}(\gamma) \subseteq \mathcal{K} \times \mathcal{K}.$$

Proof. For the first claim, note that all neighborhoods can be defined in terms of eigenvalues of $\mathbf{w}$ and by proposition 21 $\mathbf{w}$ and $\widetilde{\mathbf{w}}$ have the same eigenvalues.

From the eigenvalue characterization of the centrality measures, it is easy to see that

$$d_F(\mathbf{x}, \mathbf{s}) \geq d_2(\mathbf{x}, \mathbf{s}) \geq d_{-\infty}(\mathbf{x}, \mathbf{s}),$$

which proves the first two inclusions. The last follows from the definition of $\mathcal{F}^0(P)$ and $\mathcal{F}^0(D)$. □

Lemma 30. *Let* $(\mathbf{x}, \mathbf{s}) \in \operatorname{Int}\mathcal{K} \times \operatorname{Int}\mathcal{K}$. *Then the following relations hold.*

$$\|\mathbf{x} \circ \mathbf{s} - \mu\mathbf{e}\|_{\mathrm{F}} \geq \|\mathbf{w} - \mu\mathbf{e}\|_{\mathrm{F}} \tag{30}$$

$$\|\mathbf{x} \circ \mathbf{s} - \mu\mathbf{e}\|_2 \geq \|\mathbf{w} - \mu\mathbf{e}\|_2 \tag{31}$$

$$\min_i \lambda_i(\mu\mathbf{e} - \mathbf{x} \circ \mathbf{s}) \geq \min_i \lambda_i(\mu\mathbf{e} - \mathbf{w}) \tag{32}$$

with equality holding if $\mathbf{x}$ *and* $\mathbf{s}$ *operator commute.*

Proof. For (30), note that $L(\mathbf{x})$ and $L(\mathbf{x}^2)$ commute and the matrix $L(\mathbf{x}^2) - L(\mathbf{x})^2 = L(\mathbf{x})^2 - \mathbf{Q}_{\mathbf{x}}$ which by lemma 12, has eigenvalues

$$\frac{(\lambda_i + \lambda_j)^2}{2} - \lambda_i\lambda_j \geq 0.$$

Thus, $L(\mathbf{x}^2) \succcurlyeq L(\mathbf{x})^2$ and we have:

$$\begin{aligned}
\|\mathbf{x} \circ \mathbf{s} - \mu\mathbf{e}\|_{\mathrm{F}}^2 &= \langle \mathbf{x} \circ \mathbf{s}, \mathbf{x} \circ \mathbf{s}\rangle - 2\mu\langle \mathbf{x} \circ \mathbf{s}, \mathbf{e}\rangle + \mu^2\langle \mathbf{e}, \mathbf{e}\rangle \\
&= \langle \mathbf{s}, L(\mathbf{x})^2\mathbf{s}\rangle - 2\mu\langle \mathbf{x}, \mathbf{s}\rangle + \mu^2\langle \mathbf{e}, \mathbf{e}\rangle \qquad \text{by associativity of } \langle\cdot,\cdot\rangle \\
&= \langle \mathbf{s}, 2L(\mathbf{x})^2\mathbf{s}\rangle - \langle \mathbf{s}, L(\mathbf{x})^2\mathbf{s}\rangle - 2\mu\langle \mathbf{x}, \mathbf{s}\rangle + \mu^2\langle \mathbf{e}, \mathbf{e}\rangle \\
&\geq \langle \mathbf{s}, 2L(\mathbf{x})^2\mathbf{s}\rangle - \langle \mathbf{s}, L(\mathbf{x}^2)\mathbf{s}\rangle - 2\mu\langle \mathbf{x}, \mathbf{s}\rangle + \mu^2\langle \mathbf{e}, \mathbf{e}\rangle \\
&= \langle \mathbf{s}, (2L(\mathbf{x})^2 - L(\mathbf{x}^2))\mathbf{s}\rangle - 2\mu\langle \mathbf{Q}_{\mathbf{x}^{1/2}}\mathbf{e}, \mathbf{s}\rangle + \mu^2\langle \mathbf{e}, \mathbf{e}\rangle \quad \text{by } \mathbf{Q}_{\mathbf{x}^{1/2}}\mathbf{e} = \mathbf{x} \\
&= \langle \mathbf{s}, \mathbf{Q}_{\mathbf{x}}\mathbf{s}\rangle - 2\mu\langle \mathbf{Q}_{\mathbf{x}^{1/2}}\mathbf{e}, \mathbf{s}\rangle + \mu^2\langle \mathbf{e}, \mathbf{e}\rangle \\
&= \langle \mathbf{Q}_{\mathbf{x}^{1/2}}\mathbf{s}, \mathbf{Q}_{\mathbf{x}^{1/2}}\mathbf{s}\rangle - 2\mu\langle \mathbf{Q}_{\mathbf{x}^{1/2}}\mathbf{s}, \mathbf{e}\rangle + \mu^2\langle \mathbf{e}, \mathbf{e}\rangle \qquad \text{by } \mathbf{Q}_{\mathbf{x}^{1/2}}^2 = \mathbf{Q}_{\mathbf{x}} \\
&= \|\mathbf{w} - \mu\mathbf{e}\|_{\mathrm{F}}^2
\end{aligned}$$

For the last two inequalities, it is enough to show that $\lambda_{\min}(\mathbf{x} \circ \mathbf{s}) \leq \lambda_{\min}(\mathbf{w})$ and $\lambda_{\max}(\mathbf{x} \circ \mathbf{s}) \geq \lambda_{\max}(\mathbf{w})$. By part 3 of lemma 8 $\mathbf{Q}_{\mathbf{x}^{1/2},\mathbf{x}^{-1/2}}\mathbf{Q}_{\mathbf{x}^{1/2}} = L(\mathbf{x})$ and therefore we have:

$$\mathbf{x} \circ \mathbf{s} = L(\mathbf{x})\mathbf{s} = \mathbf{Q}_{\mathbf{x}^{1/2},\mathbf{x}^{-1/2}}\mathbf{w}.$$

Furthermore

$$\begin{aligned}
\operatorname{tr}\left(\mathbf{Q}_{\mathbf{x}^{1/2},\mathbf{x}^{-1/2}}\mathbf{u}\right) &= 2\operatorname{tr}(L(\mathbf{x}^{1/2})L(\mathbf{x}^{-1/2})\mathbf{u}) - tr(L(\mathbf{e})\mathbf{u}) \\
&= 2\operatorname{tr}(\mathbf{x}^{1/2} \circ (\mathbf{x}^{-1/2} \circ \mathbf{u})) - \operatorname{tr}(\mathbf{u}) \\
&= 2\langle \mathbf{x}^{1/2}, \mathbf{x}^{-1/2} \circ \mathbf{u}\rangle - \operatorname{tr}(\mathbf{u}) \\
&= 2\operatorname{tr}((\mathbf{x}^{1/2} \circ \mathbf{x}^{-1/2}) \circ \mathbf{u}) - \operatorname{tr}(\mathbf{u}) = \operatorname{tr}(\mathbf{u}).
\end{aligned}$$

