Mathematical Programming (2022) 194:341–370
https://doi.org/10.1007/s10107-021-01631-4

FULL LENGTH PAPER

Series A

A primal-dual interior-point algorithm for nonsymmetric exponential-cone optimization

Joachim Dahl[1] · **Erling D. Andersen**[1]

Received: 15 August 2019 / Accepted: 15 February 2021 / Published online: 9 March 2021
© The Author(s) 2021

Abstract
A new primal-dual interior-point algorithm applicable to nonsymmetric conic optimization is proposed. It is a generalization of the famous algorithm suggested by Nesterov and Todd for the symmetric conic case, and uses primal-dual scalings for nonsymmetric cones proposed by Tunçel. We specialize Tunçel's primal-dual scalings for the important case of 3 dimensional exponential-cones, resulting in a practical algorithm with good numerical performance, on level with standard symmetric cone (*e.g.*, quadratic cone) algorithms. A significant contribution of the paper is a novel higher-order search direction, similar in spirit to a Mehrotra corrector for symmetric cone algorithms. To a large extent, the efficiency of our proposed algorithm can be attributed to this new corrector.

Keywords Nonsymmetric cone optimization · Exponential-cone optimization · Interior-point methods

Mathematics Subject Classification 90C99

1 Introduction

In 1984 Karmarkar [11] presented an interior-point algorithm for linear optimization with polynomial complexity. This triggered the interior-point revolution which gave rise to a vast amount research on interior-point methods. A particularly important result was the analysis of so-called *self-concordant* barrier functions, which led to polynomial-time algorithms for linear optimization over a convex domain with a self-concordant barrier, provided that the barrier function can be evaluated in polynomial

The material in this manuscript has not published or submitted elsewhere.

✉ Joachim Dahl
joachim.dahl@mosek.com

[1] MOSEK ApS, Copenhagen, Denmark

time. This was proved by Nesterov and Nemirovski [19], and as a consequence convex optimization problems with such barriers can be solved efficently by interior-point methods, at least in theory.

However, numerical studies for linear optimization quickly demonstrated that primal-dual interior-point methods were superior in practice, which led researchers to generalize the primal-dual algorithm to general smooth convex problems. A major breakthrough in that direction is the seminal work by Nesterov and Todd (NT) [20,21] who generalized the primal-dual algorithm for linear optimization to *self-scaled* cones, with the same complexity bound as in the linear case. Güler [8] later showed that the self-scaled cones correspond to the class of symmetric cones, which has since become the most commonly used term. The good theoretical performance of primal-dual methods for symmetric cones has since been confirmed computionally, *e.g.*, by [2,28].

The class of symmetric cones has been completely characterized and includes 5 different cones, where the three most interesting ones are the nonnegative orthant, the quadratic cone, and the cone of symmetric positive semidefinite matrices, as well as products of those three cones. Although working exclusively with the symmetric cones is a limitation, they cover a great number of important applications, see, *e.g.*, Nemirovski [16] for an excellent survey of conic optimization over symmetric cones.

Some convex sets with a symmetric cone representation are more naturally characterized using nonsymmetric cones, *e.g.*, semidefinite matrices with chordal sparsity patterns, see [32] for an extensive survey. Thus algorithmic advancements for handling nonsymmetric cones directly could hopefully lead to both simpler modeling and reductions in computational complexity. Many other important convex sets cannot readily be modeled using symmetric cones, *e.g.*, convex sets involving exponentials or logarithms, for which no representation using symmetric cones is known. In terms of practical importance, the three dimensional power- and exponential-cones are perhaps the most important nonsymmetric cones; Lubin et al. [12] showed how all instances in a large benchmark library can modeled using the three symmetric cones as well the three dimensional power- and exponential cone.

Generalizing methods from symmetric to nonsymmetric cones is not straightforward, however. In [22] Nesterov et al. suggest a long-step algorithm using both the primal and dual barriers effectively doubling the size of the linear system solved at each iteration. For small-dimensional cones (such as the exponential cone) this overhead might be acceptable.

More recently Nesterov [18] proved the existence of a NT-like primal-dual scaling in the vicinity of the central path, leading to an algorithm that uses only a single barrier, but is restricted to following the central path closely. Compared to [22] this method has a main advantage of reducing the size of the linear system to the same size of the algorithms for symmetric cones; a similar advantage is shared by algorithms using explicit primal-dual scalings, *e.g.*, scalings by Tunçel [30] considered later. At each iteration Nesterov's method [18] has a centering phase, which brings the current iterate close to the central path, which is followed by one affine step which brings the iterate closer to optimum. From a practical perspective this is a significant drawback of the method since the centering phase is computationally costly.

Hence, the centering and affine steps should be combined as in the symmetric cone case. Also both algorithms [18,22] are feasible methods, *i.e.*, they require either a

strictly feasible known starting point or some sort of phase-I method to get a feasible starting point. Skajaa and Ye [27] extended Nesterov's method [18] with a homogeneous model, which also simplifies infeasibility detection. In practice, however, the method of Skaajaa and Ye is not competitive with other methods [3] due to the centering steps.

In more recent work Serrano and Ye [26] improves the algorithm in [27] such that explicit centering steps are not needed, but instead restricting iterates to a vicinity of the central path. This method has been implemented as part of the ECOS solver [6] and will be used for comparison in Sect. 8.

A different approach to non-symmetric conic optimization was proposed by Tunçel [30], who extends the concepts of the NT algorithm in a more direct fashion. Nesterov and Todd [20,21] showed existence of a primal-dual scaling defined by a single scaling point w satifying two secant equations $s = F''(w)x$ and $F'(x) = F''(w)F'_*(s)$, where F' and F'' denotes the first- and second-order derivatives of the barrier. Furthermore, the scaling $F''(w)$ is shown to be bounded, which is a key property of establishing polynomial-time complexity of the NT algorithm. Tunçel later showed in [30] that such a scaling point w exists for *any* convex cone K, but satisfying only one secant equation $s = F''(w)x$, and if the barrier has negative curvature then w is unique [23]. To enforce both secant equations in the nonsymmetric case, Tunçel [30] considered a sequence of low-rank quasi-Newton updates to a general positive definite matrix.

These ideas were further explored by Myklebust and Tunçel [15] and also in [14] and form the basis of our proposed algorithm. Following this line of work, the essential difference from a symmetric NT algorithm is the computation of a general positve definite scaling matrix, satisfying the same secant equations, but without relying on a given scaling-point w. Furthermore, such scaling matrices should be bounded to ensure polynomial-time complexity.

For three-dimensional cones these scaling matrices are particularly simple and characterized by a single scalar, as shown in Sect. 5. Uniform boundedness of the scaling matrices is not established in this correspondence, but we comment on an efficient method for computing the most bounded scaling matrix using the formulations developed in Sect. 5.

It is also possible to develop algorithms for convex optimization specified on functional form. This has been done by several authors, for example by [3,4,7,9] who all solve the KKT optimality conditions of the problem in functional form. The algorithms all require a sort of merit or penalty function, which often require problem specific parameter tuning to work well in practice. Another strain of research is the work Nemirovski and Tunçel [17] and very recently Karimi and Tunçel [10] who advocate a non-linear convex formulation instead of a conic formulations, explicitly using self-concordant barriers for the convex domains.

From a theoretical point of view the different algorithms for nonsymmetric cones (including the functional formulations) all share the same best-known complexity bounds as the symmetric counterparts. Whether these methods are competetive in practice with algorithms for symmetric cones is still an unanswered question, though.

