J Optim Theory Appl (2013) 158:816–858
DOI 10.1007/s10957-013-0278-8

A Full Nesterov–Todd Step Infeasible Interior-Point Method for Second-Order Cone Optimization

M. Zangiabadi · G. Gu · C. Roos

Received: 8 January 2012 / Accepted: 4 February 2013 / Published online: 15 February 2013

Abstract After a brief introduction to Jordan algebras, we present a primal–dual interior-point algorithm for second-order conic optimization that uses full Nesterov–Todd steps; no line searches are required. The number of iterations of the algorithm coincides with the currently best iteration bound for second-order conic optimization. We also generalize an infeasible interior-point method for linear optimization to second-order conic optimization. As usual for infeasible interior-point methods, the starting point depends on a positive number. The algorithm either finds a solution in a finite number of iterations or determines that the primal–dual problem pair has no optimal solution with vanishing duality gap.

Keywords Feasible interior-point method · Infeasible interior-point method · Second-order conic optimization · Jordan algebra · Polynomial complexity

Communicated by Anil Rao.

M. Zangiabadi (✉)
Department of Applied Mathematics, Faculty of Mathematical Sciences, Shahrekord University, P.O. Box 115, Shahrekord, Iran
e-mail: Zangiabadi-m@sci.sku.ac.ir

G. Gu
Department of Mathematics, Nanjing University, Nanjing, China
e-mail: ggu@nju.edu.cn

C. Roos
Department of Electrical Engineering, Mathematics and Computer Science,
Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands
e-mail: C.Roos@tudelft.nl

1 Introduction

Second-order conic optimization (SOCO) problems are convex optimization (CO) problems that minimize a linear objective function over the intersection of an affine linear manifold and the Cartesian product of a finite number of second-order (or Lorentz or ice-cream) cones. SOCO problems are nonlinear and convex problems, which include linear optimization (LO) problems, convex quadratic optimization problems and quadratically constrained convex quadratic optimization problems as special cases, and arise in many engineering problems [1–3].

On the other hand, SOCO problems are essentially a specific case of Semidefinite Optimization (SDO) problems. Therefore SOCO problems can be solved via the algorithms for SDO problems. Several algorithms are presented in [4–7] for solving SDO problems. However, it has been pointed out [8] that an interior-point method (IPM) that solves the SOCO problem directly has much better complexity than an IPM applied to the semidefinite formulation of the SOCO problem.

Several authors have discussed IPMs for SOCO. Nesterov and Todd [9, 10] considered linear cone optimization problems in which the cone is *self-scaled*. They presented a primal–dual IPM for optimization over such cones. It has become clear later that self-scaled cones are precisely the cones of squares in Euclidean Jordan algebras. Adler and Alizadeh [11] studied the relationship between SDO and SOCO problems and presented a unified approach to these problems. Alizadeh and Goldfarb [12] and Schmieta and Alizadeh [13, 14] showed that Euclidean Jordan algebras underly the analysis of IPMs for optimization over symmetric cones. Faybusovich [15] used Euclidean Jordan Algebras to analyze when the search directions in the MZ-family are well-defined.

Peng et al. [16, 17] presented primal–dual feasible IPMs by using self-regular proximity functions for LO, SDO and SOCO. They obtained the complexity bounds for small-update and large-update methods, which are currently the best known iteration bounds for SOCO problems. Recently, Bai et al. [18] designed a primal–dual feasible IPM for SOCO problems based on a kernel function. They obtained the same complexity bounds as in [17].

In so-called *feasible* IPMs, it is assumed that the starting point is feasible and lies in the interior of the cone. Such a starting point is called strictly feasible. All the points generated by feasible IPMs are also strictly feasible. In practice, however, it is sometimes difficult to obtain an initial strictly feasible point. Infeasible IPMs (IIPMs) do not require that the starting point be feasible, but only that it be in the interior of the cone. IIPMs are used in most practical implementations. Global convergence of a primal–dual IIPM for LO was first established by Kojima et al. [19]. Subsequently, Zhang [20], Mizuno [21] and Potra [22, 23] presented polynomial iteration complexity results for variants of this algorithm. Later, Zhang [24] extended it to SDO. Rangarajan [25] established polynomial-time convergence of IIPMs for conic programs over symmetric cones using a wide neighborhood of the central path. Recently, Roos [26] established a new IIPM which uses full Newton steps. Later, Mansouri et al. [27] generalized it to SDO.

The aim of this paper is to generalize the IIPM for LO of Roos to SOCO. Since its analysis requires a quadratic convergence result for the feasible case, we first present a primal–dual (feasible) IPM with full NT-steps for SOCO and its analysis. To our knowledge, this is the first time that a full NT-step IPM for SOCO is considered. We use the Nesteorv–Todd (NT) direction. We obtain the same complexity bound as in [17, 18], which is the currently best bound. Then we extend Roos's IIPM for LO to SOCO. We prove that the complexity bound of our IIPM coincides with the currently best iteration bound for SOCO.

The paper is organized as follows. In Sect. 2, we briefly review some properties of the second-order cone and its associated Euclidean Jordan algebra, focusing on what is needed in the rest of the paper. We derive some new inequalities that are crucial for the analysis of our algorithms. Then, in Sect. 3, we present a feasible IPM for SOCO, and in Sect. 4 our IIPM. Section 5 contains some conclusions and topics for further research.

2 Preliminaries

Mathematically, a typical second-order cone in $\mathbb{R}^n$ has the form

$$\mathcal{L} = \left\{ (x_1, x_2; \dots; x_n) \in \mathbb{R}^n : x_1^2 \geq \sum_{i=2}^{n} x_i^2, x_1 \geq 0 \right\}, \tag{1}$$

where $n \geq 2$ is some natural number.

Let $\mathcal{K} \subseteq \mathbb{R}^n$ be the Cartesian product of several second-order cones, i.e.,

$$\mathcal{K} = \mathcal{L}^1 \times \mathcal{L}^2 \times \cdots \times \mathcal{L}^N, \tag{2}$$

where $\mathcal{L}^j \subseteq \mathbb{R}^{n_j}$ for each j, $j = 1, 2, \dots, N$. A second-order conic optimization (SOCO) problem has the form

$$\min\left\{ c^T x : Ax = b, x \in \mathcal{K} \right\}, \tag{3}$$

where $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$, and $n = \sum_{j=1}^{N} n_j$. Without loss of generality, we assume that A has full row rank, i.e., $\operatorname{rank}(A) = m$. Due to the fact that $\mathcal{K}$ is self-dual, the dual problem of (3) is given by

$$\max\left\{ b^T y : A^T y + s = c, s \in \mathcal{K} \right\}. \tag{4}$$

Some notations used throughout the paper are as follows. The superscript ^{T} is used to denote the transpose of a vector or matrix. $\mathbb{R}^n$, $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$ denote the set of real vectors with n components, the set of nonnegative vectors and the set of positive vectors, respectively. We follow the convention of some high level programming languages, such as MATLAB, and use ";" for adjoining vectors in a column. Thus for instance for column vectors x, y and z we have

$$(x; y; z) = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Superscripted vectors such as x^j usually represent the jth block of x. It should be noted that sometimes the notation x^j refers to the jth power of x. The meaning is always clear from the context. $\mathbb{R}^{m\times n}$ is the space of all $m \times n$ matrices. $\mathbf{S}^n$, $\mathbf{S}^n_+$ and $\mathbf{S}^n_{++}$ denote the set of symmetric, symmetric positive semidefinite and symmetric positive definite $n \times n$ matrices, respectively. For any symmetric matrix A, $\lambda_{\min}(A)$ ($\lambda_{\max}(A)$) denotes the minimal (maximal) eigenvalue of A. As usual, $\|\cdot\|$ denotes the 2-norm for vectors and matrices. We denote the trace of a matrix as $\mathrm{Tr}(\cdot)$ and the trace of a vector as $\mathrm{tr}(\cdot)$. The Löwner partial ordering $\succeq_{\mathcal{K}}$ of $\mathbb{R}^n$ defined by a cone $\mathcal{K}$ is defined by $x \succeq_{\mathcal{K}} s$ if $x - s \in \mathcal{K}$. The interior of $\mathcal{K}$ is denoted as $\mathcal{K}_+$. We write $x \succ_{\mathcal{K}} s$ if $x - s \in \mathcal{K}_+$. $\mathcal{P}$ and $\mathcal{D}$ denote the feasible sets of the primal and the dual problem, respectively. In this paper, we assume that both the primal problem and its dual are feasible. Finally, E_n denotes the $n \times n$ identity matrix.

2.1 Euclidean Jordan Algebras

We recall certain basic notions and well-known facts concerning Jordan algebras. For omitted proofs the reader is refereed to the given references and also to [28–30].

Definition 2.1 A map $h : \mathcal{J} \times \mathcal{J} \longmapsto \mathcal{J}$, $\mathcal{J}$ is an n-dimensional vector space over $\mathbb{R}$, is called bilinear iff for all $x, y, z \in \mathcal{J}$ and $\alpha, \beta \in \mathbb{R}$:

(i) $h(\alpha x + \beta y, z) = \alpha h(x, z) + \beta h(y, z)$;
(ii) $h(z, \alpha x + \beta y) = \alpha h(z, x) + \beta h(z, y)$.

Definition 2.2 Let $\mathcal{J}$ be an n-dimensional vector space over $\mathbb{R}$ along with a bilinear map $\circ : (x, y) \mapsto x \circ y \in \mathcal{J}$. Then $(\mathcal{J}, \circ)$ is called a Euclidean Jordan algebra iff for all $x, y \in \mathcal{J}$:

(i) $x \circ y = y \circ x$ (commutativity);
(ii) $x \circ (x^2 \circ y) = x^2 \circ (x \circ y)$, where $x^2 = x \circ x$ (Jordan identity);
(iii) there exists an inner product, denoted by $\langle x, y\rangle$, such that $\langle x \circ y, z\rangle = \langle x, y \circ z\rangle$ (associativity).

We call $x \circ y$ the Jordan product of x and y. In addition, we assume that there is an element $e \in \mathcal{J}$ such that $e \circ x = x \circ e = x$ for all $x \in \mathcal{J}$, which is called the identity element in $\mathcal{J}$. The Jordan product is not necessarily associative, but it is power associative, i.e., the subalgebra generated by a single element $x \in \mathcal{J}$ is associative (Proposition II.1.2 of [28]).

For $x \in \mathcal{J}$, let r be the smallest number such that the set $\{e, x, x^2, \ldots, x^r\}$ is linearly dependent. Then r is called the degree of x and is denoted by $\deg(x)$. The rank of $\mathcal{J}$, denoted by $\mathrm{rank}(\mathcal{J})$, is defined as the maximum of $\deg(x)$ over all $x \in \mathcal{J}$. An element $x \in \mathcal{J}$ is called *regular* iff $\deg(x) = \mathrm{rank}(\mathcal{J})$.

For an element x of degree d, since $\{e, x, x^2, \ldots, x^d\}$ is linearly dependent, there exist real numbers $a_1(x), a_2(x), \ldots, a_d(x)$ such that

$$x^d - a_1(x)x^{d-1} + a_2(x)x^{d-2} + \cdots + (-1)^d a_d(x) = 0,$$

where 0 is the zero vector. Then the polynomial

$$\lambda^d - a_1(x)\lambda^{d-1} + a_2(x)\lambda^{d-2} + \cdots + (-1)^d a_d(x) = 0$$

is called the *minimum polynomial* of x. The minimum polynomial of a regular element x is called the *characteristic polynomial* of x. Since the regular elements are dense in $\mathcal{J}$, by continuity we may extend the polynomials $a_i(x)$ and consequently the characteristic polynomial to all elements of $\mathcal{J}$. The characteristic polynomial is a polynomial of degree r in λ, where r is the rank of $\mathcal{J}$. The roots $\lambda_1, \ldots, \lambda_r$ of the characteristic polynomial of x are called the eigenvalues (spectral values) of x [28].

Definition 2.3 Let $x \in \mathcal{J}$ and $\lambda_1, \ldots, \lambda_r$ be the eigenvalues of x (including multiple eigenvalues). Then,

(i) $\operatorname{tr}(x) := \lambda_1 + \cdots + \lambda_r$, is called the trace of x;
(ii) $\det(x) := \lambda_1 \cdots \lambda_r$ is called the determinant of x.

As known, a nonzero element c of $\mathcal{J}$ is called idempotent iff $c^2 = c$. A complete system of orthogonal idempotent is a set $\{c_1, \ldots, c_k\}$ of idempotents, where

$$c_i \circ c_j = 0 \quad \text{for all } i \neq j \text{ and } c_1 + \cdots + c_k = e.$$

An idempotent is called primitive iff it is not the sum of two other orthogonal idempotents. A complete system of orthogonal primitive idempotents is called a Jordan frame. Jordan frames always contain r primitive idempotents, where r is the rank of $\mathcal{J}$ [28].

The spectral decomposition theorem (Theorem III.1.2 of [28]) of an Euclidean Jordan algebra $\mathcal{J}$ states that for $x \in \mathcal{J}$ there exists a Jordan frame $c_1, \ldots, c_r$ (r is the rank of $\mathcal{J}$) and real numbers $\lambda_1, \ldots, \lambda_r$ (the eigenvalues of x) such that

$$x = \lambda_1 c_1 + \cdots + \lambda_r c_r.$$

Using this, for each $x \in \mathcal{J}$ we can define the following [12]:

square root: $x^{\frac{1}{2}} := \lambda_1^{\frac{1}{2}} c_1 + \cdots + \lambda_r^{\frac{1}{2}} c_r$, whenever all $\lambda_i \geq 0$, and undefined otherwise;
inverse: $x^{-1} := \lambda_1^{-1} c_1 + \cdots + \lambda_r^{-1} c_r$, whenever all $\lambda_i \neq 0$, and undefined otherwise;
square: $x^2 := \lambda_1^2 c_1 + \cdots + \lambda_r^2 c_r$.

Indeed, one has $x^2 = x \circ x$ and $(x^{\frac{1}{2}})^2 = x$. If x^{-1} is defined, then $x \circ x^{-1} = e$, and we call x invertible.[1] Also note that, since e has eigenvalue 1, with multiplicity r, $\operatorname{tr}(e) = r$ and $\det(e) = 1$.

[1] One should be careful here. If $x \circ s = e$, then s is not necessarily equal to x^{-1}. For an example, see [31, Example 2.1.8].

The square x^2 of any $x \in \mathcal{J}$ has nonnegative eigenvalues, and, vice versa, if an element $x \in \mathcal{J}$ has nonnegative eigenvalues, it is a square. The set of squares in $\mathcal{J}$ is given by

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

It is well-known that this set is a convex cone with nonempty interior. It is called the *cone of squares* in $\mathcal{J}$. Below we denote this cone simply as $\mathcal{K}$. We have $x \in \mathcal{K}_+$ iff all eigenvalues of x are positive. For each $x \in \mathcal{J}$, $L(x)$ denotes the linear map of $\mathcal{J}$ defined by

$$L(x)y := x \circ y, \tag{5}$$

and

$$P(x) := 2L(x)^2 - L(x^2), \tag{6}$$

where $L(x)^2 = L(x)L(x)$. The map P is called the quadratic representation of $\mathcal{J}$. Due to Definition 2.2(ii), the maps $L(x)$ and $L(x^2)$ commute. Hence, also $P(x)$ commutes with $L(x)$. One has

$$P(x^2) = P(x)^2, \qquad P(x^{\frac{1}{2}}) = P(x)^{\frac{1}{2}}, \qquad P(x)e = P(x^{\frac{1}{2}})x = x^2, \tag{7}$$

where the relations involving $x^{\frac{1}{2}}$ only hold if $x \in \mathcal{K}$. The automorphism group of (any convex cone) $\mathcal{K}$ is defined by

$$\mathrm{Aut}(\mathcal{K}) = \{g \in \mathrm{Gl}(\mathcal{K}) : g(\mathcal{K}) = \mathcal{K}\},$$

where $\mathrm{Gl}(\mathcal{K})$ is the set of invertible linear maps g from $\mathcal{J}$ into itself. The cone $\mathcal{K}$ is called homogeneous iff $\mathrm{Aut}(\mathcal{K})$ acts transitively on the interior of $\mathcal{K}$, i.e., for all x, y in $\mathcal{K}_+$, there exists $g \in \mathrm{Aut}(\mathcal{K})$ such that $gx = y$. The cone $\mathcal{K}$ is called symmetric iff it is homogeneous and self-dual. The cone of squares $\mathcal{K} = \mathcal{K}_{\mathcal{J}}$ is self-dual (cf. [31, Proposition 2.5.3]). Therefore, the next two results imply that the cone of squares be symmetric.