So we obtain

$$
\begin{aligned}
\lambda_{\min}(\mathbf{x} \circ \mathbf{s}) &= \min_{\mathbf{u}} \frac{\langle \mathbf{u}, (\mathbf{x} \circ \mathbf{s}) \circ \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \\
&= \min_{\langle \mathbf{u}, \mathbf{u} \rangle = 1} \langle \mathbf{u}^2, \mathbf{x} \circ \mathbf{s} \rangle = \min_{\mathrm{tr}(\mathbf{u}^2)=1} \langle \mathbf{u}^2, \mathbf{Q}_{\mathbf{x}^{1/2}, \mathbf{x}^{-1/2}} \mathbf{w} \rangle \\
&= \min_{\mathrm{tr}(\mathbf{u}^2)=1} \langle \mathbf{Q}_{\mathbf{x}^{1/2}, \mathbf{x}^{-1/2}} \mathbf{u}^2, \mathbf{w} \rangle \\
&\leq \min\{\langle t^2, \mathbf{w} \rangle \mid \mathbf{Q}_{\mathbf{x}^{1/2}, \mathbf{x}^{-1/2}} \mathbf{u}^2 = t^2, \mathrm{tr}(t^2) = 1\} \\
&= \min_{\mathrm{tr}(t^2)=1} \langle t^2, \mathbf{w} \rangle = \lambda_{\min}(\mathbf{w}).
\end{aligned}
$$

Similarly

$$
\begin{aligned}
\lambda_{\max}(\mathbf{x} \circ \mathbf{s}) &= \max_{\mathrm{tr}(\mathbf{u}^2)=1} \langle \mathbf{Q}_{\mathbf{x}^{1/2}, \mathbf{x}^{-1/2}} \mathbf{u}^2, \mathbf{w} \rangle \\
&\geq \max\{\langle t^2, \mathbf{w} \rangle \mid \mathbf{Q}_{\mathbf{x}^{1/2}, \mathbf{x}^{-1/2}} \mathbf{u}^2 = \mathbf{t}^2, \mathrm{tr}(\mathbf{t}^2) = 1\} \\
&= \max_{\mathrm{tr}(\mathbf{t}^2)=1} \langle \mathbf{t}^2, \mathbf{w} \rangle = \lambda_{\max}(\mathbf{w}).
\end{aligned}
$$

For the equality part, we note that if $\mathbf{x}$ and $\mathbf{s}$ operator commute then so do $\mathbf{x}^{1/2}$ and $\mathbf{s}$, which implies

$$
\begin{aligned}
\mathbf{w} = \mathbf{Q}_{\mathbf{x}^{1/2}} \mathbf{s} &= [2L(\mathbf{x}^{1/2})^2 - L(\mathbf{x})]\mathbf{s} \\
&= 2L(\mathbf{x}^{1/2})^2 L(\mathbf{s})\mathbf{e} - \mathbf{x} \circ \mathbf{s} \\
&= 2L(\mathbf{s})L(\mathbf{x}^{1/2})^2 \mathbf{e} - \mathbf{x} \circ \mathbf{s} \\
&= 2L(\mathbf{s})\mathbf{x} - \mathbf{x} \circ \mathbf{s} = \mathbf{x} \circ \mathbf{s}.
\end{aligned}
$$

In the following we assume that $\mathbf{x} \in \mathcal{F}^0(P)$ and $(\mathbf{s}, \mathbf{y}) \in \mathcal{F}^0(D)$, that we have fixed a $\mathbf{p} \in \mathfrak{C}(\mathbf{x}, \mathbf{s})$, and that $(\Delta\mathbf{x}, \Delta\mathbf{s}, \Delta\mathbf{y})$ is the solution to (23). We use the following notation

$$
\begin{gathered}
\widetilde{\mathbf{x}}(\alpha) = \widetilde{\mathbf{x}} + \alpha \widetilde{\Delta\mathbf{x}} \quad \underset{\sim}{\mathbf{s}}(\alpha) = \underset{\sim}{\mathbf{s}} + \alpha \underset{\sim}{\Delta}\,\mathbf{s} \\
\mathbf{x}(\alpha) = \mathbf{x} + \alpha \Delta\mathbf{x} \quad \mathbf{s}(\alpha) = \mathbf{s} + \alpha \Delta\mathbf{s} \\
\mu(\alpha) = \mu(\mathbf{x}(\alpha), \mathbf{s}(\alpha)) = \frac{\langle \mathbf{x}(\alpha), \mathbf{s}(\alpha) \rangle}{r} \\
\widetilde{\mathbf{w}}(\alpha) = \mathbf{Q}_{\widetilde{\mathbf{x}}(\alpha)^{1/2}} \underset{\sim}{\mathbf{s}}(\alpha)
\end{gathered}
$$

Lemma 31. *We have*

$$
\langle \widetilde{\Delta\mathbf{x}}, \underset{\sim}{\Delta}\,\mathbf{s} \rangle = \langle \Delta\mathbf{x}, \Delta\mathbf{s} \rangle = 0 \tag{33}
$$

$$
\mu(\alpha) = (1 - \alpha + \sigma\alpha)\mu. \tag{34}
$$

Proof. Feasibility of $(\mathbf{x}, \mathbf{s}, \mathbf{y})$ implies $\langle \Delta\mathbf{x}, \Delta\mathbf{s} \rangle = 0$. And by symmetry of $\mathbf{Q}_{\mathbf{p}}$ we have

$$
\langle \mathbf{Q}_{\mathbf{p}} \Delta\mathbf{x}, \mathbf{Q}_{\mathbf{p}^{-1}} \Delta\mathbf{s} \rangle = \langle \Delta\mathbf{x}, \mathbf{Q}_{\mathbf{p}} \mathbf{Q}_{\mathbf{p}^{-1}} \Delta\mathbf{s} \rangle = \langle \Delta\mathbf{x}, \Delta\mathbf{s} \rangle.
$$