The algorithm we consider herein uses the scaling matrices by Tunçel [30], resulting in an algorithm that is similar to the symmetric counterpart; the linear system solved at each iterations is very similar and both the residuals and the complementarity gap

decrease at the same rate. The algorithm is a natural extension of the Nesterov-Todd algorithm implemented for symmetric cones in, *e.g.*, SeDuMi [28] and MOSEK [3].

It is well known that the Mehrotra predictor-corrector idea [13] leads to vastly improved computational performance in the symmetric cone case. One of our main contributions is a new corrector for nonsymmetric cones, closely tied to Tunçel's primal dual scalings. It is derived from a second-order approximation of the centrality condition $s_\mu = -\mu F'(x_\mu)$ and thus involves third-order directional derivatives. The proposed corrector loosely follows the central path, without restricting the size of the neighborhood, thereby allowing the algorithm to take longer steps, and it shares similarities with the standard Mehrotra corrector; in particular they coincide if $F(x) = -\sum_i \log x_i$ denotes the standard barrier for linear optimization.

We demonstrate numerically that the proposed corrector offers a substantial and consistent reduction in the number of iterations required to solve the problems in all our numerical studies. Skajaa and Ye [27] suggest a different corrector by characterizating the full central path as a differential equation, solved by a Runge-Kutta method. An immediate draw-back of Skajaa's Runge-Kutta method is that it requires additional factorizations of the full KKT system, which significantly adds to the overall solution time.

The remaining paper is structured as follows. We define basic properties for the exponential cone in Sect. 2, and we discuss the homogeneouous model, the central path and related metrics in Sect. 3. In Sect. 4 we discuss search-directions assuming a primal-dual scaling is known, and we derive a our new corrector and provide a simple numerical example that illustrates how the corrector algorithm makes more steady progress. In Sect. 5 we discuss new characterizations of the primal-dual scalings from [15,30], which reduce to univariate characterizations for three-dimensional cones.

In Sect. 6 we give a collected overview of the suggested path-following algorithm. This is followed by a discussion of some implementation details. Next in Sect. 8 we present numerical results on a moderately large collection of exponential-cone problems. We conclude in Sect. 9, and in the appendix we give details of the first-, second- and third-order derivatives of the barrier for the exponential cone.

2 Preliminaries

In this section we list well-known properties of self-concordant and self-scaled barriers, which are used in the remainder of the paper. The proofs can be found in references such as [20,21]. We consider a pair of primal and dual linear conic problems

$$\begin{aligned} &\text{minimize } \langle c, x\rangle \\ &\text{subject to } Ax = b \\ &\qquad\quad x \in K, \end{aligned} \tag{P}$$

and

$$\begin{aligned} &\text{maximize } \langle b, y\rangle \\ &\text{subject to } c - A^T y = s \\ &\qquad\quad s \in K^*, \end{aligned} \tag{D}$$

where $y \in \mathbf{R}^m$, $s \in \mathbf{R}^n$ and $K \subset \mathbf{R}^n$ is a proper cone, *i.e.*, a pointed, closed, convex cone with non-empty interior. We assume throughout the paper that A has full rank, *i.e.*, the rows A are linearly independent. The dual cone K^* is

$$K^* = \{z \in \mathbf{R}^n \mid \langle x, z\rangle \geq 0, \forall x \in K\}.$$

If K is proper then K^* is also proper. A cone K is called self-dual if there is positive definite map between K and K^*, *i.e.*, if $TK = K^*$, $T \succ 0$. A function $F : \mathbf{int}(K) \mapsto \mathbf{R}$, $F \in C^3$ is a ϑ-logarithmically homogeneouos self-concordant barrier (ϑ-LHSCB) for $\mathbf{int}(K)$ if

$$|F'''(x)[u, u, u]| \leq 2(F''(x)[u, u])^{3/2}$$

and

$$F(\tau x) = F(x) - \vartheta \log \tau$$

holds for all $x \in \mathbf{int}(K)$ and for all $u \in \mathbf{R}^n$. For a pointed cone $\vartheta \geq 1$. We will refer to a LHSCB simply as a self-concordant barrier. If F_1 and F_2 are ϑ_1 and ϑ_2-self-concordant barriers for K_1 and K_2, respectively, then $F_1(x_1) + F_2(x_2)$ is a $(\vartheta_1 + \vartheta_2)$-self-concordant barrier for $K_1 \times K_2$. Some straightforward consequences of the homogeneouos property include

$$\begin{aligned} F'(\tau x) &= \frac{1}{\tau} F'(x), \quad F''(\tau x) = \frac{1}{\tau^2} F''(x), \\ F''(x)x &= -F'(x), \quad F'''(x)x = -2F''(x), \\ \langle F'(x), x\rangle &= -\vartheta. \end{aligned}$$

If F is a ϑ-self-concordant barrier for K, then the Fenchel conjugate

$$F_*(s) = \sup_{x \in \mathbf{int}(K)} \{-\langle s, x\rangle - F(x)\} \tag{1}$$

is a ϑ-self-concordant barrier for K^*. Futhermore, if $(x, s) \in \mathbf{int}(K) \times \mathbf{int}(K^*)$ then $(-F'(x), -F'_*(s)) \in \mathbf{int}(K^*) \times \mathbf{int}(K)$.

For a ϑ-self-concordant barrier, the so-called Dikin ellipsoid

$$E(x; r) = \{z \in \mathbf{R}^n \mid \langle F''(x)(z - x), z - x\rangle \leq r\}$$

is included in the cone for $r < 1$, *i.e.*, $E(x, r) \subset \mathbf{int}(K)$ for $r < 1$, and F is almost quadratic inside this ellipsoid,

$$(1 - r)^2 F''(x) \preceq F''(z) \preceq \frac{1}{(1 - r)^2} F''(x)$$

for all $z \in E(x, r)$.

A cone is called self-scaled if it has a ϑ-self-concordant barrier F such that for all $w, x \in \mathbf{int}(K)$,

$$F''(w)x \in \mathbf{int}(K^*)$$

and

$$F_*(F''(w)) = F(x) - 2F(w) - \vartheta.$$

Self-scaled cones are equivalent to symmetric cones and they satisfy the stronger long-step Hessian estimation property

$$\frac{1}{(1+\alpha\sigma_x(-p))^2}F''(x) \preceq F''(x-\alpha p) \preceq \frac{1}{(1-\alpha\sigma_x(p))^2}F''(x)$$

for any $\alpha \in [0; \sigma_x(p)^{-1})$ where

$$\sigma_x(p) := \frac{1}{\sup\{\alpha \ : \ x - \alpha p \in K\}}$$

denotes the distance to the boundary. Many properties of symmetric cones follow from the fact that the barriers have negative curvature $F'''(x)[u] \preceq 0$ for all $x \in \mathbf{int}(K)$ and all $u \in K$. An interesting property proven in [23] is that if both the primal and dual barrier has negative curvature then the cone is symmetric.