Proposition 2.1 (Proposition 2.2 in [32]) *For each $x \in \mathcal{K}_+$, $P(x)$ is an automorphism of $\mathcal{K}$ and $P(x)\mathcal{K}_+ = \mathcal{K}_+$. Furthermore, $P(x)$ is positive definite for each $x \in \mathcal{K}_+$.*

Proposition 2.2 (Proposition 2.4 in [32]) *Suppose that $x, s \in \mathcal{K}_+$. Then there exists a unique $w \in \mathcal{K}_+$ such that*

$$P(w)s = x.$$

Moreover,

$$w = P(s^{-\frac{1}{2}})(P(s^{\frac{1}{2}})x)^{\frac{1}{2}} = P(x^{\frac{1}{2}})(P(x^{-\frac{1}{2}})s^{-1})^{-\frac{1}{2}}. \tag{8}$$

The point w is called the scaling point of x and s (in this order). For the equality of the two different expressions for w we refer to, e.g., [32, Theorem 2.8].

We recall a few more results that will be needed in the sequel. Recall that two square matrices A and B (of the same size) are similar iff $B = PAP^{-1}$ for some invertible matrix P; in this case, we write $A \sim B$. If A and B are symmetric, then $A \sim B$ holds if and only if A and B share the same set of eigenvalues (taking multiplicities into account). Analogously, we say that two elements x and y in $\mathcal{J}$ are similar, denoted as $x \sim y$, if and only if x and y share the same set of eigenvalues. For more details we refer to [33].

Proposition 2.3 (Proposition 19 in [14]) *Two elements x and y of an Euclidean Jordan algebra are similar iff $L(x)$ and $L(y)$ are similar.*

Proposition 2.4 (Corollary 20 in [14]) *Let x and y be two elements in $\mathcal{K}_+$. Then x and y are similar iff $P(x)$ and $P(y)$ are similar.*

Proposition 2.5 (Proposition 2.1 in [32]) *The following holds for any x and s in $\mathbb{R}^n$.*

(i) *x is invertible iff $P(x)$ is invertible. In this case*:

$$P(x)x^{-1} = x, \qquad P(x)^{-1} = P\left(x^{-1}\right).$$

(ii) *If x and s are invertible, then $P(x)s$ is invertible and $(P(x)s)^{-1} = P(x^{-1})s^{-1}$.*

(iii) *For any two elements x and s*:

$$P\big(P(x)s\big) = P(x)P(s)P(x).$$

(iv) *If $x, s \in \mathcal{K}_+$, then $P(x^{\frac{1}{2}})s \sim P(s^{\frac{1}{2}})x$.*

The third identity is far from trivial; it is known as the *fundamental formula* for Jordan algebras. Since $P(e) = E_n$, taking $s = e$ it gives

$$P\left(x^2\right) = P(x)^2.$$

The fourth item follows from the *fundamental formula*. The proof is simple. It also uses Proposition 2.4 and goes as follows:

$$P\left(P\left(x^{\frac{1}{2}}\right)s\right) = P\left(x^{\frac{1}{2}}\right)P(s)P\left(x^{\frac{1}{2}}\right) \sim P(x)P(s) \sim P\left(s^{\frac{1}{2}}\right)P(x)P\left(s^{\frac{1}{2}}\right) = P\left(P\left(s^{\frac{1}{2}}\right)x\right).$$

A for our goal very important generalization is the following result. Because of its importance, we include the proof.

Lemma 2.1 (Proposition 21 in [14]) *Let $x, s, p \in \mathcal{K}_+$. Defining $\tilde{x} = P(p)x$ and $\tilde{s} = P(p^{-1})s$, one has*

$$P\left(\tilde{x}^{\frac{1}{2}}\right)\tilde{s} \sim P\left(x^{\frac{1}{2}}\right)s.$$

Proof Since $P(P(x^{\frac{1}{2}})s) \sim P(x)P(s)$, and similarly, $P(P(\tilde{x}^{\frac{1}{2}})\tilde{s}) \sim P(\tilde{x})P(\tilde{s})$, it suffices to show that $P(\tilde{x})P(\tilde{s}) \sim P(x)P(s)$. Using the fundamental formula we obtain

$$\begin{aligned} P(\tilde{x})P(\tilde{s}) &= P\big(P(p)x\big)P\big(P\big(p^{-1}\big)s\big) = P(p)P(x)P(p)P\big(p^{-1}\big)P(s)P\big(p^{-1}\big) \\ &= P(p)P(x)P(s)P\big(p^{-1}\big). \end{aligned}$$

The last matrix is similar to $P(x)P(s)$. Hence the proof is complete. □

The next lemma depends on Proposition 2.5, especially on part (ii) and the fundamental formula in part (iii).

Lemma 2.2 (Proposition 3.2.4 in [31]) *Let $x, s \in \mathcal{K}_+$. If w is the scaling point of x and s, then*

$$\big(P\big(x^{\frac{1}{2}}\big)s\big)^{\frac{1}{2}} \sim P(w)^{\frac{1}{2}}s.$$

2.2 Algebraic Properties of Second-Order Cones

In this section, we briefly review some algebraic properties of the second-order cone $\mathcal{L}$ as defined by (1) and its associated Euclidean Jordan algebra. For more details and proofs we refer to, e.g., [13, 17, 18, 34]. To our knowledge, Lemmas 2.3(iii), 2.3(iv) and 2.6 are new. These results play a key role in our analysis.

For $x, s \in \mathbb{R}^n$, we define the bilinear operator $\circ$ as follows:

$$x \circ s := \big(x^T s; x_1 s_2 + s_1 x_2; \ldots; x_1 s_n + s_1 x_n\big) = \big(x^T s; x_1 \bar{s} + s_1 \bar{x}\big),$$

where $\bar{x} = (x_2; \ldots; x_n)$. One easily checks that $(\mathbb{R}^n, \circ)$ is an Euclidean Jordan algebra, with the vector

$$e = (1; 0; \ldots; 0) \in \mathbb{R}^n$$

as identity element. In the sequel, we denote the vector $(x_2; \ldots; x_n)$ briefly as $\bar{x}$. So $x = (x_1; \bar{x})$. One easily verifies that each $x \in \mathbb{R}^n$ satisfies the quadratic equation

$$x^2 - 2x_1 x + \big(x_1^2 - \|\bar{x}\|^2\big)e = 0.$$

This means that $\lambda^2 - 2x_1\lambda + (x_1^2 - \|\bar{x}\|^2) = 0$ is the characteristic polynomial of x. Hence the rank of this Jordan algebra is 2 and the two eigenvalues of x are

$$\lambda_{\max}(x) = x_1 + \|\bar{x}\|, \qquad \lambda_{\min}(x) = x_1 - \|\bar{x}\|. \tag{9}$$

Therefore, the trace and the determinant of $x \in \mathbb{R}^n$ are

$$\begin{aligned} \operatorname{tr}(x) &= \lambda_{\max}(x) + \lambda_{\min}(x) = 2x_1, \\ \det(x) &= \lambda_{\max}(x)\lambda_{\min}(x) = x_1^2 - \|\bar{x}\|^2. \end{aligned}$$

Lemma 2.3 *For all $x, s \in \mathbb{R}^n$ one has*

(i) $\operatorname{tr}(x \circ s) = 2x^T s$;
(ii) $\det(x \circ s) \le \det(x)\det(s)$; *equality holds iff* $\bar{x} = \alpha\bar{s}, \alpha > 0$;
(iii) $|\lambda_{\max}(x)| + |\lambda_{\min}(x)| = 2\max\{|x_1|, \|\bar{x}\|\}$;
(iv) $\|\overline{x \circ s}\| \le \|x\|\|s\|$.

Proof The relation (i) is obvious. For (ii) we refer to (its elementary proof in [17, Lemma 6.2.3]). Turning to the proof of (iii), we write

$$\left|\lambda_{\max}(x)\right| + \left|\lambda_{\min}(x)\right| = \left|x_1 + \|\bar{x}\|\right| + \left|x_1 - \|\bar{x}\|\right|.$$

If $x_1 \ge \|\bar{x}\|$, then the last expression equals $2x_1$, and if $x_1 \le -\|\bar{x}\|$, then it equals $-2x_1$. On the other hand, if $x_1 \in (-\|\bar{x}\|, \|\bar{x}\|)$ then it equals $2\|\bar{x}\|$. Hence (iii) follows. Finally, using the triangle inequality and also $2ab \le a^2 + b^2$, we may write

$$\begin{aligned}
\|\overline{x \circ s}\|^2 &= \|x_1\bar{s} + s_1\bar{x}\|^2 \le \left(\|x_1\bar{s}\| + \|s_1\bar{x}\|\right)^2 \\
&= x_1^2\|\bar{s}\|^2 + s_1^2\|\bar{x}\|^2 + 2|x_1||s_1|\|\bar{x}\|\|\bar{s}\| \\
&\le x_1^2\|\bar{s}\|^2 + s_1^2\|\bar{x}\|^2 + x_1^2 s_1^2 + \|\bar{s}\|^2\|\bar{x}\|^2 \\
&= \left(x_1^2 + \|\bar{x}\|^2\right)\left(s_1^2 + \|\bar{s}\|^2\right) = \|x\|^2\|s\|^2,
\end{aligned}$$

which implies (iv). This completes the proof. □

It is worth pointing out that (ii) does not always hold with equality. This is related to the fact that the second-order cone is not closed under the Jordan product. The spectral decomposition of $x \in \mathbb{R}^n$ is given by

$$x = \lambda_{\max}(x)c_1 + \lambda_{\min}(x)c_2,$$

where the Jordan frame $\{c_1, c_2\}$ is given by

$$c_1 = \frac{1}{2}\left(1; \frac{\bar{x}}{\|\bar{x}\|}\right), \qquad c_2 := \frac{1}{2}\left(1; \frac{-\bar{x}}{\|\bar{x}\|}\right).$$

Here, by convention, $\frac{-\bar{x}}{\|\bar{x}\|} = 0$ if $\bar{x} = 0$. Note that c_1 and c_2 belong to $\mathcal{L}$ (but not to $\mathcal{L}_+$).

Since $x^2 = (\|x\|^2; 2x_1\bar{x})$, one easily understands that $\{c_1, c_2\}$ is also a Jordan frame for x^2. This implies that the matrices $L(x)$ and $L(x^2)$ commute. See, e.g., [14, Theorem 27]. (It also confirms Definition 2.2(ii).) The natural inner product is given by

$$\langle x, s\rangle := \operatorname{tr}(x \circ s) = 2x^T s, \quad x, s \in \mathbb{R}^n.$$

Hence, the norm induced by this inner product, which is denoted as $\|\cdot\|_F$ (cf. [12]), satisfies

$$\|x\|_F = \sqrt{\langle x, x\rangle} = \sqrt{\operatorname{tr}\left(x^2\right)} = \left(\lambda_{\max}(x)^2 + \lambda_{\min}(x)^2\right)^{\frac{1}{2}} = \sqrt{2}\|x\|. \tag{10}$$

We proceed with some simple properties of this inner product and the induced norm.

Lemma 2.4 *Let $x \in \mathbb{R}^n$ and $s \in \mathcal{K}$. Then*

$$\lambda_{\min}(x)\,\mathrm{tr}(s) \leq \mathrm{tr}(x \circ s) \leq \lambda_{\max}(x)\,\mathrm{tr}(s).$$

Proof For any $x \in \mathbb{R}^n$ we have $\lambda_{\max}(x)e - x \in \mathcal{K}$. Since also $s \in \mathcal{K}$, it follows that $\mathrm{tr}((\lambda_{\max}(x)e - x) \circ s) \geq 0$. Hence the second inequality in the lemma follows by writing

$$\mathrm{tr}(x \circ s) \leq \mathrm{tr}\big(\lambda_{\max}(x)e \circ s\big) = \lambda_{\max}(x)\mathrm{tr}(e \circ s) = \lambda_{\max}(x)\mathrm{tr}(s),$$

the proof of the first inequality goes in the same way. □

Lemma 2.5 *For all $x, s \in \mathbb{R}^n$ one has*

(i) $\|x^2\|_F \leq \|x\|_F^2$; *equality holds iff* $|x_1| = \|\bar{x}\|$;
(ii) $\mathrm{tr}((x \circ s)^2) \leq \mathrm{tr}(x^2 \circ s^2)$;
(iii) $\|x \circ s\|_F^2 \leq \lambda_{\max}(x^2)\|s\|_F^2 \leq \|x\|_F^2\|s\|_F^2$.

Proof Using $x^2 = (\|x\|^2; 2x_1\bar{x})$ we may write, also using $2ab \leq a^2 + b^2$,

$$\|x^2\|_F^2 = 2\big(\|x\|^4 + (2x_1\|\bar{x}\|)^2\big) \leq 2\big(\|x\|^4 + \big(x_1^2 + \|\bar{x}\|^2\big)^2\big) = 4\|x\|^4 = \|x\|_F^4,$$

which implies (i). The proof of (ii) uses that

$$\mathrm{tr}\big((x \circ s)^2\big) = 2\|x \circ s\|^2 = 2\big(x^T s\big)^2 + 2\|x_1\bar{s} + s_1\bar{x}\|^2.$$

Proceeding in a similar way as in the proof of Lemma 2.3(iv), one can obtain part (ii) of the current lemma. Finally, using part (ii) we may write

$$\|x \circ s\|_F^2 = \mathrm{tr}\big((x \circ s)^2\big) \leq \mathrm{tr}\big(x^2 \circ s^2\big).$$

Due to Lemma 2.4 and part (i), this implies

$$\|x \circ s\|_F^2 \leq \lambda_{\max}\big(x^2\big)\mathrm{tr}\big(s^2\big) = \lambda_{\max}\big(x^2\big)\|s\|_F^2,$$

which is the first inequality in (iii). The second inequality in (iii) follows by applying (10). This completes the proof. □

As we mentioned before, the Jordan product is not associative. However, remarkably enough, the trace function is associative (which confirms Definition 2.2(iii)). We have (cf. Proposition II.4.3 in [28])

$$\mathrm{tr}\big((x \circ y) \circ z\big) = \mathrm{tr}\big(x \circ (y \circ z)\big). \tag{11}$$

An important consequence of the associativity of the trace function is that $L(x)$ is self-adjoint with respect to the above inner product:

$$\big\langle L(x)y, z\big\rangle = \mathrm{tr}\big((x \circ y) \circ z\big) = \mathrm{tr}\big((y \circ x) \circ z\big) = \mathrm{tr}\big(y \circ (x \circ z)\big) = \big\langle y, L(x)z\big\rangle.$$

Since $P(x)$ is a linear combination of the self-adjoint matrices $L(x)^2$ and $L(x^2)$, $P(x)$ is self-adjoint as well (cf. [25, p. 1214]). It easily can be verified that the cone of squares of the current Jordan algebra is given by (1). For each $x \in \mathbb{R}^n$, the matrices of $L(x)$ and $P(x)$ with respect to the natural basis will be denoted with the same notations as the maps themselves. As a consequence, we have

$$L(x) = \begin{bmatrix} x_1 & \bar{x}^T \\ \bar{x} & x_1 \mathbf{E}_{n-1} \end{bmatrix}, \qquad P(x) = \begin{bmatrix} \|x\|^2 & 2x_1\bar{x}^T \\ 2x_1\bar{x} & \det(x)\mathbf{E}_{n-1} + 2\bar{x}\bar{x}^T \end{bmatrix}.$$

The eigenvalues of $L(x)$ are $\lambda_{\max}(x)$ and $\lambda_{\min}(x)$, both with multiplicity 1, and x_1, with multiplicity $n-2$, and those of $P(x)$ are $\lambda_{\max}(x)^2$ and $\lambda_{\min}(x)^2$, both with multiplicity 1, and $\det(x)$, with multiplicity $n-2$ (cf. [12, Theorem 3]).[2] This implies the following two important facts.

(i) $x \in \mathcal{L}$ ($x \in \mathcal{L}_+$) if and only if $L(x)$ is positive semidefinite (positive definite);
(ii) if $x \in \mathcal{L}$ then $P(x)$ is positive semidefinite; if $x \in \mathcal{L}_+$ then $P(x)$ is positive definite.

The first property implies that SOCO be a special case of semidefinite optimization (SDO). We conclude this section with a result which is new and crucial for the purpose of this paper.

Lemma 2.6 *Let $x, s \in \mathcal{L}_+$, $u = P(x)^{\frac{1}{2}} s$ and $z = x \circ s \in \mathcal{L}_+$. Then we have*

$$\left\| u^{\frac{1}{2}} - u^{-\frac{1}{2}} \right\|_F \le \left\| z^{\frac{1}{2}} - z^{-\frac{1}{2}} \right\|_F.$$

Proof Using that $L(x)$ and $P(x)$ are self-adjoint and also Proposition 2.5 we may write:

$$\begin{aligned}
\left\| u^{\frac{1}{2}} - u^{-\frac{1}{2}} \right\|_F^2 &= \left\langle u^{\frac{1}{2}} - u^{-\frac{1}{2}}, u^{\frac{1}{2}} - u^{-\frac{1}{2}} \right\rangle \\
&= \left\langle u^{\frac{1}{2}}, u^{\frac{1}{2}} \right\rangle - 2\left\langle u^{-\frac{1}{2}}, u^{\frac{1}{2}} \right\rangle + \left\langle u^{-\frac{1}{2}}, u^{-\frac{1}{2}} \right\rangle \\
&= \langle u, e\rangle - 2\langle e, e\rangle + \left\langle u^{-1}, e \right\rangle \\
&= \left\langle P(x)^{\frac{1}{2}} s, e \right\rangle - 2\langle e, e\rangle + \left\langle P(x)^{-\frac{1}{2}} s^{-1}, e \right\rangle \\
&= \left\langle s, P(x)^{\frac{1}{2}} e \right\rangle - 2\langle e, e\rangle + \left\langle s^{-1}, P(x)^{-\frac{1}{2}} e \right\rangle \\
&= \langle s, x\rangle - 2\langle e, e\rangle + \left\langle s^{-1}, x^{-1} \right\rangle.
\end{aligned}$$

The last equality is due to (7). With $z = x \circ s \in \mathcal{L}_+$, we derive in a similar way the following:

$$\begin{aligned}
\left\| z^{\frac{1}{2}} - z^{-\frac{1}{2}} \right\|_F^2 &= \langle z, e\rangle - 2\langle e, e\rangle + \left\langle z^{-1}, e \right\rangle \\
&= \langle x \circ s, e\rangle - 2\langle e, e\rangle + \left\langle (x \circ s)^{-1}, e \right\rangle \\
&= \langle s, x\rangle - 2\langle e, e\rangle + \left\langle (x \circ s)^{-1}, e \right\rangle.
\end{aligned}$$

[2]Observe that this means that the determinant of $P(x)$, being the product of its eigenvalues, equals $\det(x)^n$.