For the second claim, we observe

$$\begin{aligned}\langle \mathbf{x}(\alpha), \mathbf{s}(\alpha)\rangle = \langle \widetilde{\mathbf{x}}(\alpha), \underset{\sim}{\mathbf{s}}(\alpha)\rangle &= \langle \widetilde{\mathbf{x}}, \underset{\sim}{\mathbf{s}}\rangle + \alpha(\langle \widetilde{\mathbf{x}}, \underset{\sim}{\Delta}\mathbf{s}\rangle + \langle \underset{\sim}{\mathbf{s}}, \widetilde{\Delta\mathbf{x}}\rangle) + \alpha^2\langle \widetilde{\Delta\mathbf{x}}, \underset{\sim}{\Delta}\mathbf{s}\rangle \\ &= (1-\alpha)\langle \widetilde{\mathbf{x}}, \underset{\sim}{\mathbf{s}}\rangle + \alpha\langle \widetilde{\mathbf{x}} \circ \underset{\sim}{\Delta}\mathbf{s} + \underset{\sim}{\mathbf{s}} \circ \widetilde{\Delta\mathbf{x}} + \widetilde{\mathbf{x}} \circ \underset{\sim}{\mathbf{s}}, \mathbf{e}\rangle \\ &= (1-\alpha)\langle \widetilde{\mathbf{x}}, \underset{\sim}{\mathbf{s}}\rangle + \alpha\langle \sigma\mu\mathbf{e}, \mathbf{e}\rangle \\ &= (1-\alpha)\langle \mathbf{x}, \mathbf{s}\rangle + \alpha\sigma\mu r \\ &= r[(1-\alpha)\mu + \alpha\sigma\mu] = r\mu[(1-\alpha) + \alpha\sigma].\end{aligned}$$

From now on we will assume that $\mathbf{p} \in \mathfrak{C}(\mathbf{x}, \mathbf{s})$. Therefore, the subsequent analysis applies only to the commutative subclass of the Monteiro-Zhang family of directions.

Lemma 32. *Let* $\delta_x = \|\widetilde{\Delta\mathbf{x}}\|_{\mathrm{F}}$ *and* $\delta_s = \left\|\underset{\sim}{\Delta}\mathbf{s}\right\|_{\mathrm{F}}$. *Then we have*

1. *If* $(\mathbf{x}, \mathbf{s}) \in \mathcal{N}_F(\gamma)$, *then* $(\mathbf{x}(\alpha), \mathbf{s}(\alpha)) \in \mathcal{N}_F(\gamma)$ *for* $0 \le \alpha \le \gamma\sigma\mu/2\delta_x\delta_s$.
2. *If* $(\mathbf{x}, \mathbf{s}) \in \mathcal{N}_2(\gamma)$, *then* $(\mathbf{x}(\alpha), \mathbf{s}(\alpha)) \in \mathcal{N}_2(\gamma)$ *for* $0 \le \alpha \le \gamma\sigma\mu/2\delta_x\delta_s$.
3. *If* $(\mathbf{x}, \mathbf{s}) \in \mathcal{N}_{-\infty}(\gamma)$, *then* $(\mathbf{x}(\alpha), \mathbf{s}(\alpha)) \in \mathcal{N}_{-\infty}(\gamma)$ *for* $0 \le \alpha \le \gamma\sigma\mu/2\delta_x\delta_s$.

Proof. We will show that the claims hold for $(\widetilde{\mathbf{x}}, \underset{\sim}{\mathbf{s}})$, which by proposition 29 is enough. From the Newton system and the fact that $\widetilde{\mathbf{x}}$ and $\underset{\sim}{\mathbf{s}}$ operator commute, we obtain

$$\begin{aligned}\widetilde{\mathbf{x}}(\alpha) \circ \underset{\sim}{\mathbf{s}}(\alpha) - \mu(\alpha)\mathbf{e} &= (\widetilde{\mathbf{x}} + \alpha\widetilde{\Delta\mathbf{x}}) \circ (\underset{\sim}{\mathbf{s}} + \alpha\underset{\sim}{\Delta}\mathbf{s}) - \mu(\alpha)\mathbf{e} \\ &= \widetilde{\mathbf{x}} \circ \underset{\sim}{\mathbf{s}} + \alpha(\widetilde{\mathbf{x}} \circ \underset{\sim}{\Delta}\mathbf{s} + \underset{\sim}{\mathbf{s}} \circ \widetilde{\Delta\mathbf{x}}) + \alpha^2\widetilde{\Delta\mathbf{x}} \circ \underset{\sim}{\Delta}\mathbf{s} - \mu(\alpha)\mathbf{e} \\ &= (1-\alpha)(\widetilde{\mathbf{x}} \circ \underset{\sim}{\mathbf{s}} - \mu\mathbf{e}) \\ &\qquad + \alpha(\widetilde{\mathbf{x}} \circ \underset{\sim}{\mathbf{s}} + \widetilde{\mathbf{x}} \circ \underset{\sim}{\Delta}\mathbf{s} + \underset{\sim}{\mathbf{s}} \circ \widetilde{\Delta\mathbf{x}} - \sigma\mu\mathbf{e}) + \alpha^2\widetilde{\Delta\mathbf{x}} \circ \underset{\sim}{\Delta}\mathbf{s} \\ &= (1-\alpha)(\widetilde{\mathbf{w}} - \mu\mathbf{e}) + \alpha^2\widetilde{\Delta\mathbf{x}} \circ \underset{\sim}{\Delta}\mathbf{s}.\end{aligned}$$

For the first statement, this implies

$$\begin{aligned}\|\widetilde{\mathbf{x}}(\alpha) \circ \underset{\sim}{\mathbf{s}}(\alpha) - \mu(\alpha)\mathbf{e}\|_{\mathrm{F}} &\le (1-\alpha)\,\|\widetilde{\mathbf{w}} - \mu\mathbf{e}\|_{\mathrm{F}} + \alpha^2\delta_x\delta_s \\ &\le (1-\alpha)\gamma\mu + \alpha^2\delta_x\delta_s\end{aligned}$$

So we have

$$\|\widetilde{\mathbf{x}}(\alpha) \circ \underset{\sim}{\mathbf{s}}(\alpha) - \mu(\alpha)\mathbf{e}\|_{\mathrm{F}} \le \gamma\mu(\alpha) \tag{35}$$

if

$$(1-\alpha)\gamma\mu + \alpha^2\delta_x\delta_s \le \gamma\mu(\alpha) = \gamma(1-\alpha+\sigma\alpha)\mu.$$