In addition to the three symmetric cones (*i.e.*, the nonnegative orthant, the quadratic cone and the cone of symmetric positive semidefinite matrices) we mainly consider the nonsymmetric exponential cone studied by Charez [5] in the present work.

$$K_{\text{exp}} = \text{cl}\{x \in \mathbf{R}^3 \mid x_1 \geq x_2 \exp(x_3/x_2), \ x_2 > 0\},$$

with a 3-self-concordant barrier,

$$F(x) = -\log(x_2 \log(x_1/x_2) - x_3) - \log x_1 - \log x_2. \tag{2}$$

The dual exponential cone is

$$K^*_{\text{exp}} = \text{cl}\{z \in \mathbf{R}^3 \mid e \cdot z_1 \geq -z_3 \exp(z_2/z_3), \ z_1 > 0, z_3 < 0\}.$$

The exponential cone is not self-dual, but $T K_{\text{exp}} = K^*_{\text{exp}}$ for

$$T = \begin{bmatrix} e & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} \succ 0.$$

For the exponential cone the conjugate barrier $F_*(s)$ or its derivatives cannot be evaluated on closed-form, but it can be evaluated numerically to high accuracy (*e.g.*, with a damped Newton's method) using the definition (1), *i.e.*, if

$$x_s = \arg\min\{-\langle s, x\rangle - F(x) \ : \ x \in \mathbf{int}(K)\}$$

then

$$F_*'(s) = -x_s, \quad F_*''(s) = \left[F''(x_s)\right]^{-1}.$$

We conclude this survey of introductory material by listing some of the many convex sets that can be represented using the exponential cone, or a combination of exponential cones and symmetric cones. The epigraph $t \geq e^x$ can be modelled as $(t, 1, x) \in K_{\exp}$ and similarly for the hypograph of the logarithm $t \leq \log x \Leftrightarrow (x, 1, t) \in K_{\exp}$. The hypograph of the entropy function, $t \leq -x \log x$ is equivalent to $(1, x, t) \in K_{\exp}$, and similarly for relative entropy $t \geq x \log(y/x) \Leftrightarrow (y, x, -t) \in K_{\exp}$. The *soft-plus function* $\log(1 + e^x)$ can be thought of as a smooth approximation of $\max\{0, x\}$. Its epigraph can be modelled as $t \geq \log(1+e^x) \Leftrightarrow u+v = 1,\ (u, 1, x-t), (v, 1, -t) \in K_{\exp}$. The epigraph of the logarithm of a sum exponentials can modelled as

$$t \geq \log(e^{x_1} + \cdots + e^{x_1}) \quad \Longleftrightarrow \quad \sum_{i=1}^{n} u_i = 1,\ (u_i, 1, x_i - t) \in K_{\exp}, i = 1, \ldots, n,$$

These examples all have auxiliary variables and constraints in their conic representations, which might suggest that an algorithm working directly with a barrier of the convex domain (*e.g.*, [10]) is more efficient. However, a conic formulation has the advantage of nicer conic duality, and it is easy to exploit the special (sparse) structure from the additional constraints and variables in the linear algebra implementation, thereby eliminating the overhead in a conic formulation.

3 The homogeneous model and central path

In the simplified homogeneous model we embed the KKT conditions for (P) and (D) into the homogeneous self-dual model

$$\begin{gathered} \begin{bmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{bmatrix} \begin{bmatrix} y \\ \hat{x} \\ \tau \end{bmatrix} - \begin{bmatrix} 0 \\ \hat{s} \\ \kappa \end{bmatrix} = 0 \\ \hat{x} \in K_1 \times \cdots \times K_k, \quad \hat{s} \in K_1^* \times \cdots \times K_k^*, \quad y \in \mathbf{R}^m, \quad \tau, \kappa \geq 0, \end{gathered} \tag{3}$$

where $\hat{x} = (x_1, \ldots, x_k)$ is a concatenation of conic variables, $x_i \in K_i$. We denote the primal-dual variables by $(\hat{x}, \hat{s})$ to distinguish them from augmented variables defined next. Let

$$x_{k+1} := \tau, \quad s_{k+1} := \kappa, \quad K_{k+1} := \mathbf{R}_+, \quad F_{k+1}(x_{k+1}) := -\log(x_{k+1}), \quad \vartheta_{k+1} = 1.$$

We than have

$$x := (x_1, \ldots, x_{k+1}) \in K, \qquad s := (s_1, \ldots, s_{k+1}) \in K^*.$$

where $K := K_1 \times \cdots \times K_{k+1}$ has a barrier

$$F(x) = \sum_{i=1}^{k+1} F_i(x_i).$$

with complexity $\vartheta = \sum_{i=1}^{k+1} \vartheta_i$. Let $z := (x, s, y)$ and define

$$G(z) := \begin{bmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{bmatrix} \begin{bmatrix} y \\ \hat{x} \\ \tau \end{bmatrix} - \begin{bmatrix} 0 \\ \hat{s} \\ \kappa \end{bmatrix}.$$

The KKT conditions can then be expressed succinctly as

$$G(z) = 0, \quad z \in \mathcal{D},$$

where $\mathcal{D} := K \times K^* \times \mathbf{R}^m$. Given an initial $z^0 \in \mathbf{int}(\mathcal{D})$ we consider a central path z_μ as the solution to

$$G(z_\mu) = \mu G(z^0) \tag{4}$$

$$s_\mu = -\mu F'(x_\mu), \quad x_\mu = -\mu F'_*(s_\mu), \tag{5}$$

parametrized by $\mu \in (0, 1]$, and on the central path we have

$$\langle x_\mu, s_\mu \rangle / \vartheta = \mu.$$

The following lemma gives an equivalent variational characterization of the central path.

Lemma 1 *Given* $z^0 \in \mathbf{int}(\mathcal{D})$. *Let*

$$\Psi(z) := \langle x^0, s \rangle + \langle x, s^0 \rangle + F(x) + F_*(s).$$

Then

$$z_\mu := \arg\min_{z \in \mathcal{D}} \{\Psi(z) \ : \ G(z) = \mu G(z^0)\}.$$

We omit the proof, which follows from the optimality conditions for minimizing $\Psi(z)$. In [22] the central path is defined from the variational characterization in Lemma 1, and they prove that the definition in (4)–(5) is equivalent. From the variational characterization z_μ is well-defined, and $\lim_{\mu \to 0} z_\mu = z^\star$ satisfies (see, *e.g.*, [22]),

1. $\langle x^\star, s^\star \rangle = 0$.
2. If $\tau^\star > 0$ then $\hat{x}^\star / \tau^\star$ is an optimal solution for (P) and $(y^\star, \hat{s}^\star)/\tau^\star$ is an optimal solution for (D).

3. If $\kappa^\star > 0$ then $\langle b, y^\star \rangle > 0$ and (P) is infeasible, or $\langle c, \hat{x}^\star \rangle < 0$ and (D) is infeasible, or both.

In the following neighborhood definitions and subsequent algorithms we consider iterates (x, s, y) that are generally not on the central path, and we define

$$\mu := \langle x, s \rangle / \vartheta.$$

A neighborhood used by Skajaa and Ye [27] is then

$$\|s + \mu F'(x)\|_x^* = \langle s + \mu F'(x), F''(x)^{-1}(s + \mu F'(x)) \rangle^{1/2} \leq \beta\mu, \tag{6}$$

which characterizes the central path for $\beta = 0$. We can think of (6) as a generalization of the standard two-norm neighborhood

$$\|XSe - \mu e\|_2 \leq \beta\mu$$

from linear optimization [33].