So the inequality in the lemma will follow if

$$\mathrm{tr}\big(x^{-1}\circ s^{-1}\big)=\big\langle x^{-1},s^{-1}\big\rangle\leq\big\langle(x\circ s)^{-1},e\big\rangle=\mathrm{tr}\big((x\circ s)^{-1}\big).$$

Since $x^{-1}=(x_1;-\bar{x})/\det(x)$, for each $x\in\mathcal{L}_+$, one has

$$\mathrm{tr}\big(x^{-1}\circ s^{-1}\big)=\frac{2(x_1s_1+\bar{x}^T\bar{s})}{\det(x)\det(s)}=\frac{2x^Ts}{\det(x)\det(s)},\qquad \mathrm{tr}\big((x\circ s)^{-1}\big)=\frac{2x^Ts}{\det(x\circ s)}.$$

The hypothesis in lemma implies that $\det(x)$, $\det(s)$ and $\det(x\circ s)$ are positive. Also $x^Ts>0$, because $z=x\circ s\in\mathcal{L}_+$. Hence the inequality in the lemma will hold if

$$\det(x\circ s)\leq\det(x)\det(s).$$

But this is true, by Lemma 2.3(ii). Hence the proof is complete. □

2.3 Rescaling the Cone $\mathcal{L}$

When defining the search direction in our algorithm, we need a rescaling of the space in which the cone lives. Let $x,s\in\mathcal{L}_+$. Since $\lambda_{\min}(x)$ and $\lambda_{\min}(s)$ are positive, x^{-1} and s^{-1} exist. By Proposition 2.2, there exists a unique $w\in\mathcal{L}_+$ such that $P(w)s=x$, with w as given in (8). Due to Proposition 2.1, $P(w)$ is an automorphism. As a consequence, there exists $\tilde{v}\in\mathcal{L}_+$ such that

$$\tilde{v}=P(w)^{-\frac{1}{2}}x=P(w)^{\frac{1}{2}}s. \tag{12}$$

We call this Nesterov–Todd (NT)-scaling of $\mathbb{R}^n$, after the inventors. In the following lemma, we recall several properties of the NT-scaling scheme. Because of their importance, we include their short proofs.

Lemma 2.7 (Cf. Proposition 6.3.3 in [17]) *Let $W=P(w^{\frac{1}{2}})$ for some $w\in\mathcal{K}_+$. Then the following holds for any two vectors $x,s\in\mathbb{R}^n$.*

(i) $\mathrm{tr}(Ws\circ W^{-1}x)=\mathrm{tr}(s\circ x)$;
(ii) $\det(Ws)=\det(w)\det(s)$, $\det(W^{-1}x)=\det(w^{-1})\det(x)$;
(iii) *if w is the scaling point of $x,s\in\mathcal{K}_+$ then* $\det(Ws\circ W^{-1}x)=\det(s)\det(x)$.

Proof The proof of (i) is a direct consequence of the fact that W is self-adjoint:

$$\mathrm{tr}\big(Ws\circ W^{-1}x\big)=\big\langle Ws,W^{-1}x\big\rangle=\big\langle s,WW^{-1}x\big\rangle=\langle s,x\rangle=\mathrm{tr}(s\circ x).$$

For the proof of (ii) we need the matrix

$$R=\mathrm{diag}(1,-1,\dots,-1)\in\mathbb{R}^{n\times n}. \tag{13}$$

Obviously, $R^2=E_n$, where E_n denotes the identity matrix of size $n\times n$. Moreover, $\det(s)=s^TRs$, for any s. It is well known that $WRW=\det(w)R$.[3] Hence we may

[3]By Proposition 3 in the appendix of [34], W is an automorphism iff $WRW=\lambda R$ for some $\lambda>0$. This condition implies $(WR)^2=\lambda E_n$. Since $\det(R)^2=1$, it follows that $\lambda^n=\det(W^2)=\det(P(w))=\det(w)^n$, whence $\lambda=\det(w)$.

write

$$\det(Ws) = (Ws)^T R(Ws) = s^T WRWs = \det(w) s^T Rs = \det(w)\det(s).$$

In a similar way, we can prove $\det(W^{-1}x) = \det(w^{-1})\det(x)$. Finally, for proving (iii), we say that, if w is the scaling point of x and s, then $Ws = W^{-1}x$. Hence, using Lemma 2.3 and part (ii) of the current lemma, we write

$$\begin{aligned}\det\bigl(Ws \circ W^{-1}x\bigr) &= \det(Ws)\det\bigl(W^{-1}x\bigr) = \det(w)\det(s)\det\bigl(w^{-1}\bigr)\det(x)\\ &= \det(s)\det(x),\end{aligned}$$

where we used that $\det(w)\det(w^{-1}) = 1$. Hence the proof is complete. □

2.4 Rescaling the Cone $\mathcal{K}$

In this section, we show how the definitions and properties in the previous sections can be adapted to the case where $N > 1$, when the cone underlying the given problems (3) and (4) is the Cartesian product of N cones $\mathcal{L}^j$, as given in (2).

First, we partition any vector $x \in \mathbb{R}^n$ according to the dimensions of the successive cones $\mathcal{L}^j$, so

$$x = \bigl(x^1; \ldots; x^N\bigr), \quad x^j \in \mathbb{R}^{n_j},$$

and we define the algebra $(\mathbb{R}^n, \circ)$ as a direct product of the Jordan algebras $(\mathbb{R}^{n_j}, \circ)$, by defining

$$x \circ s := \bigl(x^1 \circ s^1; \ldots; x^N \circ s^N\bigr).$$

Obviously, if $e^j \in \mathcal{L}^j$ is the unit element in the Jordan algebra for the jth cone, then the vector

$$e = \bigl(e^1; \ldots; e^N\bigr) \tag{14}$$

is the unit element in $(\mathbb{R}^n, \circ)$. Moreover, $\mathrm{tr}(e) = 2N$, which is the rank of $(\mathbb{R}^n, \circ)$. One easily verifies that $L(\cdot)$ and $P(\cdot)$ are given by (cf. [12]):

$$\begin{aligned}L(x) &:= \mathrm{diag}\bigl(L\bigl(x^1\bigr), \ldots, L\bigl(x^N\bigr)\bigr),\\ P(x) &:= \mathrm{diag}\bigl(P\bigl(x^1\bigr), \ldots, P\bigl(x^N\bigr)\bigr).\end{aligned}$$

The NT-scaling scheme for the general case can be obtained as follows.

For $x^j, s^j \in \mathcal{L}^j_+$, let w^j be the scaling point in $\mathcal{L}^j$. Then

$$P\bigl(w^j\bigr)^{-\frac{1}{2}} x^j = P\bigl(w^j\bigr)^{\frac{1}{2}} s^j, \quad 1 \le j \le N.$$

The scaling point of x and s in $\mathcal{K}$ is then defined by

$$w := \bigl(w^1; \ldots; w^N\bigr).$$

Since $P(w^j)$ is symmetric and positive definite for each j, the matrix

$$P(w) := \operatorname{diag}\big(P\big(w^1\big), \dots, P\big(w^N\big)\big)$$

is symmetric and positive definite as well and represents an automorphism of $\mathcal{K}$ such that $P(w)s = x$. Therefore $P(w)$ can be used to rescale x and s to the same vector

$$v := \big(v^1; \dots; v^N\big), \tag{15}$$

according to (12). Since $L(x) := \operatorname{diag}(L(x^1), \dots, L(x^N))$, one easily gets

$$\lambda_{\max}(x) = \lambda_{\max}\big(L(x)\big) = \max\big\{\lambda_{\max}\big(x^j\big) : 1 \le j \le N\big\}, \tag{16}$$

$$\lambda_{\min}(x) = \lambda_{\min}\big(L(x)\big) = \min\big\{\lambda_{\min}\big(x^j\big) : 1 \le j \le N\big\}. \tag{17}$$

Furthermore,

$$\operatorname{tr}(x) = \sum_{j=1}^{N} \operatorname{tr}\big(x^j\big) = \sum_{j=1}^{N} \big[\lambda_{\min}\big(x^j\big) + \lambda_{\max}\big(x^j\big)\big], \tag{18}$$

$$\|x\|_F^2 = \sum_{j=1}^{N} \big\|x^j\big\|_F^2 = \sum_{j=1}^{N} \big[\lambda_{\min}\big(x^j\big)^2 + \lambda_{\max}\big(x^j\big)^2\big], \tag{19}$$

$$\det(x) = \prod_{j=1}^{N} \det\big(x^j\big) = \prod_{j=1}^{N} \lambda_{\min}\big(x^j\big)\lambda_{\max}\big(x^j\big). \tag{20}$$

3 A Feasible Full NT-step Algorithm

In this section, we present a full NT-step feasible IPM and its analysis. The results of this section will be used later on, when dealing with the purpose of this paper, a full step infeasible IPM.

3.1 The Central Path for SOCO

We assume that both (3) and (4) satisfy the interior-point condition (IPC), i.e., there exists (x^0, s^0, y^0) such that

$$Ax^0 = b, \quad x^0 \in \mathcal{K}_+, \qquad A^T y^0 + s^0 = c, \quad s^0 \in \mathcal{K}_+.$$

It is well known that the IPC implies that (3) and (4) have optimal solutions with duality gap zero [12]. Under the IPC assumption, finding optimal solutions of (3) and (4), is therefore equivalent to solving the following system (see also [15]):

$$\begin{aligned} Ax &= b, \quad x \in \mathcal{K}, \\ A^T y + s &= c, \quad s \in \mathcal{K}, \\ x \circ s &= 0. \end{aligned} \tag{21}$$

The basic idea of primal–dual IPMs is to replace the third equation in (21), the so-called complementary condition for (3) and (4), by the parameterized equation $x \circ s = \mu e$, with $\mu > 0$. Thus we consider the system

$$\begin{aligned} Ax &= b, \quad x \in \mathcal{K}, \\ A^T y + s &= c, \quad s \in \mathcal{K}, \\ x \circ s &= \mu e. \end{aligned} \tag{22}$$

For each $\mu > 0$ the parameterized system (22) has a unique solution $x(\mu)$ and $(y(\mu), s(\mu))$. We call $x(\mu)$ and $(y(\mu), s(\mu))$ the μ-center of (3) and (4), respectively. Note that at the μ-center we have

$$x(\mu)^T s(\mu) = \frac{1}{2}\operatorname{tr}\bigl(x(\mu) \circ s(\mu)\bigr) = \frac{1}{2}\operatorname{tr}(\mu e) = \frac{\mu}{2}\operatorname{tr}(e) = \mu N,$$

where $\operatorname{tr}(e) = 2N$. The set of μ-centers (with μ running through all positive real numbers) gives a homotopy path, which is called the central path of (3) and (4) [15]. If $\mu \to 0$ then the limit of the central path exists and, since the limit points satisfy the complementarity condition, the limit yields optimal solutions for (3) and (4) [15].

3.2 The Nesterov–Todd Search Direction

The natural way to define a search direction is to follow the Newton approach and to linearize the third equation in (22), which leads to the system

$$\begin{aligned} A\Delta x &= 0, \\ A^T \Delta y + \Delta s &= 0, \\ x \circ \Delta s + s \circ \Delta x &= \mu e - x \circ s. \end{aligned} \tag{23}$$

This system not always has a solution, due to the fact that x and s do not operator commute in general (i.e., $L(x)L(s) \neq L(s)L(x)$). For an example of this phenomenon we refer to [17, Sect. 6.3.1]. It is now well known that this difficulty can be solved by applying a *scaling scheme*. This goes as follows. Let $u \in \mathcal{K}_+$. Then we have

$$x \circ s = \mu e \quad \Leftrightarrow \quad P(u)x \circ P\bigl(u^{-1}\bigr)s = \mu e.$$

Since $x, s \in \mathcal{K}_+$, this is an easy consequence of Proposition 2.5(ii), as becomes clear when using that $x \circ s = \mu e$ holds if and only if $x = \mu s^{-1}$ (cf. Lemma 28 in [14]). Now, replacing the third equation in (23) by $P(u)x \circ P(u^{-1})s = \mu e$, and then applying Newton's method, we obtain the system

$$\begin{aligned} A\Delta x &= 0, \\ A^T \Delta y + \Delta s &= 0, \\ P(u)x \circ P\bigl(u^{-1}\bigr)\Delta s + P\bigl(u^{-1}\bigr)s \circ P(u)\Delta x &= \mu e - P(u)x \circ P\bigl(u^{-1}\bigr)s. \end{aligned} \tag{24}$$

By choosing u appropriately, this system can be used to define search directions. In the literature, the following choices are well known: $u = s^{\frac{1}{2}}$, $u = x^{-\frac{1}{2}}$ and $u = w^{-\frac{1}{2}}$, where w is the NT-scaling point of x and s. The first two choices lead to the so-called xs-direction and sx-direction, respectively [12, 14]. In this paper, we focus on the third choice, which gives rise to the NT-direction. For that case we define

$$v := \frac{P(w)^{-\frac{1}{2}} x}{\sqrt{\mu}} \qquad \left[= \frac{P(w)^{\frac{1}{2}} s}{\sqrt{\mu}}\right], \tag{25}$$

and

$$\bar{A} := \sqrt{\mu} A P(w)^{\frac{1}{2}}, \qquad d_x := \frac{P(w)^{-\frac{1}{2}} \Delta x}{\sqrt{\mu}}, \qquad d_s := \frac{P(w)^{\frac{1}{2}} \Delta s}{\sqrt{\mu}}. \tag{26}$$

This enables us to rewrite the system (24) as follows:

$$\bar{A} d_x = 0, \tag{27}$$

$$\bar{A}^T \frac{\Delta y}{\mu} + d_s = 0, \tag{28}$$

$$d_s + d_x = v^{-1} - v. \tag{29}$$

That substitution of (25) and (26) into the first two equations of (24) yields (27) and (28) is easy to verify. It is less obvious that the third equation in (24) yields (29). After the substitution we get, after dividing both sides by μ, $v \circ (d_s + d_x) = e - v^2$. This can be written as $L(v)(d_s + d_x) = e - v^2$. After multiplying of both sides from the left with $L(v)^{-1}$, while using $L(v)^{-1} e = v^{-1}$ and $L(v)^{-1} v^2 = v$, we obtain (29). It easily follows that the above system has unique solution. Since (27) requires that d_x belongs to the null space of $\bar{A}$, and (28) that d_s belongs to the row space of $\bar{A}$, it follows that system (27)–(29) determines d_x and d_s uniquely as the (mutually orthogonal) components of the vector $v^{-1} - v$ in these two spaces. From (29) and the orthogonality of d_x and d_s we obtain

$$\|d_x\|_F^2 + \|d_s\|_F^2 = \|d_x + d_s\|_F^2 = \left\|v^{-1} - v\right\|_F^2. \tag{30}$$

Therefore the displacements d_x, d_s (and since $\bar{A}$ has full row rank, also Δy) are zero if and only if $v^{-1} - v = 0$. In this case it easily follows that $v = e$, and that this implies that x, y and s coincide with the respective μ-centers.