Solving this inequality for α, we see that (35) holds if $0 \le \alpha \le \gamma\sigma\mu/\delta_x\delta_s = \alpha^*$. So in the interval $[0, \alpha^*]$, $\widetilde{\mathbf{x}}(\alpha) \circ \underset{\sim}{\mathbf{s}}(\alpha)$ is positive definite and therefore $det(\widetilde{\mathbf{x}}(\alpha) \circ \underset{\sim}{\mathbf{s}}(\alpha)) > 0$. We now need to prove that in this interval both $\widetilde{\mathbf{x}}(\alpha)$ and $\underset{\sim}{\mathbf{s}}(\alpha)$ are positive definite. So assume they are not. Then by continuity of $\lambda_{\min}$ we can without loss of generality assume that there exists a value α_0 in the interval such that $\widetilde{\mathbf{x}}(\alpha_0) \not\succcurlyeq \mathbf{0}$. But then we can note that we didn't use positive definiteness of $\mathbf{s}$ in the proof of (32) and that by continuity it also holds for $\mathbf{x} \succcurlyeq \mathbf{0}$. Therefore $\lambda_{\min}(\widetilde{\mathbf{w}}) > 0$ and hence $\det(\widetilde{\mathbf{w}}(\alpha)) > 0$. But proposition 5 tells us that $\det(\widetilde{\mathbf{w}}(\alpha)) = \det(\widetilde{\mathbf{x}}(\alpha)) \det(\underset{\sim}{\mathbf{s}}(\alpha))$ which is clearly zero. This is a contradiction, so $\widetilde{\mathbf{x}}(\alpha)$ and $\underset{\sim}{\mathbf{s}}(\alpha)$ are positive definite in the interval. Now we can invoke lemma 30 and conclude that $(\widetilde{\mathbf{x}}(\alpha), \underset{\sim}{\mathbf{s}}(\alpha)) \in \mathcal{N}_F(\gamma)$.

For the second and third claim, we use lemmas 14 and 30 and the fact that $\lambda_{\min}(\mathbf{w}) - \mu \geq -\gamma\mu$ to get

$$\begin{aligned}
\lambda_{\min}[\widetilde{\mathbf{x}}(\alpha) \circ \underset{\sim}{\mathbf{s}}(\alpha) - \mu(\alpha)\mathbf{e}] &= \lambda_{\min}[(1-\alpha)(\widetilde{\mathbf{x}} \circ \underset{\sim}{\mathbf{s}} - \mu\mathbf{e}) + \alpha^2 \widetilde{\Delta\mathbf{x}} \circ \underset{\sim}{\Delta}\, \mathbf{s}] \\
&\geq \lambda_{\min}[(1-\alpha)(\widetilde{\mathbf{x}} \circ \underset{\sim}{\mathbf{s}} - \mu\mathbf{e})] - \alpha^2 \left\| \widetilde{\Delta\mathbf{x}} \circ \underset{\sim}{\Delta}\, \mathbf{s} \right\|_{\mathrm{F}} \\
&\geq (1-\alpha)\lambda_{\min}[\widetilde{\mathbf{x}} \circ \underset{\sim}{\mathbf{s}} - \mu\mathbf{e}] - \alpha^2 \delta_x \delta_s \\
&\geq (\alpha - 1)\gamma\mu - \alpha^2 \delta_x \delta_s.
\end{aligned}$$

So we have

$$\mu(\alpha) - \lambda_{\min}[\widetilde{\mathbf{x}}(\alpha) \circ \underset{\sim}{\mathbf{s}}(\alpha] \leq \gamma\mu(\alpha)$$

if

$$(1-\alpha)\gamma\mu + \alpha^2 \delta_x \delta_s \leq \gamma\mu(\alpha) = \gamma\mu(1 - \alpha + \sigma\alpha).$$

The solution to this inequality is again $0 \leq \alpha \leq \alpha^*$. Also, $\widetilde{\mathbf{x}}(\alpha) \circ \underset{\sim}{\mathbf{s}}(\alpha) \succ \mathbf{0}$, so $\widetilde{\mathbf{x}}(\alpha), \underset{\sim}{\mathbf{s}}(\alpha) \in \operatorname{Int} \mathcal{K}$ and by lemma 30, $(\widetilde{\mathbf{x}}(\alpha), \underset{\sim}{\mathbf{s}}(\alpha)) \in \mathcal{N}_{-\infty}(\gamma)$.

For the second claim, lemma 14 also gives us a bound for the largest eigenvalue.

$$\lambda_{\max}[\widetilde{\mathbf{x}}(\alpha) \circ \underset{\sim}{\mathbf{s}}(\alpha) - \mu(\alpha)\mathbf{e}] \leq (1-\alpha)\gamma\mu + \alpha^2 \delta_x \delta_s,$$

so similarly to the smallest eigenvalue we have

$$\lambda_{\max}[\widetilde{\mathbf{x}}(\alpha) \circ \underset{\sim}{\mathbf{s}}(\alpha] - \mu(\alpha) \leq \gamma\mu(\alpha)$$

if

$$(1-\alpha)\gamma\mu + \alpha^2 \delta_x \delta_s \leq \gamma\mu(\alpha)$$

which completes the proof for $\mathcal{N}_2$.

3.1. Bounds for $\delta_x \delta_s$

Lemma 33. *Let $\mathbf{u}, \mathbf{v} \in \mathcal{J}$ and G a positive definite matrix which is symmetric with respect to the scalar product $\langle \cdot, \cdot \rangle$. Then*

$$\|\mathbf{u}\|_{\mathrm{F}} \, \|\mathbf{v}\|_{\mathrm{F}} \leq \frac{1}{2}\sqrt{\operatorname{cond}(G)} \Big(\left\| G^{1/2}\mathbf{u} \right\|_{\mathrm{F}}^2 + \left\| G^{-1/2}\mathbf{v} \right\|_{\mathrm{F}}^2 \Big), \tag{36}$$

where $\operatorname{cond}(G) = \lambda_{\max}(G)/\lambda_{\min}(G)$.

Proof. By lemma 13, we have

$$\lambda_{\min}(G) = \min_{\mathbf{x}} \frac{\langle \mathbf{x}, G\mathbf{x}\rangle}{\langle \mathbf{x}, \mathbf{x}\rangle} = \min_{\mathbf{x}} \frac{\langle G^{1/2}\mathbf{x}, G^{1/2}\mathbf{x}\rangle}{\langle \mathbf{x}, \mathbf{x}\rangle} = \min_{\mathbf{x}} \frac{\left\|G^{1/2}\mathbf{x}\right\|_F^2}{\|\mathbf{x}\|_F^2}.$$

Which implies

$$\forall \mathbf{x} \in \mathcal{J}: \quad \|\mathbf{x}\|_F^2 \leq \frac{\left\|G^{1/2}\mathbf{x}\right\|_F^2}{\lambda_{\min}(G)}.$$

Choosing G for $\mathbf{u}$ and G^{-1} for $\mathbf{v}$, we obtain

$$\begin{aligned} \|\mathbf{u}\|_F \, \|\mathbf{v}\|_F &\leq \frac{\left\|G^{1/2}\mathbf{u}\right\|_F}{\sqrt{\lambda_{\min}(G)}} \frac{\left\|G^{-1/2}\mathbf{v}\right\|_F}{\sqrt{\lambda_{\min}(G^{-1})}} \\ &= \sqrt{\frac{1}{\lambda_{\min}(G)\lambda_{\min}(G^{-1})}} \left\|G^{1/2}\mathbf{u}\right\|_F \left\|G^{-1/2}\mathbf{v}\right\|_F, \end{aligned}$$

which together with $\lambda_{\min}(G^{-1}) = 1/\lambda_{\max}(G)$ and $2ab \leq a^2 + b^2$ proves the claim.