A different neighborhood is due to Nesterov and Todd [21]. We define *shadow iterates* (following [30])

$$\tilde{x} := -F_*'(s), \quad \tilde{s} := -F'(x) \tag{7}$$

and

$$\tilde{\mu} := \langle \tilde{x}, \tilde{s} \rangle / \vartheta$$

for an iterate $(x, s, y) \in \mathcal{D}$. Nesterov and Todd [21] then showed that $\mu\tilde{\mu} \geq 1$ with equality only on the central path. This leads to a different neighborhood $\beta\mu\tilde{\mu} \leq 1$ for $\beta \in (0; 1]$, or equivalently

$$\beta\mu\langle \tilde{x}, \tilde{s} \rangle \leq \vartheta.$$

This is satisfied if

$$\beta\mu\langle \tilde{x}_i, \tilde{s}_i \rangle \leq \vartheta_i, \ i = 1, \ldots, k + 1,$$

leading to another neighborhood definition

$$\mathcal{N}(\beta) = \left\{ (x, s) \in K \times K^* \mid \vartheta_i \langle F'(x_i), F_*'(s_i) \rangle^{-1} \geq \beta\mu, \ i = 1, \ldots, k + 1 \right\},$$

which (in contrast (6)) characterizes the central path for $\beta = 1$. We use the neighborhood $\mathcal{N}(\beta)$, which can be seen as a generalization of the one-sided ∞-norm neighborhood.

Both the central path and the neighborhood depend on the initial values. A simple choice is $y^0 = 0$ and

$$x^0 = s^0 = -F'(x^0),$$

which are optimality conditions for minimizing

$$f(x) := (1/2)x^T x + F(x).$$

If $K_i = K_{\exp}$ this can be solved off-line using a backtracking Newton's method to get

$$[x^0]_i = [s^0]_i \approx (1.290928, 0.805102, -0.827838).$$

For the symmetric cones and the three-dimensional power-cone such a central starting point can be found analytically. Then $\langle x^0, s^0\rangle/\vartheta = 1$ and $(x^0, s^0) \in \mathcal{N}(1)$.

4 Search-directions using a primal-dual scaling

In this section we define search directions, assuming a primal-dual scaling is known; how to compute such scalings is discussed in Sect. 5. We consider an iterate $(x, s, y) \in \mathbf{int}(\mathcal{D})$ and nonsingular primal-dual scalings W_i satisfying double secant equations

$$W_i x_i = W_i^{-T} s_i, \qquad W_i \tilde{x}_i = W_i^{-T} \tilde{s}_i, \quad i = 1, \ldots, k+1$$

where $\tilde{x}_i$ and $\tilde{s}_i$ are the shadow iterates defined in (7). Let

$$W := \begin{bmatrix} W_1 & & \\ & \ddots & \\ & & W_{k+1} \end{bmatrix}$$

We can then express the primal-dual scaling succinctly as

$$v = Wx = W^{-T} s, \quad \tilde{v} = W\tilde{x} = W^{-T}\tilde{s}. \tag{8}$$

where $W_{k+1} := \sqrt{\kappa/\tau}$ and

$$v_{k+1} = \sqrt{\tau\kappa}, \qquad \tilde{v}_{k+1} = \frac{1}{\sqrt{\tau\kappa}}.$$

Whenever we encounter a scaling W in the remainder of this paper, it is assumed that W satisfies the double secant equation (8) for the current iterate.

To linearize the centrality condition we consider

$$s + \Delta s = -\mu F'(x + \Delta x),$$

with a linearization given by

$$s + \Delta s = -\mu F'(x) - \mu F''(x)\Delta x. \tag{9}$$

On the central path we can express (9) as

$$W\Delta x + W^{-T}\Delta s = \mu\tilde{v} - v \tag{10}$$

with $W := [\mu F''(x)]^{-1/2}$. If (x, s) is not on the central path then (10) is an *approximate* linearization of the symmetric centrality condition

$$v = \mu\tilde{v}$$

with a quality determined by the distance $\|W^T W - \mu F''(x)\|$. Thus, the assumption that $W^T W \approx \mu F''(x)$ is important for our proposed algorithm, including the corrector studied later in this section.

We next define different search-directions used in the proposed algorithm. The affine search-direction is the solution to

$$G(\Delta z^{\mathrm{a}}) = -G(z), \qquad W\Delta x^{\mathrm{a}} + W^{-T}\Delta s^{\mathrm{a}} = -v, \tag{11}$$

and is characterized by the following lemma.

Lemma 2 *The solution to* (11) *satisfies*

$$\langle s, \Delta x^a\rangle + \langle x, \Delta s^a\rangle = -\langle x, s\rangle, \qquad \langle \Delta x^a, \Delta s^a\rangle = 0, \tag{12}$$

and for all $\alpha \in \mathbf{R}$,

$$\langle x + \alpha\Delta x^a, s + \alpha\Delta s^a\rangle = (1-\alpha)\langle x, s\rangle.$$

Proof It follows from (11) that

$$\langle s, \Delta x^{\mathrm{a}}\rangle + \langle x, \Delta s^{\mathrm{a}}\rangle = -\langle v, v\rangle = -\langle x, s\rangle,$$

and skew-symmetry implies that

$$-\langle [y + \Delta y^{\mathrm{a}},\ x + \Delta x^{\mathrm{a}}], G(z + \Delta z^{\mathrm{a}})\rangle = \langle x + \Delta x^{\mathrm{a}}, s + \Delta s^{\mathrm{a}}\rangle = 0,$$

which combined shows that $\langle \Delta x^{\mathrm{a}}, \Delta s^{\mathrm{a}}\rangle = 0$. The last part follows directly from (12). □

Lemma 2 shows that a full affine step step $(z + \Delta z^{\mathrm{a}})$ satisfies both $G(z + \Delta z^{\mathrm{a}}) = 0$ and $\langle x + \Delta x^{\mathrm{a}}, s + \Delta s^{\mathrm{a}}\rangle = 0$. Thus, if $(z + \Delta z^{\mathrm{a}}) \in \mathcal{D}$ then $(z + \Delta z^{\mathrm{a}})$ is optimal (*i.e.*, a solution to (3)).

We next consider a higher-order corrector for nonsymmetric cones, with some similarities to a Mehrotra corrector for symmetric cones. We consider the first- and second-order derivatives of $s_\mu = -\mu F'(x_\mu)$ with respect to μ,

$$\begin{aligned} \dot{s}_\mu + \mu F''(x_\mu)\dot{x}_\mu &= -F'(x_\mu), \\ \ddot{s}_\mu + \mu F''(x_\mu)\ddot{x}_\mu &= -2F''(x_\mu)\dot{x}_\mu - \mu F'''(x_\mu)[\dot{x}_\mu, \dot{x}_\mu] \end{aligned} \tag{13}$$

From (13) we have

$$\mu\dot{x}_\mu = -[F''(x_\mu)]^{-1}(F'(x_\mu) + \dot{s}_\mu) = x_\mu - [F''(x_\mu)]^{-1}\dot{s}_\mu,$$

resulting in an expression for the third-order directional derivative,

$$\mu F'''(x_\mu)[\dot{x}_\mu, \dot{x}_\mu] = F'''(x_\mu)[\dot{x}_\mu, x_\mu] - F'''(x_\mu)[\dot{x}_\mu, (F''(x_\mu))^{-1}\dot{s}_\mu] \tag{14}$$

$$= -2F''(x_\mu)\dot{x}_\mu - F'''(x_\mu)[\dot{x}_\mu, (F''(x_\mu))^{-1}\dot{s}_\mu] \tag{15}$$

where (15) follows from the homogeneity property $F'''(x)[x] = -2F''(x)$. This results in an alternative expression for the second-order derivative of the centrality condition, *i.e.*,

$$\ddot{s}_\mu + \mu F''(x_\mu)\ddot{x}_\mu = F'''(x_\mu)[\dot{x}_\mu, (F''(x_\mu))^{-1}\dot{s}_\mu].$$

Assuming that $W^T W \approx \mu F''(x)$ and comparing (11) and (13) we interpret $\Delta s^{\mathrm{a}} = -\dot{s}_\mu/\mu$ and $\Delta x^{\mathrm{a}} = -\dot{x}_\mu/\mu$, leading to the definition of our corrector term

$$\eta := -\frac{1}{2}F'''(x)[\Delta x^{\mathrm{a}}, (F''(x))^{-1}\Delta s^{\mathrm{a}}]. \tag{16}$$

Thus a pure corrector search-direction can be defined as the solution to

$$G(\Delta z^{\mathrm{c}}) = 0, \qquad W\Delta x^{\mathrm{c}} + W^{-T}\Delta s^{\mathrm{c}} = -W^{-T}\eta, \tag{17}$$

and satisfies properties given in the following lemma.