To get the search directions Δx and Δs in the original, we simply transform the scaled search directions back to the x- and s-space by using (26):

$$\Delta x = \sqrt{\mu} P(w)^{\frac{1}{2}} d_x, \qquad \Delta s = \sqrt{\mu} P(w)^{-\frac{1}{2}} d_s. \tag{31}$$

The new iterates are obtained by taking a full step, as follows.

$$\begin{aligned} x_+ &= x + \Delta x, \\ y_+ &= y + \Delta y, \\ s_+ &= s + \Delta s. \end{aligned} \tag{32}$$

Using definition (25) and Lemma 2.7(i), it readily follows that

$$\mu \operatorname{tr}\bigl(v^2\bigr) = \operatorname{tr}(x \circ s). \tag{33}$$

3.3 Proximity Measure

In the analysis of the algorithm, we need a measure for the distance of the iterates (x, y, s) to the current μ-center $(x(\mu), y(\mu), s(\mu))$. The aim of this section is to present such a measure and to show how it depends on the eigenvalues of the vector v. The proximity measure that we are going to use is defined as follows:

$$\delta(x, s; \mu) \equiv \delta(v) := \frac{1}{2}\bigl\|v^{-1} - v\bigr\|_F = \frac{1}{2}\sqrt{\sum_{j=1}^{N} \bigl\|\bigl(v^j\bigr)^{-1} - v^j\bigr\|_F^2}. \tag{34}$$

In the sequel, we will often use the following relation:

$$4\delta(v)^2 = \bigl\|v - v^{-1}\bigr\|_F^2 = \operatorname{tr}\bigl(v^2\bigr) + \operatorname{tr}\bigl(v^{-2}\bigr) - 2\operatorname{tr}(e), \tag{35}$$

which makes clear that $\delta(v)^2$ depends only on the eigenvalues of v^2 and its inverse.

3.4 The Feasible Algorithm

The full NT-step feasible algorithm is given in Algorithm 1. We show below (cf. Lemma 3.3) that after a full NT-step (targeting at the μ-center), the duality gap $x^T s$ gets its target value $N\mu$. Hence, if the algorithm stops, then the duality gap equals $N\mu/(1-\theta)$, which by then is less than $\varepsilon/(1-\theta)$.

Algorithm 1 PRIMAL–DUAL FEASIBLE IPM

Input:
  accuracy parameter $\varepsilon > 0$;
  barrier update parameter θ, $0 < \theta < 1$;
  threshold parameter $\tau > 0$;
  strictly feasible pair (x^0, s^0) and $\mu^0 > 0$ such that
  ${x^0}^T s^0 = N\mu^0$ and $\delta(x^0, s^0; \mu^0) \le \tau$.
begin
  $x := x^0$; $s := s^0$; $\mu := \mu^0$;
  while $N\mu \ge \varepsilon$
    $(x, y, s) := (x, y, s) + (\Delta x, \Delta y, \Delta s)$;
    $\mu := (1-\theta)\mu$;
  endwhile
end

3.5 Analysis of the Full NT-step

3.5.1 Feasibility of the Full NT-step

Our aim is to find a condition that guarantees feasibility of the iterates after a full NT-step. As before, let $x, s \in \mathcal{K}_+$, $\mu > 0$ and let w be the scaling point of x and s. Using (25), (31) and (32), we obtain

$$x_+ = x + \Delta x = \sqrt{\mu} P(w)^{\frac{1}{2}}(v + d_x), \tag{36}$$

$$s_+ = s + \Delta s = \sqrt{\mu} P(w)^{-\frac{1}{2}}(v + d_s). \tag{37}$$

Since $P(w)^{\frac{1}{2}}$ and its inverse $P(w)^{-\frac{1}{2}}$ are automorphisms of $\mathcal{K}$, x_+ and s_+ will belong to $\mathcal{K}_+$ if and only if $v + d_x$, and $v + d_s$ belong to $\mathcal{K}_+$. For the proof of our main result in this section, which is Lemma 3.2, we need the following lemma.

Lemma 3.1 *If $\delta(v) \le 1$ then $e + d_x \circ d_s \in \mathcal{K}$. Moreover, if $\delta(v) < 1$, then $e + d_x \circ d_s \in \mathcal{K}_+$.*

Proof Since d_x and d_s are orthogonal, Lemma A.2(i) implies that the absolute values of the eigenvalues of $d_x \circ d_s$ do not exceed $\frac{1}{4}\|d_x + d_s\|_F^2$. Since

$$d_x + d_s = v^{-1} - v, \qquad \left\|v^{-1} - v\right\|_F^2 = 4\delta(v)^2,$$

it follows that the absolute values of the eigenvalues of $d_x \circ d_s$ do not exceed $\delta(v)^2$. This implies that $1 - \delta(v)^2$ be a lower bound for the eigenvalues of $e + d_x \circ d_s$. Hence, if $\delta(v) \le 1$ then $e + d_x \circ d_s \in \mathcal{K}$, and if $\delta(v) < 1$, then $e + d_x \circ d_s \in \mathcal{K}_+$. This proves the lemma. $\square$

Lemma 3.2 *The full NT-step is feasible iff $\delta(v) \le 1$ and strictly feasible iff $\delta(v) < 1$.*

Proof We introduce a step length α, $0 \le \alpha \le 1$, and define

$$v_x^\alpha = v + \alpha d_x, \qquad v_s^\alpha = v + \alpha d_s.$$

We then have $v_x^0 = v$, $v_x^1 = v + d_x$ and $v_s^0 = v$, $v_s^1 = v + d_s$. Since $d_x + d_s = v^{-1} - v$, it follows that

$$\begin{aligned} v_x^\alpha \circ v_s^\alpha &= (v + \alpha d_x) \circ (v + \alpha d_s) = v^2 + \alpha v \circ (d_x + d_s) + \alpha^2 d_x \circ d_s \\ &= v^2 + \alpha v \circ \left(v^{-1} - v\right) + \alpha^2 d_x \circ d_s = (1-\alpha)v^2 + \alpha e + \alpha^2 d_x \circ d_s. \end{aligned}$$

Since $\delta(v) \le 1$, Lemma 3.1 implies that $d_x \circ d_s \succeq_{\mathcal{K}} -e$. Substitution gives

$$v_x^\alpha \circ v_s^\alpha \succeq_{\mathcal{K}} (1-\alpha)v^2 + \alpha e - \alpha^2 e = (1-\alpha)\left(v^2 + \alpha e\right).$$

If $0 \le \alpha < 1$, the last vector belongs to $\mathcal{K}_+$. Hence we then have $\det(v_x^\alpha \circ v_s^\alpha) > 0$. By Lemma 2.3(ii), this implies that $\det(v_x^\alpha)\det(v_s^\alpha) > 0$, for each $\alpha \in [0, 1)$. It follows that $\det(v_x^\alpha)$ and $\det(v_s^\alpha)$ do not vanish for $\alpha \in [0, 1)$. Since we have

$\det(v_x^0) = \det(v_s^0) = \det(v) > 0$, by continuity, $\det(v_x^\alpha)$ and $\det(v_s^\alpha)$ stay positive for all $\alpha \in [0, 1)$. Again by continuity, we also find that $\det(v_x^1)$ and $\det(v_s^1)$ are nonnegative. This proves that if $\delta(v) \le 1$, then $v + d_x \in \mathcal{K}$ and $v + d_s \in \mathcal{K}$. If $\delta(v) < 1$ then we have $d_x \circ d_s \succ_{\mathcal{K}} -e$ and the same arguments imply that $\det(v_x^\alpha)\det(v_s^\alpha) > 0$, for each $\alpha \in [0, 1]$, whence $v + d_x \in \mathcal{K}_+$ and $v + d_s \in \mathcal{K}_+$. This proves the lemma. □

An important consequence of the above lemma is

$$\delta(v) < 1 \quad \Rightarrow \quad (v + d_x) \circ (v + d_s) \in \mathcal{K}_+. \tag{38}$$

The next lemma shows that the target duality gap is attained after a full NT-step.

Lemma 3.3 *Let* $(x, s) \in \mathcal{K}$ *and* $\mu > 0$. *Then*

$$x_+^T s_+ = N\mu.$$

Proof Due to (36) and (37), we may write

$$x_+^T s_+ = \left(\sqrt{\mu} P(w)^{\frac{1}{2}}(v + d_x)\right)^T \left(\sqrt{\mu} P(w)^{-\frac{1}{2}}(v + d_s)\right) = \mu (v + d_x)^T (v + d_s).$$

Using the third equation in (28), we obtain

$$\begin{aligned}(v + d_x)^T (v + d_s) &= v^T v + v^T (d_x + d_s) + d_x^T d_s = v^T v + v^T \left(v^{-1} - v\right) + d_x^T d_s \\ &= e^T e + d_x^T d_s.\end{aligned}$$

Since d_x and d_s are orthogonal, and $e^T e = N$, the lemma follows. □

3.5.2 Quadratic Convergence

In this section, we prove quadratic convergence to the target point $(x(\mu), s(\mu))$ when taking full NT-steps. According to (25), the v-vector after the step is given by

$$v_+ := \frac{P(w_+)^{-\frac{1}{2}} x_+}{\sqrt{\mu}} \qquad \left[= \frac{P(w_+)^{\frac{1}{2}} s_+}{\sqrt{\mu}}\right], \tag{39}$$

where w_+ is the scaling point of x_+ and s_+.

Lemma 3.4 (Proposition 5.9.3 in [31]) *One has*

$$v_+ \sim \left(P(v + d_x)^{\frac{1}{2}}(v + d_s)\right)^{\frac{1}{2}}.$$

Proof It readily follows from (39) and Lemma 2.2 that

$$\sqrt{\mu}\, v_+ = P(w_+)^{\frac{1}{2}} s_+ \sim \left(P(x_+)^{\frac{1}{2}} s_+\right)^{\frac{1}{2}}.$$

Due to (36), (37) and Lemma 2.1, with $p = w^{\frac{1}{2}}$, we may write

$$P(x_+)^{\frac{1}{2}} s_+ = \mu P\left(P(w)^{\frac{1}{2}}(v + d_x)\right)^{\frac{1}{2}} P(w)^{-\frac{1}{2}}(v + d_s) \sim \mu P(v + d_x)^{\frac{1}{2}}(v + d_s).$$

From this the lemma follows. □

The above lemma implies

$$v_+^2 \sim P(v+d_x)^{\frac{1}{2}}(v+d_s).$$

Now let $\delta(v) < 1$. Due to (38), we have

$$z = (v+d_x) \circ (v+d_s) \in \mathcal{K}_+.$$

By Lemma 3.2, we also have $v+d_x, v+d_s \in \mathcal{K}_+$. Hence we may use Lemma 2.6. This yields the following inequality:

$$4\delta(v_+)^2 = \left\| v_+ - v_+^{-1} \right\|_F^2 \leq \left\| z^{\frac{1}{2}} - z^{-\frac{1}{2}} \right\|_F^2. \tag{40}$$

Using $d_x + d_s = v^{-1} - v$, we obtain

$$z = v^2 + v \circ (d_x + d_s) + d_x \circ d_s = v^2 + v \circ \left(v^{-1} - v\right) + d_x \circ d_s = e + d_x \circ d_s. \tag{41}$$

Lemma 3.5 *If $\delta := \delta(v) < 1$, then the full NT-step is strictly feasible and*

$$\delta(v_+) \leq Q(\delta) := \frac{\delta^2}{\sqrt{2(1-\delta^4)}}. \tag{42}$$

That the iterates are strictly feasible after a full NT-step follows from Lemma 3.2. To simplify the notation, we denote the eigenvalues of $d_x \circ d_s$ as λ_i, $1 \leq i \leq 2N$. We deduce from (40) that $4\delta(v_+)^2 \leq \operatorname{tr}(z) + \operatorname{tr}(z^{-1}) - 2\operatorname{tr}(e)$. Since the eigenvalues of $z = e + d_x \circ d_s$ are $1 + \lambda_i$, we get

$$4\delta(v_+)^2 \leq \sum_{i=1}^{2N} \left(1 + \lambda_i + \frac{1}{1+\lambda_i} - 2\right) = \sum_{i=1}^{2N} \frac{\lambda_i^2}{1+\lambda_i},$$

We now write

$$\sum_{i=1}^{2N} \frac{\lambda_i^2}{1+\lambda_i} = \sum_{i,\lambda_i \geq 0} \frac{\lambda_i^2}{1+\lambda_i} + \sum_{i,\lambda_i \leq 0} \frac{\lambda_i^2}{1+\lambda_i}. \tag{43}$$

Recall that d_x and d_s are orthogonal. Hence $\operatorname{tr}(d_x \circ d_s) = 0$. This implies the existence of a nonnegative number σ such that

$$\sum_{i,\lambda_i \geq 0} \lambda_i = -\sum_{i,\lambda_i \leq 0} \lambda_i = \sigma.$$

Lemma 3.1 implies that $1 + \lambda_i > 0$ for each i. Therefore, $\frac{\lambda_i^2}{1+\lambda_i}$ is convex in λ_i. Since this function vanishes if $\lambda_i = 0$, we may apply Corollary A.1. This gives

$$\sum_{i,\lambda_i > 0} \frac{\lambda_i^2}{1+\lambda_i} \leq \frac{\sigma^2}{1+\sigma}, \qquad \sum_{i,\lambda_i < 0} \frac{\lambda_i^2}{1+\lambda_i} \leq \frac{\sigma^2}{1-\sigma}.$$

Substituting these bounds into (43), we obtain

$$4\delta(v_+)^2 \le \frac{\sigma^2}{1+\sigma} + \frac{\sigma^2}{1-\sigma} = \frac{2\sigma^2}{1-\sigma^2}.$$

The last expression is monotonically increasing in σ. Hence we may replace it by an upper bound. Applying part (iii) and part (iv) of Lemma 2.3, we may write

$$\begin{aligned}\sigma &= \frac{1}{2}\sum_{i=1}^{2N}\left|\lambda_i(d_x \circ d_s)\right| = \frac{1}{2}\sum_{j=1}^{N}\left(\left|\lambda_{\max}\left(d_x^j \circ d_s^j\right)\right| + \left|\lambda_{\min}\left(d_x^j \circ d_s^j\right)\right|\right)\\ &= \frac{1}{2}\sum_{j=1}^{N} 2\max\left\{\left|\left(d_x^j\right)^T d_s^j\right|, \left\|\overline{d_x^j \circ d_s^j}\right\|\right\} \le \sum_{j=1}^{N}\left\|d_x^j\right\|\left\|d_s^j\right\| = \frac{1}{2}\sum_{j=1}^{N}\left\|d_x^j\right\|_F\left\|d_s^j\right\|_F.\end{aligned}$$

Now using $2ab \le a^2 + b^2$ and the orthogonality of d_x and d_s, we obtain

$$\sigma \le \frac{1}{4}\sum_{j=1}^{N}\left(\left\|d_x^j\right\|_F^2 + \left\|d_s^j\right\|_F^2\right) = \frac{1}{4}\left(\|d_x\|_F^2 + \|d_s\|_F^2\right) = \frac{1}{4}\|d_x + d_s\|_F^2 = \delta(v)^2.$$

With $\delta = \delta(v)$, we thus proved that $\sigma \le \delta^2$. Substitution of this bound for σ yields

$$4\delta(v_+)^2 \le \frac{2\delta^4}{1-\delta^4},$$

which implies the lemma. □

Corollary 3.1 *If $\delta(v) \le \frac{1}{\sqrt[4]{2}}$ then $\delta(v_+) \le \delta(v)^2$, showing that the NT-process converges quadratically fast to the μ-center.*

3.5.3 Updating the Barrier Parameter μ

In this section, we establish a simple relation for our proximity measure just before and after a μ-update.

Lemma 3.6 *Let $(x,s) \in \mathcal{K}_+$, $x^T s = N\mu$, and $\delta = \delta(x,s;\mu)$. If $\mu^+ = (1-\theta)\mu$ for some $0 < \theta < 1$, then*

$$\delta\left(x,s;\mu^+\right)^2 = \frac{\theta^2 N}{2(1-\theta)} + (1-\theta)\delta^2.$$

Proof When updating μ to μ^+, the vector v is divided by the factor $\sqrt{1-\theta}$. Hence we may write

$$4\delta\left(x,s;\mu^+\right)^2 = \left\|\sqrt{1-\theta}v^{-1} - \frac{v}{\sqrt{1-\theta}}\right\|_F^2 = \left\| -\frac{\theta v}{\sqrt{1-\theta}} + \sqrt{1-\theta}\left(v^{-1} - v\right)\right\|_F^2.$$

Yet we observe that the vectors v and $v^{-1} - v$ are orthogonal. This is due to $\operatorname{tr}(x \circ s) = 2N\mu$, which by (33) implies that $\operatorname{tr}(v^2) = 2N$. Hence we have

$$\operatorname{tr}\bigl(v \circ \bigl(v^{-1} - v\bigr)\bigr) = \operatorname{tr}\bigl(e - v^2\bigr) = \operatorname{tr}(e) - \operatorname{tr}\bigl(v^2\bigr) = 2N - 2N = 0.$$

Therefore, using $\|v\|_F^2 = \operatorname{tr}(v^2) = 2N$, we may proceed as follows:

$$4\delta\bigl(x, s; \mu^+\bigr)^2 = \frac{\theta^2}{1-\theta}\|v\|_F^2 + (1-\theta)\bigl\|v^{-1} - v\bigr\|_F^2 = \frac{2\theta^2 N}{1-\theta} + 4(1-\theta)\delta^2.$$

This implies the lemma. □

3.6 Iteration Bound

We conclude this section with an iteration bound for the Algorithm 1. Since we are mainly interested in the question how fast the number of iterations grows as a function of N, it suffices to find a valid upper bound for large values of N. The next theorem only holds for $N \geq 6$, but one can easily show that for smaller values of N the bound remains $O(\sqrt{N})$.