Lemma 34. *Let* $G = L(\underset{\sim}{\mathbf{s}})^{-1} L(\widetilde{\mathbf{x}})$. *Then*

$$\begin{aligned} \delta_x \delta_s &\leq \frac{1}{2}\sqrt{\operatorname{cond}(G)} \left(\sigma^2\mu^2 \operatorname{tr}(\widetilde{\mathbf{w}}^{-1}) - 2\sigma\mu \operatorname{tr}(\mathbf{e}) + \operatorname{tr}(\widetilde{\mathbf{w}})\right) \\ &= \frac{1}{2}\sqrt{\operatorname{cond}(G)} \sum_{i=1}^{r} \frac{(\sigma\mu - \lambda_i(\widetilde{\mathbf{w}}))^2}{\lambda_i(\widetilde{\mathbf{w}})}. \end{aligned}$$

Proof. Since $\widetilde{\mathbf{x}}$ and $\underset{\sim}{\mathbf{s}}$ operator commute, G is a symmetric matrix and

$$\begin{aligned} (L(\widetilde{\mathbf{x}})L(\underset{\sim}{\mathbf{s}}))^{-1/2} L(\widetilde{\mathbf{x}}) &= L(\underset{\sim}{\mathbf{s}})^{-1/2} L(\widetilde{\mathbf{x}})^{1/2} = G^{1/2} \\ (L(\widetilde{\mathbf{x}})L(\underset{\sim}{\mathbf{s}}))^{-1/2} L(\underset{\sim}{\mathbf{s}}) &= L(\underset{\sim}{\mathbf{s}})^{1/2} L(\widetilde{\mathbf{x}})^{-1/2} = G^{-1/2}. \end{aligned}$$

Multiplying the last equation of (23) by $(L(\widetilde{\mathbf{x}})L(\underset{\sim}{\mathbf{s}}))^{-1/2}$, we obtain

$$G^{1/2} \underset{\sim}{\Delta}\,\mathbf{s} + G^{-1/2} \widetilde{\Delta\mathbf{x}} = \sigma\mu \, (L(\widetilde{\mathbf{x}})L(\underset{\sim}{\mathbf{s}}))^{-1/2} e - G^{1/2}\underset{\sim}{\mathbf{s}}.$$

Taking norm-squared on both sides, we have

$$\begin{aligned}
\left\| G^{1/2} \underset{\sim}{\Delta}\, \mathbf{s} + G^{-1/2} \widetilde{\Delta \mathbf{x}} \right\|_{\mathrm{F}}^2 &= \left\| G^{1/2} \underset{\sim}{\Delta}\, \mathbf{s} \right\|_{\mathrm{F}}^2 + \left\| G^{-1/2} \widetilde{\Delta \mathbf{x}} \right\|_{\mathrm{F}}^2 \\
&= \left\| \sigma\mu \left(L(\widetilde{\mathbf{x}})L(\underset{\sim}{\mathbf{s}})\right)^{-1/2} \mathbf{e} - G^{1/2} \underset{\sim}{\mathbf{s}} \right\|_{\mathrm{F}}^2 \\
&= \sigma^2\mu^2 \langle (L(\widetilde{\mathbf{x}})L(\underset{\sim}{\mathbf{s}}))^{-1/2}\, \mathbf{e}, (L(\widetilde{\mathbf{x}})L(\underset{\sim}{\mathbf{s}}))^{-1/2}\, \mathbf{e} \rangle \\
&\quad - 2\sigma\mu \langle (L(\widetilde{\mathbf{x}})L(\underset{\sim}{\mathbf{s}}))^{-1/2}\, \mathbf{e}, (L(\widetilde{\mathbf{x}})L(\underset{\sim}{\mathbf{s}}))^{-1/2}\, L(\widetilde{\mathbf{x}})\underset{\sim}{\mathbf{s}} \rangle \\
&\quad + \langle G^{1/2}\underset{\sim}{\mathbf{s}}, G^{1/2}\underset{\sim}{\mathbf{s}} \rangle \\
&= \sigma^2\mu^2 \langle \mathbf{e}, (L(\widetilde{\mathbf{x}})L(\underset{\sim}{\mathbf{s}}))^{-1}\, \mathbf{e} \rangle \\
&\quad - 2\sigma\mu \langle (L(\widetilde{\mathbf{x}})L(\underset{\sim}{\mathbf{s}}))^{-1}\, \mathbf{e}, L(\widetilde{\mathbf{x}})\underset{\sim}{\mathbf{s}} \rangle + \langle G\underset{\sim}{\mathbf{s}}, \underset{\sim}{\mathbf{s}} \rangle \\
&= \sigma^2\mu^2 \langle \widetilde{\mathbf{x}}^{-1}, \underset{\sim}{\mathbf{s}}^{-1} \rangle - 2\sigma\mu \langle \mathbf{e}, \mathbf{e} \rangle + \langle \widetilde{\mathbf{x}}, \underset{\sim}{\mathbf{s}} \rangle \\
&= \sigma^2\mu^2 \langle \mathbf{e}, \widetilde{\mathbf{w}}^{-1} \rangle - 2\sigma\mu \langle \mathbf{e}, \mathbf{e} \rangle + \langle \mathbf{e}, \widetilde{\mathbf{w}} \rangle \\
&= \sigma^2\mu^2 \operatorname{tr}(\widetilde{\mathbf{w}}^{-1}) - 2\sigma\mu \operatorname{tr}(\mathbf{e}) + \operatorname{tr}(\widetilde{\mathbf{w}}) \\
&= \sum_{i=1}^{r} \frac{(\sigma\mu - \lambda_i(\widetilde{\mathbf{w}}))^2}{\lambda_i(\widetilde{\mathbf{w}})}
\end{aligned}$$

which combined with lemma 33 proves the claim.