Lemma 3 *The solution to* (17) *satisfies*

$$\langle s, \Delta x^c\rangle + \langle x, \Delta s^c\rangle = 0, \qquad \langle \Delta x^c, \Delta s^c\rangle = 0.$$

Proof From (17) we have that

$$\langle s, \Delta x^{\mathrm{c}}\rangle + \langle x, \Delta s^{\mathrm{c}}\rangle = (1/2)\langle x, F'''(x)[\Delta x^{\mathrm{a}}, (F''(x))^{-1}\Delta s^{\mathrm{a}}]\rangle = -\langle \Delta x^{\mathrm{a}}, \Delta s^{\mathrm{a}}\rangle = 0,$$

using the homogeneity property $F'''(x)[x] = -2F''(x)$, and skewsymmetry implies that $\langle \Delta x^{\mathrm{c}}, \Delta s^{\mathrm{c}}\rangle = 0$. □

We note that

$$W_{k+1}^2 \Delta x_{k+1}^{\mathrm{c}} - \Delta s_{k+1}^{\mathrm{c}} = -\eta_{k+1}$$

reduces to the familiar expression $\kappa\Delta\tau^{\mathrm{c}} + \tau\Delta\kappa^{\mathrm{c}} = -\Delta\tau^{\mathrm{a}}\Delta\kappa^{\mathrm{a}}$.

For a given centering parameter $\gamma > 0$ we define a combined search-direction as the solution to

$$G(\Delta z) = -(1-\gamma)G(z), \qquad W\Delta x + W^{-T}\Delta s = -v + \gamma\mu\tilde{v} - W^{-T}\eta, \tag{18}$$

with properties given in the following lemma.

Lemma 4 *The solution to* (18) *satisfies*

$$\langle s, \Delta x\rangle + \langle x, \Delta s\rangle = 0, \qquad \langle \Delta x, \Delta s\rangle = 0$$

and for all $\alpha \in \mathbf{R}$,

$$\begin{aligned} G(z + \alpha\Delta z) &= (1 - \alpha(1-\gamma))G(z), \\ \langle x + \alpha\Delta x, s + \alpha\Delta s\rangle &= (1 - \alpha(1-\gamma))\langle x, s\rangle. \end{aligned}$$

Proof From (18) and Lemma 3 we have that

$$\begin{aligned} \langle s, \Delta x\rangle + \langle x, \Delta s\rangle &= -(1-\gamma)\langle x, s\rangle - (1/2)\langle x, F'''(x)[\Delta x^{\mathrm{a}}, (F''(x))^{-1}\Delta s^{\mathrm{a}}]\rangle \\ &= -(1-\gamma)\langle x, s\rangle, \end{aligned}$$

and skewsymmetry implies that

$$\begin{aligned} \langle (1-\gamma)x + \Delta x, (1-\gamma)s + \Delta s\rangle = (1-\gamma)^2\langle x, s\rangle + (1-\gamma)[\langle s, \Delta x\rangle + \langle x, \Delta s\rangle] \\ + \langle \Delta x, \Delta s\rangle = 0, \end{aligned}$$

i.e., $\langle \Delta x, \Delta s\rangle = 0$. The last part now follows. □

The search-direction (18) forms the basis of our algorithm. For a given step-size $\alpha \in (0, 1]$ the residuals and the complementarity gap decrease at the same rate. More explicitly, for $\mu^k := \langle x^k, s^k\rangle/\vartheta$ the residuals are the kth iteration are

$$G(z^k) = \mu^k G(z^0),$$

which should be compared with central path definition in Lemma 1. Similarly, the complementarity gap at the kth iteration is

$$\langle x^k, s^k\rangle = \mu^k\vartheta.$$

This is in contrast with other methods [18,26,27], which do not decrease the complementarity gap at the same rate. Also, no explicit merit function as in [9] is required to ensure a balanced decrease of the residuals and the complementarity gap.

Tunçel [30] showed polynomial complexity of an infeasible method (without a corrector) assuming boundedness of the scaling matrices. These results were further extended by Myklebust and Tunçel [15] to include an analysis for scaling matrices obtained by a BFGS scaling (such scalings are considered in the next section). Although we use slightly different scaling matrices, their analysis could be be applied in a small neighborhood around the central path. Thus a short-step algorithm using the search directions herein would likely inherit good theoretical performance. It is also possibility that future studies will prove a conjucture that the scaling matrices considered herein are bounded, which would simplify a complexity analysis.

To gain some insight into the corrector η we note that in the case of the nonnegative orthant we have the familiar expression

$$\frac{1}{2}F'''(x)[\Delta x^{\mathrm{a}}, (F''(x))^{-1}\Delta s^{\mathrm{a}}] = -\mathbf{diag}(x)^{-1}\mathbf{diag}(\Delta x^{\mathrm{a}})\Delta s^{\mathrm{a}},$$

and similarly for the semidefinite cone we have

$$\begin{aligned}\frac{1}{2}F'''(x)[\Delta x^{\mathrm{a}}, (F''(x))^{-1}\Delta s^{\mathrm{a}}] &= -\frac{1}{2}x^{-1}\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}} - \frac{1}{2}\Delta s^{\mathrm{a}}\Delta x^{\mathrm{a}}x^{-1} \\ &= -(x^{-1}) \circ (\Delta x^{\mathrm{a}}\Delta s^{\mathrm{a}}),\end{aligned}$$

using the generalized product associated with the Euclidean Jordan algebra, see, *e.g.*, [31] for a discussion of Euclidean Jordan algebra in a similar context as ours. For the Lorentz cone we have

$$F'''(x)[(F''(x))^{-1}u] = -\frac{2}{x^T Qx}(ux^T Q + Qxu^T - (x^T u)Q)$$

with $Q = \mathbf{diag}(1, -1, \ldots, -1)$. Let e_k denote the kth standard basis-vector (*i.e.*, the vector with value 1 in position k and 0 elsewhere). Then

$$F'''(x)[(F''(x))^{-1}u]e_1 = -2(x^{-1}) \circ u,$$

again using the notation of the generalized product [31]. We defer the derivation and implementation specific details for the exponential cone to the appendix.

As an illustration of the proposed corrector we consider a simple example,

$$\begin{aligned}&\text{minimize } x_1 + x_2 \\ &\text{subject to } x_1 + x_2 + x_3 = 1 \\ &\qquad (x_1, x_2, x_3) \in K_{\exp}.\end{aligned} \tag{19}$$

In Table 1 we list the complementarity gap $\langle x^k, s^k\rangle$ for the iterates produced using the search direction (18) *with* and *without* the proposed corrector, using the scaling matrices defined in Sect. 5.

In Fig. 1 we plot the same iterates $x^k/\tau^\star$ projected onto the hyperplane $x_1+x_2+x_3 = 1$. Although some iterates appear close to the boundary they are all within the defined neighborhood, *i.e.*, no cut-back is performed to stay within the neighborhood. We see how the corrector algorithm makes significantly more progress and thereby reduces the required number of iterations. A similar observation is made for a wide selection of test-problems in Sect. 8.