Theorem 3.1 *If $N \geq 6$, $\theta = \frac{1}{\sqrt{2N}}$ and $\tau = \frac{1}{\sqrt{2(1-\theta)}}$, then the number of iterations of Algorithm 1 does not exceed*

$$\sqrt{2N}\log\frac{N\mu^0}{\varepsilon}.$$

Proof Let x, s be iterates at the start of an iteration and $x^T s = N\mu$. At the start of the first iteration we have $\delta(x, s; \mu) \leq \tau$. We claim that this property is maintained during the course of the algorithm. In other words, if $\delta(x, s; \mu) \leq \tau$ at the start of some iteration, then after the NT-step and the μ-update we have $\delta(x_+, s_+; \mu^+) \leq \tau$. Due to Lemmas 3.5 and 3.6, this will hold if

$$\frac{\theta^2 N}{2(1-\theta)} + (1-\theta)Q(\tau)^2 \leq \tau^2.$$

Assuming for the moment that $\tau \leq \frac{1}{\sqrt[4]{2}}$, we have $Q(\tau) \leq \tau^2$. Hence the above inequality will hold if

$$\frac{\theta^2 N}{2(1-\theta)} + (1-\theta)\tau^4 - \tau^2 \leq 0.$$

Since $\theta = \frac{1}{\sqrt{2N}}$, this inequality reduces to

$$\frac{1}{4(1-\theta)} + (1-\theta)\tau^4 - \tau^2 = (1-\theta)\left(\tau^2 - \frac{1}{2(1-\theta)}\right)^2 \leq 0,$$

and this holds for the value of τ in the lemma. It remains to deal with the above assumption that $\tau \leq \frac{1}{\sqrt[4]{2}}$. Using the definition of τ, one easily verifies that the assumption holds if $N \geq 6$. The iteration bound follows from [35, Lemma I.36]. Hence the proof is complete. □

4 An Infeasible Full NT-step Algorithm

In this section, we present our infeasible interior-point algorithm. As has become usual for infeasible IPMs, we start the algorithm with a triple (x^0, y^0, s^0) and $\mu^0 > 0$, such that

$$x^0 = \zeta e, \qquad y^0 = 0, \qquad s^0 = \zeta e, \qquad \mu^0 = \zeta^2, \tag{44}$$

where ζ is a (positive) number such that

$$x^* + s^* \preceq_{\mathcal{K}} \zeta e, \tag{45}$$

for some optimal solutions (x^*, y^*, s^*) of (3) and (4). We call a triple (x, y, s) an ε-solution of (3) and (4) iff the duality gap and the norms of the residual vectors $b - Ax$ and $c - A^T y - s$ do not exceed ε. The algorithm generates an ε-solution of (P) and (D), or it establishes that there do not exist optimal solutions satisfying (45).

The initial values of the primal and dual residual vectors are denoted as r_b^0 and r_c^0, respectively. So we have

$$r_b^0 := b - Ax^0, \tag{46}$$

$$r_c^0 := c - A^T y^0 - s^0. \tag{47}$$

In general, we have $r_b^0 \neq 0$ and $r_c^0 \neq 0$. In other words, the initial iterates are not feasible. The iterates generated by the algorithm will (in general) be infeasible for (3) and (4) as well, but they will be feasible for perturbed versions of (3) and (4) that we introduce in the next subsection.

4.1 Perturbed Problems

For any ν with $0 < \nu \leq 1$, we consider the perturbed problem (P_ν), defined by

$$\min\bigl\{\bigl(c - \nu r_c^0\bigr)^T x : b - Ax = \nu r_b^0, x \in \mathcal{K}\bigr\}, \tag{P_ν}$$

and its dual problem (D_ν), which is given by

$$\max\bigl\{\bigl(b - \nu r_b^0\bigr)^T y : c - A^T y - s = \nu r_c^0, s \in \mathcal{K}\bigr\}. \tag{D_ν}$$

Note that these problems are defined in such a way that if (x, y, s) is feasible for (P_ν) and (D_ν), then the residual vectors for the given triple (x, y, s), with respect to the original problems (3) and (4) are νr_b^0 and νr_c^0, respectively. If $\nu = 1$, then $x = x^0$ yields a strictly feasible solution of (P_ν), and $(y, s) = (y^0, s^0)$ is a strictly feasible solution of (D_ν). This means that if $\nu = 1$, then (P_ν) and (D_ν) satisfy the IPC.

Lemma 4.1 *Let* (3) *and* (4) *be feasible and* $0 < \nu \leq 1$. *Then the perturbed problems* (P_ν) *and* (D_ν) *satisfy the IPC.*

Proof Let $\bar{x}$ be a feasible solution of (3) and $(\bar{y}, \bar{s})$ a feasible solution of (4). Then $A\bar{x} = b$ and $A^T\bar{y} + \bar{s} = c$, with $\bar{x} \in \mathcal{K}$ and $\bar{s} \in \mathcal{K}$. Consider

$$x = (1-\nu)\bar{x} + \nu x^0, \qquad y = (1-\nu)\bar{y} + \nu y^0, \qquad s = (1-\nu)\bar{s} + \nu s^0.$$

Since x is the sum of the vectors $(1-\nu)\bar{x} \in \mathcal{K}$ and $\nu x^0 \in \mathcal{K}_+$, we have $x \in \mathcal{K}_+$. Moreover,

$$b - Ax = b - A\big[(1-\nu)\bar{x} + \nu x^0\big] = b - (1-\nu)b - \nu A x^0 = \nu\big(b - Ax^0\big) = \nu r_b^0,$$

showing that x is strictly feasible for (P_ν). In precisely the same way, one shows that (y, s) is strictly feasible for (D_ν). Thus we have shown that (P_ν) and (D_ν) satisfy the IPC. □

It should be mentioned that the problems (P_ν) and (D_ν) have been studied first in [36], and later also in [37].

4.2 The Central Path of the Perturbed Problems

Let (3) and (4) be feasible and $0 < \nu \le 1$. Then Lemma 4.1 implies that the problems (P_ν) and (D_ν) satisfy the IPC, and therefore their central paths exist. This means that the system

$$b - Ax = \nu r_b^0, \qquad x \in \mathcal{K}, \tag{48}$$

$$c - A^T y - s = \nu r_c^0, \qquad s \in \mathcal{K}, \tag{49}$$

$$x \circ s = \mu e.$$

has a unique solution, for every $\mu > 0$. This solution is denoted as $x(\mu, \nu)$ and $(y(\mu, \nu), s(\mu, \nu))$. These are the μ-centers of the perturbed problems (P_ν) and (D_ν). In the sequel, the parameters μ and ν will always be in a one-to-one correspondence, according to

$$\mu = \nu\mu^0 = \nu\zeta^2,$$

and, therefore, we feel free to denote $x(\mu, \nu)$ and $(y(\mu, \nu), s(\mu, \nu))$ simply as $x(\nu)$ and $(y(\nu), s(\nu))$. Due to the choice of the initial iterates, according to (44), we have $x^0 \circ s^0 = \mu^0 e$. Hence x^0 is the μ^0-center of the perturbed problem (P_1), and (y^0, s^0) the μ^0-center of (D_1). In other words, $(x(1), y(1), s(1)) = (x^0, y^0, s^0)$.

4.3 An Iteration of our Algorithm

We just established that if $\nu = 1$ and $\mu = \mu^0$, then $x = x^0$, and $(y, s) = (y^0, s^0)$ are the μ-center of (P_ν) and (D_ν), respectively. These are our initial iterates.

We measure proximity to the μ-center of the perturbed problems by the quantity $\delta(x, s; \mu)$, as defined in (34). So, initially we have $\delta(x, s; \mu) = 0$. In the sequel, we assume that at the start of each iteration, just before the μ-update, $\delta(x, s; \mu)$ is smaller than or equal to a (small) threshold value $\tau > 0$. Since we then have

$\delta(x, s; \mu) = 0$, this condition is certainly satisfied at the start of the first iteration, and also $x^T s = N\mu^0$.

Now we describe one (main) iteration of our algorithm. Suppose that for some $\nu \in (0, 1]$ we have x, y and s satisfying the feasibility conditions (48) and (49) for $\mu = \nu\mu^0$, and such that $x^T s = N\mu$ and $\delta(x, s; \mu) \le \tau$. Now, we reduce ν to $\nu^+ = (1-\theta)\nu$, with $\theta \in (0, 1)$, and find new iterates x_+, y_+ and s_+ that satisfy (48) and (49), with ν replaced by ν^+ and μ by $\mu^+ = \nu^+\mu^0 = (1-\theta)\mu$, and such that $x_+^T s_+ = N\mu^+$ and $\delta(x_+, s_+; \mu^+) \le \tau$.

One (main) iteration consists of a feasibility step and a few centering steps. The feasibility step serves to get iterates (x_f, y_f, s_f) that are strictly feasible for (P_{ν^+}) and (D_{ν^+}), and such that for some positive number τ_f, one conclude that $\delta(x_f, s_f; \mu^+) \le \tau_f < \sqrt[4]{2}$. Because the NT-step is then quadratically convergent, a few centering steps, starting at (x_f, y_f, s_f) and targeting at the μ^+-centers of (P_{ν^+}) and (D_{ν^+}), will generate iterates (x_+, y_+, s_+) that are feasible for (P_{ν^+}) and (D_{ν^+}), and that satisfy $\delta(x_+, s_+; \mu^+) \le \tau$. By Lemma 3.5, after k centering steps, we obtain iterates (x_+, y_+, s_+) that are still feasible for (P_{ν^+}) and (D_{ν^+}) and such that

$$\delta\bigl(x_+, s_+; \mu^+\bigr) \le Q^k(\tau_f),$$

with the function Q as defined in (42). Hence we will have $\delta(x_+, s_+; \mu^+) \le \tau$ if

$$Q^k(\tau_f) := \underbrace{Q\bigl(Q\bigl(\dots Q}_{k \text{ times}}(\tau_f)\bigr)\bigr) \le \tau. \tag{50}$$

From this one easily obtains an upper bound for the required number of centering steps.

Since each main iteration reduces the duality gap $x^T s$ with the factor $1-\theta$, and the size of the residual vectors are reduced with the same factor, given θ we can also compute the number of main iterations that is necessary to satisfy the stopping criterion in the algorithm. If our aim is to get the duality gap and the norms of the residual vectors less than or equal to some small number ε, then this number is given by

$$\frac{1}{\theta} \log \frac{\max\{N\zeta^2, \|r_b^0\|, \|r_c^0\|\}}{\varepsilon}. \tag{51}$$

It may be clear that we only need to define and analyze the feasibility step. This is the most difficult part of the analysis in this paper. In essence, we follow the same chain of arguments as in [26], but at several places the analysis is more tight and also more elegant.

In the rest of this section we describe the feasibility step in detail. The analysis will follow in subsequent sections. Suppose we have strictly feasible iterates (x, y, s) for (P_ν) and (D_ν). This means that (x, y, s) satisfies (48) and (49), with $\mu = \nu\zeta^2$. We need displacements $\Delta^f x$, $\Delta^f y$ and $\Delta^f s$ such that

$$\begin{aligned} x_f &:= x + \Delta^f x, \\ y_f &:= y + \Delta^f y, \\ s_f &:= s + \Delta^f s, \end{aligned} \tag{52}$$

are feasible for (P_{ν^+}) and (D_{ν^+}). One may easily verify that (x_f, y_f, s_f) satisfies (48) and (49), with ν replaced by ν^+ and μ by $\mu^+ = \nu^+\mu^0 = (1-\theta)\mu$, only if the first two equations in the following system are satisfied.

$$\begin{aligned} A\Delta^f x &= \theta\nu r_b^0, \\ A^T\Delta^f y + \Delta^f s &= \theta\nu r_c^0, \\ P(u)x \circ P\big(u^{-1}\big)\Delta^f s + P\big(u^{-1}\big)s \circ P(u)\Delta^f x &= (1-\theta)\mu e - P(u)x \circ P\big(u^{-1}\big)s. \end{aligned} \tag{53}$$

The third equation is inspired by the third equation in the system (24) that we used to define search directions for the feasible case, except that we target at the μ^+-centers of (P_{ν^+}) and (D_{ν^+}). As in the feasible case, we use the NT-scaling scheme to guarantee that the above system has a unique solution. So we take $u = w^{-\frac{1}{2}}$, where w is the NT-scaling point of x and s. Then the third equation becomes

$$P(w)^{-\frac{1}{2}}x \circ P(w)^{\frac{1}{2}}\Delta^f s + P(w)^{\frac{1}{2}}s \circ P(w)^{-\frac{1}{2}}\Delta^f x = (1-\theta)\mu e - P(w)^{-\frac{1}{2}}x \circ P(w)^{\frac{1}{2}}s. \tag{54}$$

Due to this choice of u, the coefficient matrix of the resulting system is exactly the same as in the feasible case, and hence it defines the feasibility step uniquely.

By its definition, after the feasibility step, the iterates satisfy the affine equations in (48) and (49), with $\nu = \nu^+$. The hard part in the analysis will be to guarantee that $x_f, s_f \in \mathcal{K}_+$ and to guarantee that the new iterates satisfy $\delta(x_f, s_f; \mu^+) \le \tau_f$.

4.4 The Infeasible Algorithm

A formal description of the algorithm is given in Algorithm 2. Recall that after each iteration, the residuals and the duality gap are reduced by the factor $1-\theta$. The algorithm stops if the norms of the residuals and the duality gap are less than the accuracy parameter ε.

4.5 Analysis of the Feasibility Step

Let x, y and s denote the iterates at the start of an iteration with $x^T s = N\mu$ and $\delta(x, s; \mu) \le \tau$. Recall that at the start of the first iteration this is certainly true, because $(x^0)^T s^0 = N\mu^0$ and $\delta(x^0, s^0; \mu^0) = 0$.

We scale the matrix A and the search directions, just as we did in the feasible case (cf. (26)), by defining

$$\bar{A} := \sqrt{\mu} A P(w)^{\frac{1}{2}}, \qquad d_x^f := \frac{P(w)^{-\frac{1}{2}}\Delta^f x}{\sqrt{\mu}}, \qquad d_s^f := \frac{P(w)^{\frac{1}{2}}\Delta^f s}{\sqrt{\mu}}, \tag{55}$$

with w denoting the scaling point of x and s, as defined in (8). With the vector v as defined before (cf. (25)), equation (54) can be restated as

$$\mu v \circ \big(d_x^f + d_s^f\big) = (1-\theta)\mu e - \mu v^2.$$

Algorithm 2 PRIMAL–DUAL INFEASIBLE IPM

Input:

accuracy parameter $\varepsilon > 0$;
barrier update parameter θ, $0 < \theta < 1$;
threshold parameter $\tau > 0$;
initialization parameter $\zeta > 0$.

begin

$x^0 := \zeta e, s^0 := \zeta e, y^0 := 0; \mu^0 := \zeta^2$;
while $\max(x^T s, \nu\|r_b^0\|, \nu\|r_c^0\|) \geq \varepsilon$
feasibility step:
$(x, y, s) := (x, y, s) + (\Delta^f x, \Delta^f y, \Delta^f s)$;
update of μ and ν:
$\mu := (1-\theta)\mu, \nu := (1-\theta)\nu$;
centering steps:
while $\delta(x, s; \mu) \geq \tau$
$(x, y, s) := (x, y, s) + (\Delta x, \Delta y, \Delta s)$;
endwhile
endwhile

end

By multiplying both sides of this equation from left with $\mu^{-1} L(v)^{-1}$, this equation gets the form

$$d_x^f + d_s^f = (1-\theta)v^{-1} - v.$$

Thus we arrive at the following system for the scaled search directions in the feasibility step:

$$\begin{aligned} \bar{A} d_x^f &= \theta \nu r_b^0, \\ \frac{1}{\mu} \bar{A}^T \Delta^f y + d_s^f &= \frac{1}{\sqrt{\mu}} \theta \nu P(w)^{\frac{1}{2}} r_c^0, \\ d_x^f + d_s^f &= (1-\theta)v^{-1} - v. \end{aligned} \tag{56}$$

To get the search directions $\Delta^f x$ and $\Delta^f s$ in the x- and s-space we use (55), which gives

$$\Delta^f x = \sqrt{\mu} P(w)^{\frac{1}{2}} d_x^f, \qquad \Delta^f s = \sqrt{\mu} P(w)^{-\frac{1}{2}} d_s^f. \tag{57}$$

The new iterates are obtained by taking a full step, as given by (52). Hence we have

$$x_f = x + \Delta^f x = \sqrt{\mu} P(w)^{\frac{1}{2}} \left(v + d_x^f\right), \tag{58}$$

$$s_f = s + \Delta^f s = \sqrt{\mu} P(w)^{-\frac{1}{2}} \left(v + d_s^f\right). \tag{59}$$

From the third equation in (56) we derive that

$$\left(v + d_x^f\right) \circ \left(v + d_s^f\right) = v^2 + v \circ \left[(1-\theta)v^{-1} - v\right] + d_x^f \circ d_s^f = (1-\theta)e + d_x^f \circ d_s^f.$$

As we mentioned before, the analysis of the algorithm as presented below is much more difficult than in the feasible case. The main reason for this is that the scaled search directions d_x^f and d_s^f are not (necessarily) orthogonal.