Lemma 35. *Let $G = L(\underset{\sim}{\mathbf{s}})^{-1} L(\widetilde{\mathbf{x}})$. Then we have the following neighborhood specific bounds for $\sum_{i=1}^{r} (\sigma\mu - \lambda_i(\widetilde{\mathbf{w}}))^2 / \lambda_i(\widetilde{\mathbf{w}})$.*

1. *If $(\mathbf{x}, \mathbf{s}) \in \mathcal{N}_F(\gamma)$ then*

$$\sum_{i=1}^{r} (\sigma\mu - \lambda_i(\widetilde{\mathbf{w}}))^2 / \lambda_i(\widetilde{\mathbf{w}}) \le \left(\frac{\gamma^2 + (1-\sigma)^2 r}{1-\gamma} \right) \mu.$$

2. *If $(\mathbf{x}, \mathbf{s}) \in \mathcal{N}_2(\gamma)$ or $(\mathbf{x}, \mathbf{s}) \in \mathcal{N}_{-\infty}$ then*

$$\sum_{i=1}^{r} (\sigma\mu - \lambda_i(\widetilde{\mathbf{w}}))^2 / \lambda_i(\widetilde{\mathbf{w}}) \le \left(1 - 2\sigma + \frac{\sigma^2}{1-\gamma} \right) \mu r.$$

Proof. If $(\mathbf{x}, \mathbf{s}) \in \mathcal{N}_F(\gamma)$ then $\lambda_i(\widetilde{\mathbf{w}}) \ge (1-\gamma)\mu$ and we have

$$\begin{aligned}
\sum_{i=1}^{r} \frac{(\sigma\mu - \lambda_i(\widetilde{\mathbf{w}}))^2}{\lambda_i(\widetilde{\mathbf{w}})} &\le \frac{1}{(1-\gamma)\mu} \sum_{i=1}^{r} (\sigma\mu - \lambda_i(\widetilde{\mathbf{w}}))^2 \\
&= \frac{1}{(1-\gamma)\mu} \sum_{i=1}^{r} (\mu - \lambda_i(\widetilde{\mathbf{w}}) - (1-\sigma)\mu)^2
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{(1-\gamma)\mu} \sum_{i=1}^{r} (\mu - \lambda_i(\widetilde{\mathbf{w}}))^2 + ((1-\sigma)\mu)^2 \\
&\le \frac{1}{(1-\gamma)\mu} \left(\gamma^2\mu^2 + (1-\sigma)^2\mu^2 r\right) \\
&\le \frac{\gamma^2 + (1-\sigma)^2 r}{1-\gamma}\mu,
\end{aligned}
$$

where in the second equality we used $\sum \lambda_i(\widetilde{\mathbf{w}}) = r\mu$. For the second statement we use that the eigenvalues of $\widetilde{\mathbf{w}}^{-1}$ are the reciprocals of the eigenvalues of $\widetilde{\mathbf{w}}$, so they are all bounded from above by $1/(1-\gamma)\mu$.

$$
\begin{aligned}
\sigma^2\mu^2 \operatorname{tr}(\widetilde{\mathbf{w}}^{-1}) - 2\sigma\mu \operatorname{tr}(\mathbf{e}) + \operatorname{tr}(\widetilde{\mathbf{w}}) &= \sigma^2\mu^2 \sum_{i=1}^{r} \left(\lambda_i(\widetilde{\mathbf{w}}^{-1})\right) - 2\sigma\mu r + \mu r \\
&\le \sigma^2\mu^2 r \frac{1}{(1-\gamma)\mu} - 2\sigma\mu r + \mu r \\
&= \left(\frac{\sigma^2}{1-\gamma} - 2\sigma + 1\right)\mu r
\end{aligned}
$$

which completes the proof.

We now focus on three well-known classes of directions in the commutative class. When $\mathbf{p} = \mathbf{s}^{1/2}$, then $\underset{\sim}{\mathbf{s}} = \mathbf{e}$. When $\mathbf{p} = \mathbf{x}^{-1/2}$, then $\widetilde{\mathbf{x}} = \mathbf{e}$. Thus both choices are in the commutative class; the first one will be called the **xs** method and the second one the **sx** method.

Now suppose that we choose $\mathbf{p}$ in such a way that $\widetilde{\mathbf{x}} = \underset{\sim}{\mathbf{s}}$. Such a choice exists and is unique and leads to the Nesterov-Todd (NT) method. In this case

$$
\mathbf{p} = \left[Q_{\mathbf{x}^{1/2}}\left(Q_{\mathbf{x}^{1/2}}\mathbf{s}\right)^{-1/2}\right]^{-1/2} = \left[Q_{\mathbf{s}^{-1/2}}\left(Q_{\mathbf{s}^{1/2}}\mathbf{x}\right)^{1/2}\right]^{-1/2}.
$$

Lemma 36 (Bounds for cond G). *For the Nesterov-Todd method, the condition number of G is always* 1. *For the* **xs** *and* **sx** *methods, we have*

(a) If $(\mathbf{x}, \mathbf{s})$ *is in* $\mathcal{N}_F(\gamma)$ *or* $\mathcal{N}_2(\gamma)$ *then* $\operatorname{cond}(G) \le 2/(1-\gamma)$.
(b) If $(\mathbf{x}, \mathbf{s})$ *is in* $\mathcal{N}_{-\infty}(\gamma)$ *then* $\operatorname{cond}(G) \le r/(1-\gamma)$.

Proof. In the Nesterov-Todd method, $\mathbf{p}$ is chosen such that $\widetilde{\mathbf{x}} = \underset{\sim}{\mathbf{s}}$. Hence $G = L(\underset{\sim}{\mathbf{s}})^{-1} L(\widetilde{\mathbf{x}}) = I$ and the condition number is 1. In the **xs** method, $\mathbf{p}$ is chosen so that $\underset{\sim}{\mathbf{s}} = \mathbf{e}$. So $\widetilde{\mathbf{x}} = \widetilde{\mathbf{w}}$ and $G = L(\widetilde{\mathbf{w}})$. If $(\mathbf{x}, \mathbf{s}) \in \mathcal{N}_2(\gamma)$, we can bound $\lambda_{\max}(G) = \lambda_{\max}(\widetilde{\mathbf{w}})$ from above by $(1+\gamma)\mu$, and $\lambda_{\min}(G) = \lambda_{\min}(\widetilde{\mathbf{w}})$ from below by $(1-\gamma)\mu$. Hence

$$
\operatorname{cond}(G) = \frac{\lambda_{\max}(G)}{\lambda_{\min}(G)} \le \frac{1+\gamma}{1-\gamma} \le \frac{2}{1-\gamma}.
$$

Since $\mathcal{N}_F(\gamma) \subseteq \mathcal{N}_2(\gamma)$ this proves (a) for the xs method. For (b) we have the same bound for the smallest eigenvalue and the largest eigenvalue can trivially be bounded by $\operatorname{tr}(\widetilde{\mathbf{w}}) = \mu r$. Which leads to

$$
\operatorname{cond}(G) = \frac{\mu r}{(1-\gamma)\mu} \le \frac{r}{1-\gamma}.
$$

In the **sx** method, we have $\widetilde{\mathbf{x}} = \mathbf{e}$ and $\underset{\sim}{\mathbf{s}} = \widetilde{\mathbf{w}}$. So $G = L(\widetilde{\mathbf{w}})^{-1}$ and, since a matrix and its inverse have the same condition number, the proof is complete.