5 Primal-dual scalings

In this section we review key results for primal-dual scalings by Tunçel [30], we make connections to the theory of multiple secant-equation updates by Schnabel [25], and

Table 1 Complementarity gap for the iterates in example (19), with and without corrector

k	W. corrector	W/o. corrector
0	4.0e+00	4.0e+00
1	9.3e−01	1.2e+00
2	4.3e−02	3.6e−01
3	2.1e−04	7.6e−02
4	2.7e−08	8.7e−-03
5	5.5e−10	4.8e−03
6	8.4e−15	1.1e−03
7		5.3e−06
8		8.5e−07
9		1.3e−07
10		1.0e−08
11		1.2e−10
12		1.2e−13

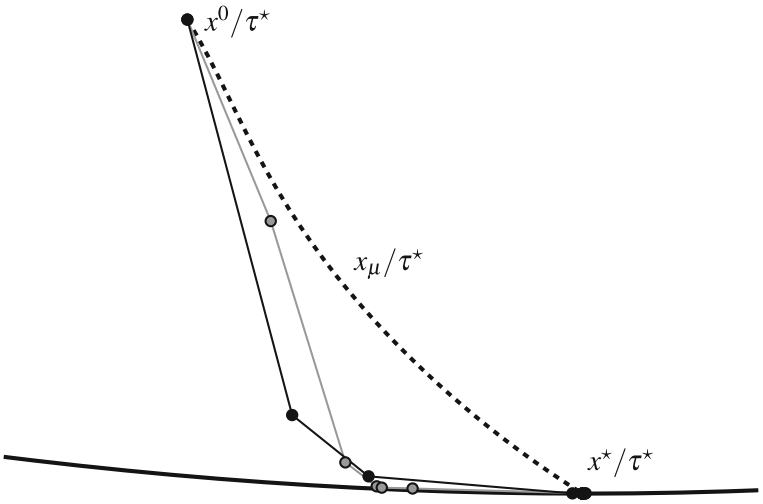

Fig. 1 Central path, boundary, and iterates for example (19), projected onto the hyperplane $x_1+x_2+x_3 = 1$. The iterates using the corrector direction are plotted in the darker solid line, and iterates without the corrector are plotted in ligher solid line

we present new formulations, which were used collaboratively in [24] to compute optimally bounded scaling matrices specifically for three-dimensional cones.

Tunçel [30] defines a set of scalings

$$\mathcal{T}_1(x,s) := \{T \ : \ T \succ 0,\ T^2 s = x,\ T^2 F'(x) = F'_*(s)\}, \tag{20}$$

i.e., positive definite scalings satisfying double secant equations. Tunçel further defines a set of bounded scalings parametrized by ξ,

$$\mathcal{T}_2(x,s,\xi) := \left\{ T \in \mathcal{T}_1(x,s) \ : \ (\xi\delta_F)^{-1} F''_*(s) \preceq T^2 \preceq \xi\delta_F \left[F''(x)\right]^{-1} \right\}, \quad (21)$$

where $\delta_F := [\vartheta(\mu\tilde{\mu} - 1) + 1]/\mu$. Let

$$\xi^\star := \inf_{\xi} \mathcal{T}_2(x,s,\xi). \quad (22)$$

In the case where $\xi^\star \in \mathcal{O}(1)$ over all $(x,s) \in \mathbf{int}(K) \times \mathbf{int}(K^*)$, Tunçel proved an iteration complexity bound of $\mathcal{O}(\sqrt{\vartheta}\log(1/\epsilon))$ for an infeasible-start primal-dual interior point algorithm (coinciding with the best known complexity bound for interior-point methods). The parameter δ_F and set $\mathcal{T}_2(x,s,\xi)$ are further addressed towards the end of this section in the context of finding optimally bounded scaling matrices.

A self-scaled cone has a unique Nesterov-Todd scaling point w satisfying double secant equations,

$$s = F''(w)x, \qquad F'(x) = F''(w)F'_*(s).$$

Furthermore, $F''(w)$ is bounded (see [20,21,30]) in the sense that

$$[F''(w)]^{-1} \in \mathcal{T}_2(x,s,4/3).$$

A barrier $F(x)$ is said to have negative curvature if for all $x \in \mathbf{int}(K)$ and for $u \in K$ we have

$$F'''(x)[u] \preceq 0$$

and $F(x)$ is self-scaled if and only if $F(x)$ and $F_*(s)$ both have negative curvature [23]. Barriers with negative curvature (but which are not self-scaled) still have a unique scaling point w satisfying exactly one secant equation

$$s = F''(w)x$$

but not the other. The exponential-cone barrier does not have negative curvature, as can be seen be considering $\hat{x} := (1, e^{-2}, 0) \in K_{\exp}$ and $\hat{u} := (1,0,0) \in K_{\exp} \setminus \mathbf{int}(K_{\exp})$. Then it can be verified using the expressions in the appendix that

$$F'''(\hat{x})[\hat{u}] = \begin{bmatrix} -4 & \frac{e^2}{2} & \frac{e^2}{2} \\ \frac{e^2}{2} & 0 & 0 \\ \frac{e^2}{2} & 0 & -\frac{e^4}{4} \end{bmatrix},$$

which is indefinite, for example

$$F'''(\hat{x})[\hat{u},\hat{v},\hat{v}] = 4$$

for $\hat{v} := (1, 8e^{-2}, 4e^{-2})$.

Thus for general nonsymmetric cones the essential question is how to define bounded scaling matrices satisfying both secant equations, without relying on a scaling point w. Tunçel [30] partly answers that question by deriving scalings $T \in \mathcal{T}_1(x, s)$ using BFGS update equations, resulting in a rank-4 update to a given positive definite matrix. It is still instructive, however, to review work by Schnabel on quasi-Newton methods with multiple secant equations. In particular, the following theorem from [25] is used repeatedly in the following discussion.

Theorem 1 *Let $S, Y \in \mathbf{R}^{n\times p}$ have full rank p. Then there exists $H \succ 0$ such that $HS = Y$ if and only if $Y^T S \succ 0$.*

As a consequence we can write any such $H \succ 0$ as

$$H = Y(Y^T S)^{-1} Y^T + ZZ^T, \quad S^T Z = 0, \quad \mathbf{rank}(Z) = n - p$$

or in factored form $H = W^T W$ with

$$W = \left[\, Y(Y^T S)^{-1/2},\ Z \,\right]^T.$$

One may verify that given any $\Omega \succ 0$ satisfying $\Omega S = Y$, the following identify holds

$$W^{-1} = \left[\, S(Y^T S)^{-1/2},\ \Omega^{-1} Z(Z^T \Omega^{-1} Z)^{-1} \,\right],$$

and therefore

$$H^{-1} = S(Y^T S)^{-1} S^T + RR^T, \quad Y^T R = 0, \quad \mathbf{rank}(R) = n - p,$$

where $RR^T = \Omega^{-1} Z(Z^T \Omega^{-1} Z)^{-2} Z^T \Omega^{-1}$. One of the most popular quasi-Newton update rules is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) step in (23), where $H \succ 0$ denotes an approximation of the Hessian.

$$H_{\mathrm{BFGS}} := Y(Y^T S)^{-1} Y^T + H - HS(S^T HS)^{-1} S^T H \tag{23}$$

It is well known (see, *e.g.*, [25]) that the update (23) is the solution to

$$\begin{array}{ll} \text{minimize} & \|\Omega^{1/2}(H_+^{-1} - H^{-1})\Omega^{1/2}\|_F \\ \text{subject to} & H_+^{-1} Y = S, \\ & H_+^{-1} \succ 0 \end{array} \tag{24}$$

for any $\Omega \succ 0$ satisfying $\Omega S = Y$. In the following theorem we show how (23) can be computed using a sequence of updates, similar to quasi-Newton updates with a single secant equation.