4.5.1 Feasibility of the Feasibility Step

By the same arguments as in Sect. 3.5.1, it follows from (58) and (59) that x_f and s_f are strictly feasible if and only if $v + d_x^f$ and $v + d_s^f$ belong to $\mathcal{K}_+$. Using this, we have the following result.

Lemma 4.2 *The iterates* (x_f, y_f, s_f) *are strictly feasible if*

$$(1-\theta)e + d_x^f \circ d_s^f \in \mathcal{K}_+.$$

Proof Just as in the proof of Lemma 3.2, we introduce a step length α, $0 \le \alpha \le 1$, and we define

$$v_x^\alpha = v + \alpha d_x^f, \qquad v_s^\alpha = v + \alpha d_s^f.$$

We then have $v_x^0 = v$, $v_x^1 = v + d_x^f$ and $v_s^0 = v$, $v_s^1 = v + d_s^f$.

Since $d_x^f + d_s^f = (1-\theta)v^{-1} - v$, it follows that

$$\begin{aligned} v_x^\alpha \circ v_s^\alpha &= \left(v + \alpha d_x^f\right) \circ \left(v + \alpha d_s^f\right) \\ &= v^2 + \alpha v \circ \left(d_x^f + d_s^f\right) + \alpha^2 d_x^f \circ d_s^f \\ &= v^2 + \alpha v \circ \left[(1-\theta)v^{-1} - v\right] + \alpha^2 d_x \circ d_s \\ &= (1-\alpha)v^2 + \alpha(1-\theta)e + \alpha^2 d_x \circ d_s. \end{aligned}$$

The hypothesis in the lemma implies that $d_x^f \circ d_s^f \succ_{\mathcal{K}} -(1-\theta)e$. Substitution gives

$$v_x^\alpha \circ v_s^\alpha \succ_{\mathcal{K}} (1-\alpha)v^2 + \alpha(1-\theta)e - \alpha^2(1-\theta)e = (1-\alpha)\left(v^2 + \alpha(1-\theta)e\right). \quad (60)$$

Since $v^2 \in \mathcal{K}$ and $\alpha(1-\theta)e \in \mathcal{K}$, we have $v^2 + \alpha(1-\theta)e \in \mathcal{K}$. Hence, if $0 \le \alpha \le 1$, then $(1-\alpha)(v^2 + \alpha(1-\theta)e) \in \mathcal{K}$. Due to (60), this implies that $v_x^\alpha \circ v_s^\alpha \in \mathcal{K}_+$. Therefore, all eigenvalues of $v_x^\alpha \circ v_s^\alpha$ are positive, whence we have $\det(v_x^\alpha \circ v_s^\alpha) > 0$, for each $\alpha \in [0, 1]$. By Lemma 2.3(ii), this implies that $\det(v_x^\alpha)\det(v_s^\alpha) > 0$, for each $\alpha \in [0, 1]$. It follows that $\det(v_x^\alpha)$ and $\det(v_s^\alpha)$ do not vanish for $\alpha \in [0, 1]$. Since $\det(v_x^0) = \det(v_s^0) = \det(v) > 0$, by continuity, $\det(v_x^\alpha)$ and $\det(v_s^\alpha)$ stay positive for all $\alpha \in [0, 1]$. Since $\det(v_x^\alpha)$ and $\det(v_s^\alpha)$, do not vanish for all $\alpha \in [0, 1]$, it follows that the eigenvalues of v_x^α and v_s^α stay positive for all $\alpha \in [0, 1]$. In particular, the eigenvalues of v_x^1 and v_s^1 are positive, which means that $v + d_x^f$ and $v + d_s^f$ belong to $\mathcal{K}_+$. Hence the proof of the lemma is complete. □

Clearly, from the above lemma that the feasibility of the iterates (x_f, y_f, s_f) highly depends on the eigenvalues of the vector $d_x^f \circ d_s^f$. It will be convenient to denote the $2N$ eigenvalues of any vector $x \in \mathbb{R}^n$ as $\lambda_i(x)$, $1 \le i \le 2N$. Then it follows from Lemma 4.2 that (x_f, y_f, s_f) is strictly feasible if

$$(1-\theta) + \lambda_i\big(d_x^f \circ d_s^f\big) > 0, \quad i = 1, \dots, 2N. \tag{61}$$

We assume below that these inequalities hold.

4.5.2 Proximity After the Feasibility Step

We proceed by deriving an upper bound for $\delta(x_f, s_f; \mu^+)$. Let w_f be the scaling point of x_f and s_f. When denoting the v-vector after the feasibility step, with respect to the μ^+-center, as v_f, according to (25), this vector is given by

$$v_f := \frac{P(w_f)^{-\frac{1}{2}} x_f}{\sqrt{\mu(1-\theta)}} \qquad \left[= \frac{P(w_f)^{\frac{1}{2}} s_f}{\sqrt{\mu(1-\theta)}}\right]. \tag{62}$$

Lemma 4.3 *One has*

$$\sqrt{1-\theta}\, v_f \sim \big[P\big(v + d_x^f\big)^{\frac{1}{2}}\big(v + d_s^f\big)\big]^{\frac{1}{2}}.$$

Proof It follows from (62) and Lemma 2.2 that

$$\sqrt{\mu(1-\theta)}\, v_f = P(w_f)^{\frac{1}{2}} s_f \sim \big(P(x_f)^{\frac{1}{2}} s_f\big)^{\frac{1}{2}}.$$

Due to (58), (59) and Lemma 2.1, with $p = w^{\frac{1}{2}}$, we may write

$$\begin{aligned} P(x_f)^{\frac{1}{2}} s_f &= \mu P\big(P(w)^{\frac{1}{2}}\big(v + d_x^f\big)\big)^{\frac{1}{2}} P(w)^{-\frac{1}{2}}\big(v + d_s^f\big) \\ &\sim \mu P\big(v + d_x^f\big)^{\frac{1}{2}}\big(v + d_s^f\big). \end{aligned}$$

Thus we obtain

$$\sqrt{\mu(1-\theta)}\, v_f \sim \sqrt{\mu}\big[P\big(v + d_x^f\big)^{\frac{1}{2}}\big(v + d_s^f\big)\big]^{\frac{1}{2}}.$$

From this the lemma follows. □

The above lemma implies that

$$v_f^2 \sim P\left(\frac{v + d_x^f}{\sqrt{1-\theta}}\right)^{\frac{1}{2}}\left(\frac{v + d_s^f}{\sqrt{1-\theta}}\right).$$

In the sequel, we denote $\delta(x_f, s_f; \mu^+)$ also briefly by $\delta(v_f)$. By Lemma 2.6 (with $x = \frac{v+d_x^f}{\sqrt{1-\theta}}$, $s = \frac{v+d_s^f}{\sqrt{1-\theta}}$, $u = P(x)^{\frac{1}{2}} s$ and $z = x \circ s$) this implies the inequality below:

$$4\delta(v_f)^2 = \big\|v_f - v_f^{-1}\big\|_F^2 = \big\|u^{\frac{1}{2}} - u^{-\frac{1}{2}}\big\|_F^2 \le \big\|z^{\frac{1}{2}} - z^{-\frac{1}{2}}\big\|_F^2.$$

Since $d_x + d_s = (1-\theta)v^{-1} - v$, one has

$$(1-\theta)z = (v+d_x)\circ(v+d_s) = v^2 + v\circ(d_x+d_s) + d_x\circ d_s$$
$$= v^2 + v\circ\big((1-\theta)v^{-1} - v\big) + d_x\circ d_s = (1-\theta)e + d_x\circ d_s.$$

So we have

$$4\delta(v_f)^2 \le \big\|z^{\frac{1}{2}} - z^{-\frac{1}{2}}\big\|_F^2 = \mathrm{tr}(z) + \mathrm{tr}\big(z^{-1}\big) - 2\mathrm{tr}(e), \qquad z = e + \frac{d_x\circ d_s}{1-\theta}. \tag{63}$$

In what follows, we denote the eigenvalues $\lambda_i(d_x\circ d_s)$ of $d_x\circ d_s$ simply as λ_i, $1\le i\le 2N$, and λ will denote the vector in $\mathbb{R}^{2N}$ with the eigenvalues λ_i as entries (in some arbitrary order). We can prove the following result. In this result $\|\lambda\|_1$ denotes the 1-norm of λ, i.e., the sum of the absolute values of the eigenvalues λ_i.

Lemma 4.4 *If* $(1-\theta)e + d_x^f\circ d_s^f \in \mathcal{K}_+$, *then*

$$4\delta(v_f)^2 \le f\left(\frac{\|\lambda\|_1}{1-\theta}\right),$$

where

$$f(t) := 1 - t + \frac{1}{1-t} - 2 = \frac{t^2}{1-t}, \qquad |t| < 1. \tag{64}$$

Proof First note that the eigenvalues of z are $1 + \lambda_i/(1-\theta)$, and by (61) these are positive. For the moment, define $\sigma_i = \lambda_i/(1-\theta)$. Then $1+\sigma_i > 0$, for each i. Using (63), we obtain

$$4\delta(v_f)^2 \le \sum_{i=1}^{2N}\left(1+\sigma_i + \frac{1}{1+\sigma_i} - 2\right) = \sum_{i=1}^{2N}\frac{\sigma_i^2}{1-\sigma_i} \le \sum_{i=1}^{2N}\frac{\sigma_i^2}{1-\|\sigma\|_\infty}$$
$$= \frac{\|\sigma\|^2}{1-\|\sigma\|_\infty} \le \frac{\|\sigma\|_1^2}{1-\|\sigma\|_1}.$$

For the last inequality we used that $\|\sigma\|_\infty \le \|\sigma\| \le \|\sigma\|_1$. This implies the lemma. □

Using Lemma 2.3, an upper bound for $\|\lambda\|_1$ can be obtained as follows:

$$\|\lambda\|_1 = \sum_{i=1}^{2N}\big|\lambda_i\big(d_x^f\circ d_s^f\big)\big| = \sum_{j=1}^{N}\big(\big|\lambda_{\max}\big(\big(d_x^f\big)^j\circ\big(d_s^f\big)^j\big)\big| + \big|\lambda_{\min}\big(\big(d_x^f\big)^j\circ\big(d_s^f\big)^j\big)\big|\big)$$
$$= 2\sum_{j=1}^{N}\max\big\{\big|\big(\big(d_x^f\big)^j\big)^T\big(d_s^f\big)^j\big|, \big\|\overline{\big(d_x^f\big)^j\circ\big(d_s^f\big)^j}\big\|\big\} \le 2\sum_{j=1}^{N}\big\|\big(d_x^f\big)^j\big\|\big\|\big(d_s^f\big)^j\big\|$$
$$= \sum_{j=1}^{N}\big\|\big(d_x^f\big)^j\big\|_F\big\|\big(d_s^f\big)^j\big\|_F \le \frac{1}{2}\sum_{j=1}^{N}\big(\big\|\big(d_x^f\big)^j\big\|_F^2 + \big\|\big(d_s^f\big)^j\big\|_F^2\big)$$
$$= \frac{1}{2}\big(\big\|d_x^f\big\|_F^2 + \big\|d_s^f\big\|_F^2\big).$$

In the present case, contrary to the case of a feasible method, the scaled search directions d_x^f and d_s^f are not orthogonal. As has become clear in the case of LO, this fact complicates the analysis drastically [26]. To deal with this complication it will be convenient to define

$$\omega(v) := \frac{1}{2}\sqrt{\|d_x^f\|_F^2 + \|d_s^f\|_F^2}.$$

Then, since $\|\lambda\|_1 \leq 2\omega(v)^2$, it follows from Lemma 4.4 that

$$4\delta(v_f)^2 \leq f\left(\frac{2\,\omega(v)^2}{1-\theta}\right).$$

We need to have $\delta(v_f) \leq \tau_f$. Clearly, from the above inequality, it suffices for this if

$$f\left(\frac{2\,\omega(v)^2}{1-\theta}\right) \leq 4\tau_f^2. \tag{65}$$

Obviously, $f(t)$ is (strict) monotonically increasing for $t \in [0, 1)$, and therefore has an inverse function $g(s)$, $s \in [0, \infty)$. Defining

$$\rho(\delta) := \delta + \sqrt{1+\delta^2}, \tag{66}$$

it can be easily verified that

$$g(s) = \frac{\sqrt{s}}{\rho(\frac{1}{2}\sqrt{s})}, \qquad s \geq 0.$$

Therefore, by applying g to both sides of (65), we obtain the equivalent inequality

$$\frac{2\,\omega(v)^2}{1-\theta} \leq g\left(4\tau_f^2\right) = \frac{2\tau_f}{\rho(\tau_f)}.$$

Hence we should find θ such that it is positive (and as large as possible) and such that it satisfies

$$\omega(v)^2 \leq \frac{(1-\theta)\tau_f}{\rho(\tau_f)}. \tag{67}$$

It should be noted that by its definition, $\omega(v)$ depends on d_x^f and d_s^f, and hence on θ itself. In the next section we investigate this dependence.

4.5.3 Upper Bound for $\omega(v)$

Recall that the scaled search directions d_x^f and d_s^f are determined by the system (56). Let us define the linear space $\mathcal{S}$ as follows:

$$\mathcal{S} := \left\{\xi \in \mathbb{R}^n : \bar{A}\xi = 0\right\}.$$

Clearly, from the first equation in (56) the affine space

$$\left\{\xi \in \mathbb{R}^n : \bar{A}\xi = \theta \nu r_b^0\right\}$$

equals $d_x^f + \mathcal{S}$. Also, from linear algebra, we know that the orthogonal complement of the linear space $\mathcal{S}$ is the row space of $\bar{A}$, i.e.

$$\mathcal{S}^\perp = \left\{\bar{A}^T \vartheta : \vartheta \in \mathbb{R}^m\right\}.$$

From the second equation in (56), Clearly, the affine space

$$\left\{\frac{1}{\sqrt{\mu}}\theta \nu P(w)^{\frac{1}{2}} r_c^0 + \bar{A}^T \vartheta : \vartheta \in \mathbb{R}^m\right\}$$

equals $d_s^f + \mathcal{S}^\perp$. Since $\mathcal{S} \cap \mathcal{S}^\perp = \{0\}$, the spaces $d_x^f + \mathcal{S}$ and $d_s^f + \mathcal{S}^\perp$ meet in a unique point. We call this point q. So q is uniquely determined by the system

$$\bar{A} q = \theta \nu r_b^0, \tag{68}$$

$$\bar{A}^T \vartheta + q = \frac{1}{\sqrt{\mu}}\theta \nu P(w)^{\frac{1}{2}} r_c^0. \tag{69}$$

Lemma 4.5 *One has*

$$4\,\omega(v)^2 \leq \|q\|_F^2 + \left(\|q\|_F + \sqrt{4(1-\theta)^2\delta(v)^2 + 2\theta^2 N}\right)^2.$$

Proof To simplify the notation, in this proof we denote $r = (1-\theta)v^{-1} - v$. Using exactly the same arguments as in the proof of Lemma 4.6 in [26], one shows that

$$4\omega(v)^2 = \|d_x^f\|_F^2 + \|d_s^f\|_F^2 = \|q\|_F^2 + \|q - r\|_F^2.$$

Since $\|q - r\|_F \leq \|q\|_F + \|r\|_F$, by the triangle inequality, we get

$$4\omega(v)^2 \leq \|q\|_F^2 + (\|q\|_F + \|r\|_F)^2. \tag{70}$$

Recall that v is the v-vector of vectors x and s that are feasible for (P_ν) and (D_ν). These vectors are obtained after a full-NT-step for a feasible problem, whence $\mu\|v\|_F^2 = 2x^T s = 2N\mu$. The latter means that v is orthogonal to $v - v^{-1}$. So we may write

$$\begin{aligned}\|r\|_F^2 &= \left\|(1-\theta)v^{-1} - v\right\|_F^2 = \left\|(1-\theta)\left(v^{-1} - v\right) - \theta v\right\|_F^2 \\ &= (1-\theta)^2\left\|v^{-1} - v\right\|_F^2 + \theta^2\|v\|_F^2.\end{aligned}$$

Since $\|v^{-1} - v\|_F^2 = 4\delta(v)^2$ and $\|v\|_F^2 = 2N$, we obtain

$$\|r\|_F^2 = 4(1-\theta)^2\delta(v)^2 + 2\theta^2 N.$$

Substitution into (70) yields the lemma. □

4.5.4 Upper Bound for $\|q\|$

Recall that the vector q is determined by Eqs. (68) and (69), where $\bar{A} = \sqrt{\mu} A P(w)^{\frac{1}{2}}$, with w denoting the scaling point of x and s, as defined in (8). So we have

$$\sqrt{\mu} A P(w)^{\frac{1}{2}} q = \theta \nu r_b^0, \tag{71}$$

$$\sqrt{\mu} P(w)^{\frac{1}{2}} A^T \vartheta + q = \frac{1}{\sqrt{\mu}} \theta \nu P(w)^{\frac{1}{2}} r_c^0. \tag{72}$$

We proceed by proving the following upper bound for $\|q\|_F$.