3.2. The path-following algorithm

The short-, semi-long-, and long-step path-following algorithms for the commutative class can be described as follows:

1. Choose $\epsilon \in (0, 1)$, $\sigma \in (0, 1)$, and $\gamma \in (0, 1)$, and let $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{s}^0) \in \mathcal{N}_\bullet(\gamma)$ be an interior feasible solution for a chosen neighborhood.
   Set $k = 0$, and $\mu_0 = \langle \mathbf{x}^0, \mathbf{s}^0 \rangle / r$.
2. Repeat until $\mu_k \le \epsilon \mu_0$:
   (a) Choose a scaling element $\mathbf{p} \in \mathfrak{C}(\mathbf{x}^k, \mathbf{s}^k)$ and compute $(\widetilde{\mathbf{x}}, \mathbf{y}, \underset{\sim}{\mathbf{s}})$.
   (b) Compute the Newton direction $(\Delta \mathbf{x}^k, \Delta \mathbf{y}^k, \Delta \mathbf{s}^k)$ by solving the scaled Newton system (23) and applying the inverse scaling to $(\widetilde{\Delta \mathbf{x}}, \underset{\sim}{\Delta}\, \mathbf{s})$.
   (c) Choose the largest step-size α_k such that
   $$(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{s}^{k+1}) = (\mathbf{x}^k, \mathbf{y}^k, \mathbf{s}^k) + \alpha_k (\Delta \mathbf{x}^k, \Delta \mathbf{y}^k, \Delta \mathbf{s}^k) \in \mathcal{N}_\bullet(\gamma).$$
   (d) Set $\mu_{k+1} = \langle \mathbf{x}^{k+1}, \mathbf{s}^{k+1} \rangle / r$ and increment k by 1.

In this algorithm, ϵ is the desired accuracy of the solution. The choice of σ, γ, and the neighborhood determines the general type of the algorithm. With $\mathcal{N}_F(\gamma)$ and $\sigma = 1 - \delta/\sqrt{r}$, $\delta \in (0, 1)$ we obtain the short-step algorithm. The semi-long-step algorithm is obtained by choosing $\mathcal{N}_2(\gamma)$ as the neighborhood and $\sigma \in (0, 1)$. Finally, choosing $\mathcal{N}_{-\infty}(\gamma)$ and $\sigma \in (0, 1)$ we obtain the long-step algorithm.

Our main result is the following theorem which describes the iteration complexity for the three path-following algorithms just described.

Theorem 37. *Assume that $\sqrt{\operatorname{cond}(G)}$ can be bounded from above by $\kappa < \infty$ for all iterations of the algorithm. Then the short-step algorithm terminates in $\mathcal{O}(\kappa\sqrt{r} \log \epsilon^{-1})$ iterations. The semi-long-, and long-step algorithms will terminate in $\mathcal{O}(\kappa r \log \epsilon^{-1})$ iterations.*

Proof. If we chose the neighborhood to be $\mathcal{N}_F(\gamma)$ and $\sigma = 1 - \delta/\sqrt{r}$ then by lemmas 34 and 35, we have
$$\delta_x \delta_s \le \frac{\kappa}{2} \left(\frac{\gamma^2 + (1-\sigma)^2 r}{1-\gamma} \right) \mu$$
and by lemma 32
$$\begin{aligned}\alpha^* = \frac{\gamma\sigma\mu}{2\delta_x\delta_s} &\ge \frac{\gamma\sigma\mu 2(1-\gamma)}{2(\gamma^2 + (1-\sigma)^2 r)\kappa\mu} = \frac{\gamma\sigma(1-\gamma)}{(\gamma^2 + (1-\sigma)^2 r)\kappa} \\ &= \frac{\gamma(1-\delta/\sqrt{r})(1-\gamma)}{(\gamma^2+\delta^2)\kappa} \ge \frac{\gamma(1-\gamma)}{2(\gamma^2+\delta^2)\kappa}.\end{aligned}$$

So the reduction of μ at each iteration is $1 - \mathcal{O}(1/\sqrt{r}\kappa)$ and the short-step algorithm has the claimed iteration complexity. Similarly, for $\mathcal{N}_2(\gamma)$ and $\mathcal{N}_{-\infty}(\gamma)$ the reduction is $1 - \mathcal{O}(1/\kappa r)$ per iteration and the algorithms will terminate in the number of iterations claimed.

We can specialize the algorithms further by prescribing the scaling element $\mathbf{p}$. Choosing $\mathbf{p}$ such that $\widetilde{\mathbf{x}} = \underset{\sim}{\mathbf{s}}$, we obtain the Nesterov-Todd search direction and have as an immediate corollary of theorem 37:

Corollary 38. *If in the path-following algorithm* $\mathbf{p} \in \mathfrak{C}(\mathbf{x}, \mathbf{s})$ *is chosen such that* $\widetilde{\mathbf{x}} = \underset{\sim}{\mathbf{s}}$, *the iteration complexities are*

- $\mathcal{O}(\sqrt{r} \log \epsilon^{-1})$, *for the short-step algorithm,*
- $\mathcal{O}(r \log \epsilon^{-1})$, *for the semi-long- and long-step algorithm.*

Proof. Immediate consequence of theorem 37 and the fact that $\kappa = 1$ by lemma 36.

The $\mathbf{xs}$ and $\mathbf{sx}$ search directions are obtained by choosing $\mathbf{p} = \mathbf{s}^{1/2}$ and $\mathbf{p} = \mathbf{x}^{-1/2}$ respectively. For these we have the following iteration complexities.

Corollary 39. *If in the path-following algorithm* $\mathbf{p} \in \mathfrak{C}(\mathbf{x}, \mathbf{s})$ *is chosen such that* $\underset{\sim}{\mathbf{s}} = \mathbf{e}$ *or* $\widetilde{\mathbf{x}} = \mathbf{e}$, *the iteration complexities are*

- $\mathcal{O}(\sqrt{r} \log \epsilon^{-1})$, *for the short-step algorithm,*
- $\mathcal{O}(r \log \epsilon^{-1})$, *for the semi-long-step algorithm,*
- $\mathcal{O}(r^{1.5} \log \epsilon^{-1})$, *for the long-step algorithm.*

Proof. Immediate consequence of theorem 37 and the bounds for κ given by lemma 36 for these two search directions.