Theorem 2 *Given $Y_0, S_0 \in \mathbf{R}^{n\times p}$ with $Y_0^T S_0 \succ 0$. Then*

$$\begin{aligned} Y_0(Y_0^T S_0)^{-1} Y_0^T &= VV^T \\ S_0(Y_0^T S_0)^{-1} S_0^T &= UU^T, \end{aligned}$$

where $V := \begin{pmatrix} v_1 & \cdots & v_p \end{pmatrix}$, $U := \begin{pmatrix} u_1 & \cdots & u_p \end{pmatrix}$ *and*

$$v_k := \frac{Y_{k-1}e_k}{\langle Y_{k-1}e_k, S_{k-1}e_k\rangle^{1/2}}, \qquad Y_k := Y_{k-1} - v_k v_k^T S_{k-1} \tag{25}$$

$$u_k := \frac{S_{k-1}e_k}{\langle Y_{k-1}e_k, S_{k-1}e_k\rangle^{1/2}}, \qquad S_k := S_{k-1} - u_k u_k^T Y_{k-1} \tag{26}$$

for $k = 1, \ldots, p$.

Proof The proof is constructive and follows from a Cholesky factorization of $Y_0^T S_0$. Let

$$L := \left(\frac{Y_0^T S_0 e_1}{\langle Y_0 e_1, S_0 e_1\rangle^{1/2}}, \cdots, \frac{Y_{p-1}^T S_{p-1} e_p}{\langle Y_{p-1}e_p, S_{p-1}e_p\rangle^{1/2}} \right).$$

In the proof we first show that $Y_0^T S_0 = LL^T$, and we then show second that $LV^T = Y_0^T$, which in turn implies (25).

We start by showing that the recursion (25), (26) is well-defined. Let Ψ_k be the principal submatrix obtained from the last $p - k$ rows and columns of $Y_k^T S_k$, where $\Psi_0 = Y_0^T S_0 \succ 0$ by assumption. Expanding $Y_k^T S_k$ we have

$$Y_k^T S_k = Y_{k-1}^T S_{k-1} - \frac{(Y_{k-1}^T S_{k-1}e_k)(Y_{k-1}^T S_{k-1}e_k)^T}{\langle Y_{k-1}e_k, S_{k-1}e_k\rangle}, \tag{27}$$

i.e., Ψ_k is the Schur-complement of the first element of Ψ_{k-1} and therefore positive definite.

We next make a simplifying observation, namely that the first k columns of Y_k and S_k are zero. From (25), (26) we immediately have that $Y_k e_k = S_k e_k = 0, k = 1, \ldots p$, and that sparsity propagates to subsequent steps, *i.e.*, $Y_j e_k = S_j e_k = 0$, $j > k$.

We can now prove that $LL^T = Y_0^T S_0$. We have

$$\begin{aligned} LL^T e_k &= \sum_{i=0}^{k-1} \frac{Y_i^T S_i e_{i+1} (Y_i^T S_i e_{i+1})^T e_k}{\langle Y_i e_{i+1}, S_i e_{i+1}\rangle} \\ &= \sum_{i=0}^{k-2} \frac{Y_i^T S_i e_{i+1} (Y_i^T S_i e_{i+1})^T e_k}{\langle Y_i e_{i+1}, S_i e_{i+1}\rangle} + Y_{k-1}^T S_{k-1} e_k. \end{aligned} \tag{28}$$

Repeated use of (27) in (28) then shows that

$$\begin{aligned} LL^T e_k &= \sum_{i=0}^{k-2} \frac{Y_i^T S_i e_{i+1} (Y_i^T S_i e_{i+1})^T e_k}{\langle Y_i e_{i+1}, S_i e_{i+1}\rangle} + Y_{k-1}^T S_{k-1} e_k \\ &= \sum_{i=0}^{k-3} \frac{Y_i^T S_i e_{i+1} (Y_i^T S_i e_{i+1})^T e_k}{\langle Y_i e_{i+1}, S_i e_{i+1}\rangle} + Y_{k-2}^T S_{k-2} e_k = \cdots = Y_0^T S_0 e_k. \end{aligned}$$

We finally show that $LV^T = Y_0^T$. From repeated use of (25) it follows that

$$\begin{aligned} LV^T &= Y_0^T - Y_0^T + \frac{S_0^T Y_0 e_1 (Y_0 e_1)^T}{\langle Y_0 e_1, S_0 e_1\rangle} + \frac{S_1^T Y_1 e_2 (Y_1 e_2)^T}{\langle Y_1 e_2, S_1 e_2\rangle} + \dots \\ &\quad + \frac{S_{p-1}^T Y_{p-1} e_p (Y_{p-1} e_p)^T}{\langle Y_{p-1}^T e_p, S_{p-1} e_p\rangle} \\ &= Y_0^T - Y_1^T + \frac{S_1^T Y_1 e_2 (Y_1 e_2)^T}{\langle Y_1 e_2, S_1 e_2\rangle} + \dots + \frac{S_{p-1}^T Y_{p-1} e_p (Y_{p-1} e_p)^T}{\langle Y_{p-1}^T e_p, S_{p-1} e_p\rangle} \\ &= \dots = Y_0^T - Y_p^T = Y_0^T. \end{aligned}$$

Since $LV^T = Y_0^T$ we have that $Y_0 (LL^T)^{-1} Y_0^T = Y_0 (Y_0^T S_0)^{-1} Y_0^T = VV^T$. Equation (26) follows similarly. □

Primal-dual scalings can then be derived similarly to Tunçel [30] and Myklebust and Tunçel [15]. In our context we derive non-singular (factored) scalings W satisfying

$$W^{-T} s = Wx, \qquad -W^{-T} F'(x) = -W F'_*(s), \tag{29}$$

i.e., $(W^T W)^{-1} \in \mathcal{T}_1(x, s)$ defined in (20). We define

$$H := \mu F''(x), \qquad S := \left[x, \tilde{x} \right], \qquad Y := \left[s, \tilde{s} \right],$$

where we remind the reader that $\tilde{x} := -F'_*(s)$ and $\tilde{s} := -F'(x)$.