Lemma 4.6 *If* (x^0, y^0, s^0), (x^*, y^*, s^*) *and* ζ *are as defined in* (44) *and* (45), *then*

$$\|q\|_F \leq \theta \sqrt{\nu \operatorname{tr}\big(w^2 + w^{-2}\big)}. \tag{73}$$

Proof To keep the notation simple, we introduce

$$D := P(w)^{\frac{1}{2}}, \qquad r_b := \theta \nu r_b^0, \qquad r_c := \theta \nu r_c^0.$$

Then equations (71) and (72) get the form

$$\sqrt{\mu} A D q = r_b,$$

$$\sqrt{\mu}\, D A^T \vartheta + q = \frac{1}{\sqrt{\mu}} D r_c.$$

Exactly the same system occurs in [26, Sect. 4.4]. There it has been shown (cf. [26, Eqn. (4.14)]) that it implies the following inequality:

$$\sqrt{\mu} \|q\|_F \leq \theta \nu \sqrt{\big\|D\big(s^0 - s^*\big)\big\|_F^2 + \big\|D^{-1}\big(x^0 - x^*\big)\big\|_F^2}, \tag{74}$$

where we used $\|\cdot\|_F = \sqrt{2}\|\cdot\|$. Since x^* is feasible for (3), we have $x^* \succeq_{\mathcal{K}} 0$. Also $s^* \in \mathcal{K}_+$. Hence we have $0 \preceq_{\mathcal{K}} x^* \preceq_{\mathcal{K}} x^* + s^* \preceq_{\mathcal{K}} \zeta e$. In a similar way we derive for $s^* \succeq_{\mathcal{K}} 0$ that $0 \preceq_{\mathcal{K}} s^* \preceq_{\mathcal{K}} \zeta e$. Therefore it follows that

$$0 \preceq_{\mathcal{K}} x^0 - x^* \preceq_{\mathcal{K}} \zeta e, \qquad 0 \preceq_{\mathcal{K}} s^0 - s^* \preceq_{\mathcal{K}} \zeta e. \tag{75}$$

We first consider the term $\|D(s^0 - s^*)\|_F^2$. Using that D is self-adjoint with respect to the inner product $\langle \cdot, \cdot \rangle$ and $D^2 e = P(w) e = w^2$, we may write

$$\begin{aligned}
\big\|D\big(s^0 - s^*\big)\big\|_F^2 &= \big\langle D\big(s^0 - s^*\big), D\big(s^0 - s^*\big)\big\rangle = \big\langle D^2\big(s^0 - s^*\big), s^0 - s^*\big\rangle \\
&= \big\langle D^2\big(s^0 - s^*\big), \zeta e\big\rangle - \big\langle D^2\big(s^0 - s^*\big), \zeta e - \big(s^0 - s^*\big)\big\rangle \\
&\leq \big\langle D^2\big(s^0 - s^*\big), \zeta e\big\rangle = \big\langle s^0 - s^*, D^2 \zeta e\big\rangle = \zeta \big\langle s^0 - s^*, w^2\big\rangle \\
&= \zeta \big\langle \zeta e, w^2\big\rangle - \zeta \big\langle \zeta e - \big(s^0 - s^*\big), w^2\big\rangle \\
&\leq \zeta \big\langle \zeta e, w^2\big\rangle = \zeta^2 \big\langle e, w^2\big\rangle = \zeta^2 \operatorname{tr}\big(w^2\big).
\end{aligned}$$

In the same way it follows that

$$\left\| D^{-1}\left(x^0 - x^*\right)\right\|_F^2 \le \zeta^2 \mathrm{tr}\left(w^{-2}\right).$$

Substitution of the last two inequalities into (74) gives

$$\sqrt{\mu}\|q\|_F \le \theta\nu\sqrt{\zeta^2\mathrm{tr}\left(w^2\right) + \zeta^2\mathrm{tr}\left(w^{-2}\right)} = \theta\nu\zeta\sqrt{\mathrm{tr}\left(w^2 + w^{-2}\right)}.$$

Finally, by using $\mu = \nu\mu^0 = \nu\zeta^2$, the inequality in the lemma follows. □

Our next task is to find an upper bound for $\mathrm{tr}(w^2 + w^{-2})$. Before doing this, we recall the following relations:

$$P\left(s^{\frac{1}{2}}\right)x \sim P\left(x^{\frac{1}{2}}\right)s \sim \left(P(w)^{\frac{1}{2}}s\right)^2 = \left(P(w)^{-\frac{1}{2}}x\right)^2 = \mu v^2; \tag{76}$$

where the similarities are due to Proposition 2.5(iv) and Lemma 2.2, and the equality to (25). We now can prove the following result.

Lemma 4.7 *Let $x, s \in \mathcal{K}$ and w the scaling point of x and s. Then*

$$\|q\|_F \le \theta\frac{\mathrm{tr}(x+s)}{\zeta\lambda_{\min}(v)}.$$

Proof For the moment, let $u := (P(x^{\frac{1}{2}})s)^{-\frac{1}{2}}$. Then, by (8), $w = P(x^{\frac{1}{2}})u$. Using that $P(x^{\frac{1}{2}})$ is self-adjoint, and also Lemma 2.4, we obtain

$$\mathrm{tr}\left(w^2\right) = \left\langle P\left(x^{\frac{1}{2}}\right)u, P\left(x^{\frac{1}{2}}\right)u\right\rangle = \left\langle u, P(x)u\right\rangle \le \lambda_{\max}(u)\mathrm{tr}\left(P(x)u\right).$$

By using the same arguments and also $P(x)e = x^2$, we may write

$$\mathrm{tr}\left(P(x)u\right) = \mathrm{tr}\left(P(x)u \circ e\right) = \left\langle P(x)u, e\right\rangle = \left\langle u, P(x)e\right\rangle = \left\langle u, x^2\right\rangle \le \lambda_{\max}(u)\mathrm{tr}\left(x^2\right),$$

where the last inequality follows from Lemma 2.4. Combining the above inequalities we obtain

$$\mathrm{tr}\left(w^2\right) \le \lambda_{\max}\left(P\left(x^{\frac{1}{2}}\right)s\right)^{-1}\mathrm{tr}\left(x^2\right).$$

Due to (76) we have

$$\lambda_{\max}\left(P\left(x^{\frac{1}{2}}\right)s\right)^{-1} = \frac{1}{\lambda_{\min}(P(x^{\frac{1}{2}})s)} = \frac{1}{\mu\lambda_{\min}(v)^2}.$$

Thus we obtain

$$\mathrm{tr}\left(w^2\right) \le \frac{\mathrm{tr}(x^2)}{\mu\lambda_{\min}(v)^2}.$$

By noting that w^{-1} is the scaling element of s and x, it follows from the above inequality, by interchanging the role of x and s, that

$$\operatorname{tr}\bigl(w^{-2}\bigr) \le \frac{\operatorname{tr}(s^2)}{\mu\lambda_{\min}(v)^2}.$$

By adding the last two inequalities we obtain

$$\operatorname{tr}\bigl(w^2 + w^{-2}\bigr) \le \frac{\operatorname{tr}(x^2) + \operatorname{tr}(s^2)}{\mu\lambda_{\min}(v)^2}. \tag{77}$$

Since $x, s \in \mathcal{K}$, we have $\operatorname{tr}(x \circ s) \ge 0$. Hence, also using that $\operatorname{tr}(z^2) \le \operatorname{tr}(z)^2$ for each $z \in \mathcal{K}$,

$$\begin{aligned} \operatorname{tr}\bigl(x^2\bigr) + \operatorname{tr}\bigl(s^2\bigr) &\le \operatorname{tr}\bigl(x^2\bigr) + \operatorname{tr}\bigl(s^2\bigr) + 2\operatorname{tr}(x \circ s) \\ &= \operatorname{tr}\bigl((x+s)^2\bigr) \le \operatorname{tr}(x+s)^2. \end{aligned} \tag{78}$$

Substituting (77) and (78) into (73), also using that $\mu = \nu\zeta^2$, yields

$$\|q\|_F \le \theta\sqrt{\nu\frac{\operatorname{tr}(x^2) + \operatorname{tr}(s^2)}{\mu\lambda_{\min}(v)^2}} \le \theta\sqrt{\frac{\operatorname{tr}(x+s)^2}{\zeta^2\lambda_{\min}(v)^2}} = \theta\frac{\operatorname{tr}(x+s)}{\zeta\lambda_{\min}(v)},$$

which completes the proof. □

Lemma 4.8 *Let $\delta = \delta(v)$ be given by (34) and $\rho(\delta)$ as defined in (66). Then we have*

$$\frac{1}{\rho(\delta)} \le \lambda_{\min}\bigl(v^j\bigr) \le \lambda_{\max}\bigl(v^j\bigr) \le \rho(\delta), \quad j \in \{1, \dots, N\},$$

Proof Using (35), the proof is easy and similar to the proof of Lemma II.60 in [35]. □

As a consequence of Lemma 4.8, we have $\lambda_{\min}(v) \ge \frac{1}{\rho(\delta)}$. Hence we obtain from Lemma 4.7 that

$$\|q\|_F \le \frac{\theta\rho(\delta)}{\zeta}\operatorname{tr}(x+s). \tag{79}$$

4.5.5 *Upper Bound for* $\operatorname{tr}(x+s)$

In this section, we compute an upper bound for $\operatorname{tr}(x+s)$.

Lemma 4.9 *Let x and (y, s) be feasible for the perturbed problems (P_ν) and (D_ν), respectively, with $x^T s = N\mu$, and (x^0, y^0, s^0) as defined in (44) and ζ as in (45). We then have*

$$\operatorname{tr}(x+s) \le 4N\zeta.$$

Proof Let (x^*, y^*, s^*) be optimal solutions satisfying (45). We define

$$x' = x - \nu x^0 - (1-\nu)x^*,$$
$$y' = y - \nu y^0 - (1-\nu)y^*,$$
$$s' = s - \nu s^0 - (1-\nu)s^*.$$

From the feasibility conditions (48) and (49) of the perturbed problems (P_ν) and (D_ν), it is easily seen that (x', y', s') satisfies

$$Ax' = 0,$$
$$A^T y' + s' = 0.$$

This implies that x' and s' are orthogonal, i.e.,

$$\operatorname{tr}\big(\big(x - \nu x^0 - (1-\nu)x^*\big) \circ \big(s - \nu s^0 - (1-\nu)s^*\big)\big) = 0.$$

By expanding this equality, we obtain

$$\begin{aligned}\nu \operatorname{tr}\big(s^0 \circ x + x^0 \circ s\big) &= \operatorname{tr}(s \circ x) + \nu^2 \operatorname{tr}\big(s^0 \circ x^0\big) + (1-\nu)^2 \operatorname{tr}\big(s^* \circ x^*\big) \\ &\quad + \nu(1-\nu)\operatorname{tr}\big(s^0 \circ x^* + x^0 \circ s^*\big) - (1-\nu)\operatorname{tr}\big(s \circ x^* + s^* \circ x\big).\end{aligned}$$

We have $\operatorname{tr}(x^* \circ s^*) = 0$, because the triple (x^*, y^*, s^*) is optimal for (3) and (4). Furthermore, $\operatorname{tr}(x \circ s) = 2N\mu$, because the triple (x, y, s) is obtained after a full NT-step with respect to the problems (P_ν) and (D_ν). Moreover, since (x^0, y^0, s^0) is as in (44), we have $\operatorname{tr}(s^0 \circ x + x^0 \circ s) = \zeta \operatorname{tr}(x + s)$, $\operatorname{tr}(s^0 \circ x^0) = 2N\zeta^2$, and $\operatorname{tr}(s^0 \circ x^* + x^0 \circ s^*) = \zeta \operatorname{tr}(x^* + s^*)$. Due to (45) we have

$$\operatorname{tr}\big(x^* + s^*\big) \le \zeta \operatorname{tr}(e) = 2N\zeta.$$

Substitution gives

$$\begin{aligned}\nu\zeta \operatorname{tr}(x + s) &= 2N\mu + 2\nu^2 N\zeta^2 + (1-\nu)\big(\nu\zeta \operatorname{tr}\big(x^* + s^*\big) - \operatorname{tr}\big(s \circ x^* + s^* \circ x\big)\big) \\ &\le 2N\mu + 2\nu^2 N\zeta^2 + 2\nu(1-\nu)N\zeta^2 - (1-\nu)\operatorname{tr}\big(s \circ x^* + s^* \circ x\big) \\ &= 2N\mu + 2\nu N\zeta^2 - (1-\nu)\operatorname{tr}\big(s \circ x^* + s^* \circ x\big).\end{aligned}$$

Since x, s, x^* and s^* belong to $\mathcal{K}$, we have $\operatorname{tr}(s \circ x^* + s^* \circ x) \ge 0$, and since $\nu\zeta^2 = \mu$, we obtain $\nu\zeta \operatorname{tr}(x + s) \le 4\nu N\zeta^2$. By dividing both sides by $\nu\zeta$, the inequality in the lemma follows. □

Substitution of the bound in Lemma 4.9 into (79) gives

$$\|q\|_F \le 4\theta\rho(\delta)N. \tag{80}$$

4.5.6 Putting All Things Together

We proved in Sect. 4.5.2 (cf. Eqn. (67)) that, in order to have $\delta(v_f) \le \tau_f$, one should have

$$\omega(v)^2 \le \frac{(1-\theta)\tau_f}{\rho(\tau_f)}. \tag{81}$$

Then, in Lemma 4.5 (Sect. 4.5.3), we showed that

$$4\,\omega(v)^2 \le \|q\|_F^2 + \left(\|q\|_F + \sqrt{4(1-\theta)^2\delta^2 + 2\theta^2 N}\right)^2. \tag{82}$$

Due to this (81), certainly holds if

$$\|q\|_F^2 + \left(\|q\|_F + \sqrt{4(1-\theta)^2\delta^2 + 2\theta^2 N}\right)^2 \le 4\frac{(1-\theta)\tau_f}{\rho(\tau_f)}. \tag{83}$$

At the start of each main iteration we have $\delta \equiv \delta(v) \le \tau$. Since $\rho(\delta)$ is monotonically increasing in δ, it follows from (80) that $\|q\|_F \le 4\theta\rho(\tau)N$. Hence we conclude that if θ and τ satisfy the following inequality, then we certainly have $\delta(v_f) \le \tau_f$.

$$\left(4\theta\rho(\tau)N\right)^2 + \left(4\theta\rho(\tau)N + \sqrt{4(1-\theta)^2\tau^2 + 2\theta^2 N}\right)^2 \le 4\frac{(1-\theta)\tau_f}{\rho(\tau_f)}.$$

Since $\rho(\tau) \ge 1$ and $\tau_f \le \frac{1}{2}\rho(\tau_f)$, this inequality implies $2(4\theta N)^2 \le 2(1-\theta)$, whence $\theta < \frac{1}{4N}$. We may therefore write

$$\theta = \frac{\alpha}{N}, \quad 0 \le \alpha \le \frac{1}{4}. \tag{84}$$

The above inequality then reduces to

$$\left(4\alpha\rho(\tau)\right)^2 + \left(4\alpha\rho(\tau) + \sqrt{4\left(1-\frac{\alpha}{N}\right)^2\tau^2 + \frac{2\alpha^2}{N}}\right)^2 \le 4\frac{(1-\frac{\alpha}{N})\tau_f}{\rho(\tau_f)}.$$

This certainly holds, for all $N \ge 1$, if

$$\left(4\alpha\rho(\tau)\right)^2 + \left(4\alpha\rho(\tau) + \sqrt{4\,\tau^2 + 2\,\alpha^2}\right)^2 \le \frac{4(1-\alpha)\tau_f}{\rho(\tau_f)}. \tag{85}$$

4.6 Iteration Bound

Assuming that τ_f and τ are given, and the function Q be as defined in (42), according to (50) the number of required centering steps in each main iteration is the smallest positive integer such that

$$Q^k(\tau_f) \le \tau. \tag{86}$$

Table 1 Best lower bound if $\tau_f = 1/\sqrt{2}$ as a function of k

	τ	α	Bound	
1	0.707107	0.408248	0.058472	34.204611 N
2	0.707107	0.119523	0.141262	21.237170 N
3	0.707107	0.010103	0.163000	24.539823 N
4	0.707107	0.000072	0.164311	30.430070 N
5	0.707107	0.000000	0.164320	36.514089 N

With k fixed, we may assume that $\tau = Q^k(\tau_f)$. Apart from the k centering steps, each main iteration contains one feasibility step. Therefore, according to (51), the total number of inner iterations is bounded above by

$$\frac{1+k}{\theta} \log \frac{\max\{N\zeta^2, \|r_b^0\|, \|r_c^0\|\}}{\varepsilon}. \tag{87}$$

Since $\theta = \alpha/N$, we need to find values of α and τ that minimize $\frac{1+k}{\alpha}$, while satisfying (83) for each $N \geq 1$. This minimization problem has been numerically solved for several choices of k and τ_f, with $\tau = Q^k(\tau_f)$. For example, when doing this for $\tau_f = 1/\sqrt{2}$ and $1 \leq k \leq 5$, we find the results in Table 1. So, for $\tau_f = 1/\sqrt{2}$ the best upper bound is obtained for $k = 2$, and then at most 22 inner iterations are needed. We did similar computations for $\tau_f \in (0.2 : 0.000001 : 0.8)$, and concluded that the best lower bound obtained in this way is 17 and this occurs for $\tau_f \in [0.25752, 0.45397]$ and $k = 1$. As a consequence we may state our main result as follows.