4. Conclusion

In this paper we showed that Tsuchiya's and Monteiro and Zhang's analyses of polynomial iteration complexity of short, semi-long and long step path-following algorithms over the class of similarly scaled (Monteiro-Zhang) directions in SDP extend verbatim to optimization over symmetric cones. Furthermore algorithms that are not in the Monteiro-Zhang family, such as those of Tseng [Tse96] and Monteiro and Tsuchiya [MT96] can also be extended to symmetric cones and in particular applied to second order cone programming problems using techniques of this paper.

We reiterate an open problem mentioned in our earlier work [SA01]. It remains to be settled whether pure Jordan algebraic techniques, as employed in the present paper, can be applied to prove polynomiality of the $XS + SX$ method, or whether the assumption of an inducing associative algebra is inherently essential for this algorithm. Finally, we hope that this and the companion paper, along with the work of Faybusovich cited above, has convinced researchers investigating theoretical properties of semidefinite programming that it is more profitable to direct their attention to the more general problem of optimization over symmetric cones.

References

[AG02] Alizadeh, F., Goldfarb, D.: Second-order cone programming. Math. Program-ming Series B, 2002. To appear

[AHO98] Alizadeh, F., Haeberly, J.P., Overton, M.L.: Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results SIAM J. Optim. **8**(3), 746–768 (1998)

[Ali95] Alizadeh, F.: Interior point methods in semidefinite programming with applications to combinatorial optimization. SIAM J. Optim. **5**(1), 13–51 (1995)

[AS97] Alizadeh, F., Schmieta, S.H.: Optimization with Semidefinite, Quadratic and Linear Constraints. Technical Report rrr 23-97, RUT-COR, Rutgers University, November 1997 (Available at URL: http://rutcor.rutgers.edu/pub/rrr/reports97/23.ps)

[AS00] Alizadeh, F., Schmieta, S.H.: Symmetric Cones, Potential Reduction Methods and Word-By-Word Extensions. In: R. Saigal, L. Vandenberghe, and H. Wolkowicz, editors, Handbook of Semidefinite Programming, Theory, Algorithms and Applica-tions, pages 195–233. Kluwer Academic Publishers, 2000

[Fay97a] Faybusovich, L.: Euclidean Jordan algebras and interior-point algorithms. Positivity **1**(4), 331–357 (1997)

[Fay97b] Faybusovich, L.: Linear systems in Jordan algebras and primal-dual interior point algorithms. J. Comput. Appl. Math. **86**, 149–175 (1997)

[Fay98] Faybusovich, L.: A Jordan-algebraic approach to potential-reduction algorithms Technical report, Department of Mathematics, University of Notre Dame, Notre Dame, IN, USA, April 1998

[FK94] Faraut, J., Koráanyi, A.: Analysis on Symmetric Cones. Oxford University Press, Oxford, UK, 1994

[Gül97] Güler, O.: Personal communication. 1997

[HJ90] Horn, R., Johnson, C.: Topics in Matrix Analysis. Cambridge University Press, Cambridge, 1990

[HRVW96] Helmberg, C., Rendl, F., Vanderbei, R.J., Wolkowicz, H.: An interior-point method for semidefinite programming. SIAM J. Optim. **6**, 342–361 (1996)

[Jac68] Jacobson, N.: Structure and Representation of Jordan Algebras, volume XXXIX of Colloqium Publications. American Mathematical Society, Providence, Rhode Island, 1968

[KMY89] Kojima, M., Mizuno, S., Yoshise, A.: A primal-dual interior-point algorithm for linear programming. In: N. Megiddo, editor, Progress in Mathematical Programming, Berlin, 1989. Springer Verlag

[KSH97] Kojima, M., Shindoh, S., Hara, S.: Interior-point methods for the monotone linear complementarity problem in symmetric matrices. SIAM J. Optim. **7**(9), 86–125 (1997)

[MA88] Monteiro, R., Adler, I.: Polynomial-time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extension. Technical Report ESRC 88–8, Industrial Engineering and Operations Research Department, University of California–Berkeley, March 1988

[Mon96] Monteiro, R.D.C.: Polynomial Convergence of Primal-Dual Algorithms for Semidefinite Programming Based on Monteiro and Zhang Family of Directions. Submitted, 1996

[MT96] Monteiro, R.D.C., Tsuchiya, T.: Polynomial Convergence of a New Family of Primal-Dual Algorithms for Semidefinite Programming. Technical Report 627, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan, November 1996

[MT98] Monteiro, R.D.C., Tsuchiya, T.: Polynomial Convergence of Primal-Dual Algorithms for the Second-Order Cone Program Based on the MZ-Family of Directions Technical report, The School of ISyE, Georgia Insitute of Technology, Atlanta, GA 30332, USA, May 1998. submitted

[MZ98] Monteiro, R.D.C., Zhang, Y.: A unified analysis for a class of path-following primal-dual interior-point algorithms for semidefinite programming. Math. Programming **81**, 281–299 (1998)

[NN94] Nesterov, Y., Nemirovski, A.: Interior Point Polynomial Methods in Convex Programming: Theory and Applications. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1994

[NS96] Nemirovski, A., Scheinberg, K.: Extension of Karmarkar's algorithm onto convex quadratically constrained quadratic programming. Math. Programming **72**, 273–289 (1996)

[NT95] Nesterov, Y.E., Todd, M.J.: Primal-Dual Interior-Point Methods for Self-Scaled Cones. Submitted, 1995

[NT97] Nesterov, Y.E., Todd, M.J.: Self-scaled barriers and interior-point methods for convex programming. Math. of Oper. Res. **22**, 1–42 (1997)

[SA01] Schmieta, S.H., Alizadeh, F.: Associative and Jordan Algebras, and Polynomial Time Interior-Point Algorithms for Symmetric Cones. Math. of Oper. Res. **26**(3), 543–564 (2001)

[Tse96] Tseng, P.: Search directions and convergence analysis of some infeasible path-following methods for the monotone semi-definite LCP. Technical Report, Department of Mathemaics, University of Washington, Seatle, WA 98195, 1996

[Tsu97] Tsuchiya, T.: A Polynomial Primal-Dual Path-Following Algorithm for Second-Order Cone Programming. Technical Report No. 649, The Institute of Statistical Mathematics, Tokyo, Japan, 1997

[Tsu98] Tsuchiya, T.: A convergence analysis of the scaling-invariant primal-dual path-following algorithms for second-order cone programming. Technical Report No 664, The Institute of Statistical Mathematics, Tokyo, Japan, 1998

[TTT98] Todd, M.J., Toh, K.C., Tütüncü, R.H.: On the Nesterov-Todd direction in semidefinite programming. SIAM J. Optim. **8**, 769–796 (1998)

[Zha98] Zhang, Y.: On extending primal-dual interior-point algorithms from linear programming to semidefinite programming. SIAM J. Optim. **8**, 356–386 (1998)