Theorem 4.1 *Let* (3) *and* (4) *have optimal solution* (x^*, y^*, s^*) *such that* $x^* + s^* \preceq \zeta e$. *Then, by taking*

$$\tau_f = \frac{1}{3}, \qquad \tau = Q(\tau_f), \qquad \theta = \frac{\alpha}{N},$$

where α *is the largest value satisfying* (85), *the algorithm needs no more than*

$$17N \log \frac{\max\{N\zeta^2, \|r_b^0\|, \|r_c^0\|\}}{\varepsilon}$$

inner iterations to generate an ε*-solution of* (3) *and* (4).

Proof Since $\tau_f = \frac{1}{3}$ and $k = 1$, we have $\tau = Q(\tau_f) = 0.079057$. Substituting these values, with $\alpha = 2/17$, the value of the left-hand side of the inequality (83) becomes 0.805148, and the value of the right-hand side 0.847952. This implies the theorem. □

The above iteration bound has been derived under the assumption of the existence of optimal solutions x^* of (3) and (y^*, s^*) of (4) with vanishing duality gap and such that $x^* + s^* \preceq \zeta e$. One might ask what happens if this condition is not satisfied. One should note that, if after each feasibility step the proximity measure $\delta(x, s; \mu)$ does not exceed the value τ_f, then the algorithm will continue to run until it has found an ε-solution of (3) and (4). If, however, during the course of the algorithm it happens that, after some feasibility step the proximity measure $\delta(x, s; \mu)$ exceeds the value τ_f, then it tells us that the above assumptions are violated, which means that either the problems (3) and (4) do not have optimal solutions with vanishing duality gap less than or equal to ε, or the value of ζ has been chosen too small. In the latter case one might run the algorithm once more with a larger value of ζ.

5 Concluding Remarks

The first contribution of this paper is a new primal–dual IPM for solving SOCO problems that uses full NT-steps only. So no line searches are required. Then, using the method proposed first in [26] for LO (see also [38, 39]), and that was extended to SDO and linear complementarity problems (LCPs) in [27, 40, 41], we extended this algorithm to an infeasible primal–dual IPM for SOCO that uses full NT-steps only. In both cases the order of the iterations bounds coincide with the currently best known iteration bounds for SOCO. Our bounds do not use the order symbol, however, and as such they are better than all known iteration bounds for SOCO. Even more, when realizing that $\mathbb{R}^n_+$ can be considered as the direct product of n one-dimensional second-order cones, the bounds turn into iteration bounds for an IPM and an IIPM for LO, and these bounds improve existing bounds.

A more challenging task is to unify the analysis for LO, SOCO and SDO by considering optimization problems over general symmetric cones. Another topic for further research is to consider large-update variants of the algorithm, since such methods are much more efficient in practice. Finally, a question that might be considered is whether full step methods (either of Newton or NT-type) can be made efficient by using dynamic updates of the barrier parameter. This will not improve the theoretical complexity, but it will enhance the practical performance of the algorithm significantly.

Acknowledgements Authors wish to thank Professor Florian Potra and four anonymous referees for useful comments and suggestions on an earlier draft of the manuscript. The first author would like to thank for the financial grant from Shahrekord University. The first author was also partially supported by the Center of Excellence for Mathematics, University of Shahrekord, Shahrekord, Iran.

Appendix A: Technical Lemmas

Lemma A.1 *For $i = 1, \ldots, n$, let $f_i : \mathbb{R}_+ \to \mathbb{R}$ denote a convex function. Then we have, for any nonzero vector $z \in \mathbb{R}^n_+$, the following inequality*:

$$\sum_{i=1}^{n} f_i(z_i) \le \frac{1}{e^T z} \sum_{j=1}^{n} z_j \Bigl(f_j\bigl(e^T z\bigr) + \sum_{i \ne j} f_i(0) \Bigr).$$

Proof We define the function $F : \mathbb{R}^n_+ \to \mathbb{R}$ by

$$F(z) = \sum_{i=1}^{n} f_i(z_i), \quad z \ge 0.$$

Letting e_j denote the jth unit vector in $\mathbb{R}^n$, we may write z as a convex combination of the vectors $(e^T z)e_j$, as follows.

$$z = \sum_{j=1}^{n} \frac{z_j}{e^T z}\big(e^T z\big)e_j,$$

Indeed, $\sum_{j=1}^{n} \frac{z_j}{e^T z} = 1$ and $z_j/e^T z \geq 0$ for each j. Since $F(z)$ is a sum of convex functions, $F(z)$ is convex in z, and hence we have

$$F(z) \leq \sum_{j=1}^{n} \frac{z_j}{e^T z} F\big(\big(e^T z\big)e_j\big) = \sum_{j=1}^{n} \frac{z_j}{e^T z} \sum_{i=1}^{n} f_i\big(\big(e^T z\big)(e_j)_i\big).$$

Since $(e_j)_i = 1$ if $i = j$ and zero if $i \neq j$, we obtain

$$F(z) \leq \sum_{j=1}^{n} \frac{z_j}{e^T z}\Bigg(f_j\big(e^T z\big) + \sum_{i \neq j} f_i(0)\Bigg).$$

Hence the inequality in the lemma follows. □

Corollary A.1 *Let $f : \mathbb{R}_+ \to \mathbb{R}$ be a convex function such that $f(0) = 0$. Then we have, for any vector $z \in \mathbb{R}^n_+$, the following inequality*:

$$\sum_{i=1}^{n} f(z_i) \leq f\Bigg(\sum_{i=1}^{n} z_i\Bigg).$$

Proof In Lemma A.1, take $f_i = f$ for each i; then the result follows. □

In the lemma below, we use that the cone $\mathcal{K}$ of squares in a Euclidean Jordan algebra defines a partial ordering $\preceq_{\mathcal{K}}$ of $\mathbb{R}^n$ according to the definition

$$x \preceq_{\mathcal{K}} s \quad \Leftrightarrow \quad s - x \in \mathcal{K}.$$

Lemma A.2 *Let $x, s \in \mathbb{R}^n$ and $x^T s = 0$, then one has*

(i) $-\frac{1}{4}\|x+s\|_F^2 e \preceq_{\mathcal{K}} x \circ s \preceq_{\mathcal{K}} \frac{1}{4}\|x+s\|_F^2 e$;
(ii) $\|x \circ s\|_F \leq \frac{1}{2\sqrt{2}}\|x+s\|_F^2$.

Proof We write

$$x \circ s = \frac{1}{4}\big((x+s)^2 - (x-s)^2\big). \tag{88}$$

Since $(x+s)^2$ is a square, it belongs to $\mathcal{K}$. This means that $x \circ s + \frac{1}{4}(x-s)^2 \in \mathcal{K}$, or, equivalently,

$$x \circ s \succeq_{\mathcal{K}} -\frac{1}{4}(x-s)^2.$$

Since $x \preceq_{\mathcal{K}} \|x\|_F e$ for every x, also using Lemma 2.5(i), we may write

$$(x-s)^2 \preceq_{\mathcal{K}} \|(x-s)^2\|_F e \preceq_{\mathcal{K}} \|x-s\|_F^2 e,$$

whence it follows that

$$x \circ s \succeq_{\mathcal{K}} -\frac{1}{4}(x-s)^2 \succeq_{\mathcal{K}} -\frac{1}{4}\|x-s\|_F^2 e.$$

In the same way one derives from (88) that

$$x \circ s \preceq_{\mathcal{K}} \frac{1}{4}(x+s)^2 \preceq_{\mathcal{K}} \frac{1}{4}\|x+s\|_F^2 e.$$

Thus we have shown that one has, for all $x, s \in \mathbb{R}^n$,

$$-\frac{1}{4}\|x-s\|_F^2 e \preceq_{\mathcal{K}} x \circ s \preceq_{\mathcal{K}} \frac{1}{4}\|x+s\|_F^2 e.$$

Since x and s are orthogonal, we have $\mathrm{tr}(x \circ s) = 2x^T s = 0$, whence $\|x+s\|_F = \|x-s\|_F$. Hence part (i) of the lemma follows.

For the proof of (ii) we return to (88). Using $\|z\|_F^2 = \mathrm{tr}(z^2)$, we obtain

$$\begin{aligned}\|x \circ s\|_F^2 &= \left\|\frac{1}{4}\big((x+s)^2 - (x-s)^2\big)\right\|_F^2 = \frac{1}{16}\mathrm{tr}\big[\big((x+s)^2 - (x-s)^2\big)^2\big] \\ &= \frac{1}{16}\big[\mathrm{tr}\big((x+s)^4\big) + \mathrm{tr}\big((x-s)^4\big) - 2\mathrm{tr}\big((x+s)^2 \circ (x-s)^2\big)\big].\end{aligned}$$

Since $(x+s)^2$ and $(x-s)^2$ belong to $\mathcal{K}$, the trace of their product is nonnegative. Thus we obtain

$$\|x \circ s\|_F^2 \leq \frac{1}{16}\big[\mathrm{tr}\big((x+s)^4\big) + \mathrm{tr}\big((x-s)^4\big)\big] = \frac{1}{16}\big[\big\|(x+s)^2\big\|_F^2 + \big\|(x-s)^2\big\|_F^2\big].$$

Using Lemma 2.5(i) and $\|x+s\|_F = \|x-s\|_F$ again, we get

$$\|x \circ s\|_F^2 \leq \frac{1}{16}\big[\|x+s\|_F^4 + \|x-s\|_F^4\big] = \frac{1}{8}\|x+s\|_F^4.$$

This implies (ii). Hence the proof of the lemma is complete. □

References

1. Lobo, M.S., Vandenberghe, L., Boyd, S.E., Lebret, H.: Applications of second-order cone programming. Linear Algebra Appl. **284**, 193–228 (1998)
2. Vanderbei, R.J., Yurttan, H.: Using LOQO to solve second-order cone programming problems. Report sor 98-09, Princeton University, Princeton, USA (1998)
3. Wolkowicz, H., Saigal, R., Vandenberghe, L. (eds.): Handbook of Semidefinite Programming. International Series in Operations Research & Management Science, vol. 27. Kluwer, Boston (2000)
4. Ji, J., Potra, F.A., Sheng, R.: On the local convergence of predictor-corrector method for semidefinite programming. SIAM J. Optim. **10**(1), 195–210 (1999)

5. Ji, J., Potra, F.A., Sheng, R.: On a general class of interior-point algorithms for semidefinite with polynomial complexity and super convergence programming. Methods Appl. Anal. **6**(4) (1999)
6. Potra, F.A., Sheng, R.: A superlinearly convergent primal–dual infeasible-interior-point algorithm for semidefinite programming. SIAM J. Optim. **8**(4), 1007–1028 (1998)
7. Potra, F.A., Sheng, R.: On homogeneous interior-point algorithms for semidefinite programming. Optim. Methods Softw. **9**, 161–184 (1998)
8. Nesterov, Y.E., Nemirovskii, A.S.: Interior Point Polynomial Methods in Convex Programming: Theory and Algorithms. SIAM, Philadelphia (1993)
9. Nesterov, Y.E., Todd, M.J.: Self-scaled barriers and interior-point methods for convex programming. Math. Oper. Res. **22**(1), 1–42 (1997)
10. Nesterov, Y.E., Todd, M.J.: Primal–dual interior-point methods for self-scaled cones. SIAM J. Optim. **8**(2), 324–364 (1998)
11. Adler, I., Alizadeh, F.: Primal–dual interior point algorithms for convex quadratically constrained and semidefinite optimization problems. Technical Report RRR-111-95, Rutger Center for Operations Research, Brunswick, NJ (1995)
12. Alizadeh, F., Goldfarb, D.: Second-order cone programming. Math. Program. **95**, 3–51 (2003)
13. Schmieta, S.H., Alizadeh, F.: Associative and Jordan algebras, and polynomial time interior-point algorithms for symmetric cones. Math. Oper. Res. **26**(3), 543–564 (2001)
14. Schmieta, S.H., Alizadeh, F.: Extension of polynimal time interior-point algorithms to symmetric cones. Math. Program. **96**(3), 409–438 (2003)
15. Faybusovich, L.: Linear system in Jordan algebras and primal–dual interior-point algorithms. J. Comput. Appl. Math. **86**(1), 149–175 (1997)
16. Peng, J., Roos, C., Terlaky, T.: Self-regular functions and new search directions for linear and semidefinite optimization. Math. Program. **93**, 129–171 (2002)
17. Peng, J., Roos, C., Terlaky, T.: Self-Regularity. A New Paradigm for Primal–Dual Interior-Point Algorithms. Princeton University Press, Princeton (2002)
18. Bai, Y.Q., Wang, G.Q., Roos, C.: Primal–dual interior-point algorithms for second-order cone optimization based on kernel functions. Nonlinear Anal. **70**(10), 3584–3602 (2009)
19. Kojima, M., Megiddo, N., Mizuno, S.: A primal–dual infeasible-interior-point algorithm for linear programming. Math. Program. **61**, 263–280 (1993)
20. Zhang, Y.: On the convergence of a class of infeasible-interior-point methods for the horizontal linear complementarity problem. SIAM J. Optim. **4**, 208–227 (1994)
21. Mizuno, S.: Polynomiality of infeasible interior point algorithms for linear programming. Math. Program. **67**, 109–119 (1994)
22. Potra, F.A.: A quadratically convergent predictor-corrector method for solving linear programs from infeasible starting points. Math. Program. **67**, 383–406 (1994)
23. Potra, F.A.: An infeasible-interior-point predictor-corrector algorithm for linear programming. SIAM J. Optim. **6**(1), 19–32 (1996)
24. Zhang, Y.: On extending primal–dual interior-point algorithms from linear programming to semidefinite programming. SIAM J. Optim. **8**(2), 365–386 (1998)
25. Rangarajan, B.K.: Polynomial convergence of infeasible-interior-point methods over symmetric cones. SIAM J. Optim. **16**(4), 1211–1229 (2006)
26. Roos, C.: A full-Newton step $O(n)$ infeasible interior-point algorithm for linear optimization. SIAM J. Optim. **16**(4), 1110–1136 (2006)
27. Mansouri, H.: Full-Newton step interior-point methods for conic optimization. PhD thesis, ISBN 978-90-9023179-2. Faculty of Electrical Engineering, Mathematics and Computer Science, The Netherlands (2008)
28. Faraut, J., Koranyi, A.: Analysis on symmetric cones. Oxford University Press, New York (1994)
29. Koecher, M.: The Minnesota Notes on Jordan Algebras and Their Applications. Lecture Notes in Mathematics, vol. 1710. Springer, Berlin (1999). Edited, annotated and with a preface by Aloys Krieg and Sebastian Walcher
30. McCrimmon, K.: A Taste of Jordan Algebras. Universitext. Springer, New York (2004)
31. Vieira, M.V.C.: Jordan algebraic approach to symmetric optimization. PhD thesis, ISBN 978-90-6464-189-3. Faculty of Electrical Engineering, Mathematics and Computer Science, The Netherlands (2007)
32. Lim, Y.: Geometric means on symmetric cones. Arch. Math. (Basel) **75**(1), 39–56 (2000)
33. Sturm, J.: Similarity and other spectral relations for symmetric cones. Linear Algebra Appl. **312**(1–3), 135–154 (2000)

34. Monteiro, R.D.C., Tsuchiya, T.: Polynomial convergence of primal–dual algorithms for the second-order program based on the MZ-family of directions. Math. Program. **88**, 61–83 (2000)
35. Roos, C., Terlaky, T., Vial, J.-P.: Theory and Algorithms for Linear Optimization. An Interior-Point Approach. Wiley, Chichester (1997) (2001). (2nd edn. Springer, Berlin (2005))
36. Mizuno, S., Todd, M.J., Ye, Y.: A surface of analytic centers and primal–dual infeasible-interior-point algorithms for linear programming. Math. Oper. Res. **20**(1), 135–162 (1995)
37. Freund, R.M.: An infeasible-start algorithm for linear programming whose complexity depends on the distance from the starting point to the optimal solution. Ann. Oper. Res. **62**, 29–57 (1996)
38. Mansouri, H., Roos, C.: A simplified $O(nL)$ infeasible interior-point algorithm for linear optimization using full Newton steps. Optim. Methods Softw. **22**(3), 519–530 (2007)
39. Mansouri, H., Zangiabadi, M.: An adaptive infeasible interior-point algorithm with full-Newton step for linear optimization. J. Optim. (2011). doi:10.1080/02331934.2011.611881
40. Mansouri, H., Roos, C.: A new full-Newton step $O(n)$ infeasible interior-point algorithm for semidefinite optimization. J. Numer. Algorithms **52**(2), 225–255 (2009)
41. Mansouri, H., Zangiabadi, M., Pirhaji, M.: A full-Newton step $O(n)$ infeasible interior-point algorithm for linear complementarity problems. Nonlinear Anal., Real World Appl. **12**, 545–561 (2011)