The condition $Y^T S \succ 0$ is equivalent to assuming that (x, s) is not on the central path[1]. We next define

$$\delta_x := x - \mu \tilde{x}, \qquad \delta_s := s - \mu \tilde{s}$$

also used in [15]. To compute (23) we first use Theorem 2 with $S_0 := S$, $Y_0 := Y$ resulting in

$$Y (S^T Y)^{-1} Y^T = \frac{s s^T}{\langle x, s\rangle} + \frac{\delta_s \delta_s^T}{\langle \delta_x, \delta_s\rangle}.$$

We next define

$$\rho_x := \tilde{x} - \frac{\langle x, H\tilde{x}\rangle}{\langle x, Hx\rangle} x.$$

If we use Theorem 2 again, this time with $S_0 := S$ and $Y_0 := HS$ we get

$$HS (S^T HS)^{-1} S^T H = \frac{Hx (Hx)^T}{\langle x, Hx\rangle} + \frac{H\rho_x (H\rho_x)^T}{\langle \rho_x, H\rho_x\rangle},$$

[1] If (x, s) is on the central path, then $W^T W = \mu F''(x)$ is a scaling with $(W^T W)^{-1} \in \mathcal{T}_2(x, s, 1)$, see (21).

leading to an expression for the BFGS update as a rank 4 update to $H \succ 0$ (for a general H),

$$H_{\text{BFGS}} = H + \frac{ss^T}{\langle x, s\rangle} + \frac{\delta_s \delta_s^T}{\langle \delta_x, \delta_s\rangle} - \frac{Hx(Hx)^T}{\langle x, Hx\rangle} - \frac{H\rho_x(H\rho_x)^T}{\langle \rho_x, H\rho_x\rangle}.$$

Considering (24), we see that the BFGS update to $H := \mu F''(x)$ has the desirable property of minimizing

$$\|W^{-1}W^{-T} - (\mu F''(x))^{-1}\|_\Omega,$$

measured in a weighted norm; for simplicity we can assume that $\Omega = W^T W$. With this choice of H we have

$$H_{\text{BFGS}} = \mu F''(x) + \frac{ss^T}{\langle x, s\rangle} + \frac{\delta_s \delta_s^T}{\langle \delta_x, \delta_s\rangle} - \frac{\mu}{\vartheta}\tilde{s}\tilde{s}^T - \mu\frac{(F''(x)\tilde{x} - \tilde{\mu}\tilde{s})(F''(x)\tilde{x} - \tilde{\mu}\tilde{s})^T}{\langle \tilde{x}, F''(x)\tilde{x}\rangle - \vartheta\tilde{\mu}^2},$$

which curiously reduces to a rank 3 update,

$$\begin{aligned} H_{\text{BFGS}} = \mu F''(x) &+ \frac{1}{2\mu\vartheta}\delta_s(s + \mu\tilde{s} + \frac{1}{\mu\tilde{\mu} - 1}\delta_s)^T + \frac{1}{2\mu\vartheta}(s + \mu\tilde{s} + \frac{1}{\mu\tilde{\mu} - 1}\delta_s)\delta_s^T \\ &-\mu\frac{(F''(x)\tilde{x} - \tilde{\mu}\tilde{s})(F''(x)\tilde{x} - \tilde{\mu}\tilde{s})^T}{\langle \tilde{x}, F''(x)\tilde{x}\rangle - \vartheta\tilde{\mu}^2}. \end{aligned} \tag{30}$$

We conclude this section by considering the three-dimensional case, for which the expressions simplify significantly. It follows from Theorem 1 that any scaling (29) has the form

$$W^T W = Y(Y^T S)^{-1}Y^T + tzz^T, \qquad W^{-1}W^{-T} = S(Y^T S)^{-1}S^T + t^{-1}rr^T$$

where $S^T z = 0, Y^T r = 0$ and $\langle r, z\rangle = 1$, *i.e.*, the scaling is essentially characterized by $t > 0$. For simplicity, we assume $\|z\| = 1$. We compute z and r using cross-products,

$$z = \frac{x \otimes \tilde{x}}{\|x \otimes \tilde{x}\|}, \qquad r = \frac{s \otimes \tilde{s}}{\langle s \otimes \tilde{s}, z\rangle}.$$

In the three dimensional case we can devise a simple algorithm for finding scalings achieving the bound (22). For notational convinience we introduce $Q := \begin{bmatrix} r, & S \end{bmatrix}$, which is nonsingular. We can then solve

$$\inf_\xi \left\{ \xi \ : \ (\xi\delta_F)^{-1} Q F''(x) Q^T \preceq \begin{bmatrix} t & 0 \\ 0 & Y^T S \end{bmatrix} \preceq (\xi\delta_F) Q[F''_*(s)]^{-1} Q^T \right\} \tag{31}$$

using a simple bisection algorithm. Consider the monotonically decreasing function,

$$\xi^l(t) := \inf_\xi \left\{ \xi \ : \ (\xi\delta_F)^{-1} Q F''(x) Q^T \preceq \begin{bmatrix} t & 0 \\ 0 & Y^T S \end{bmatrix} \right\},$$

and the monotonically increasing increasing function

$$\xi^u(t) := \inf_{\xi} \left\{ \xi \ : \ (\xi\delta_F) Q[F_*''(s)]^{-1} Q^T \succeq \begin{bmatrix} t & 0 \\ 0 & Y^T S \end{bmatrix} \right\}.$$

Given upper and lower bounds on t, the solution to (31) can then be found using bisection on t to solve $\xi^l(t) = \xi^u(t)$. Such a bisection method was considered in [24], where it was conjectured that $\xi^\star \approx 1.253$ for the exponential cone.

The BFGS scaling corresponds to

$$t = \mu \left\| F''(x) - \frac{\tilde{s}\tilde{s}^T}{\vartheta} - \frac{(F''(x)\tilde{x} - \tilde{\mu}\tilde{s})(F''(x)\tilde{x} - \tilde{\mu}\tilde{s})^T}{\langle \tilde{x}, F''(x)\tilde{x}\rangle - \vartheta\tilde{\mu}^2} \right\|_F. \tag{32}$$

We have tried both the optimal scaling and the BFGS scaling for all numerical test problems in Sect. 8, without noticing a significant difference in required number of iterations or quality of the solution. For all the problems the largest observed bound on ξ was 1.72 for the BFGS scaling; for simplicity we only report results for the simpler BFGS scaling in the following. In preliminary experiments we have observed similarly encouraging numerical results for higher dimensional non-symmetric cones using the BFGS scaling. The bisection algorithm is still valuable as a reference, however, and we hope that future studies will prove the conjecture, and possibly derive similar bounds for other three dimensional cones (including a tighter bound than $4/3$ for quadratic cones).

6 A primal-dual algorithm for exponential cone optimization

In this section we give a collected overview of the suggested path-following primal-dual algorithm. The algorithm is specialized for three-dimensional cones (in particular, the exponential cone) by using cross-products for computing the BFGS scalings. By computing the scaling matrices as a general rank 3 update (30) the algorithm is readily adapted to other non-symmetric cones.

We fix β to a constant low value, for example $\beta = 10^{-6}$. The different essential parts of the method are *i*) finding a starting point, *ii*) computing a search-direction and step-size, and *iii*) checking the stopping criteria for termination.

– *Starting point*. Find a starting point on the central path

$$x = s = -F'(x)$$

and $y = 0$, $\tau = \kappa = 1$. Then $z := (\hat{x}, \hat{s}, y, \tau, \kappa) \in \mathcal{N}(1)$.

– *Scaling matrices*. Compute BFGS scaling matrices

$$W_k = \left[\frac{s_k}{\sqrt{\langle x_k, s_k\rangle}}, \ \frac{\delta_{s_k}}{\sqrt{\langle \delta_{x_k}, \delta_{s_k}\rangle}}, \ \sqrt{t_k} \cdot z_k \right]^T, \quad W_k^{-1} = \left[\frac{x_k}{\sqrt{\langle x_k, s_k\rangle}}, \ \frac{\delta_{x_k}}{\sqrt{\langle \delta_{x_k}, \delta_{s_k}\rangle}}, \ \frac{r_k}{\sqrt{t_k}} \right]^T$$

where $z_k = (x_k \otimes \tilde{x}_k)/\|x_k \otimes \tilde{x}_k\|$, $r_k = (s_k \otimes \tilde{s}_k)/\langle s_k \otimes \tilde{s}_k, z_k\rangle$ and t_k is chosen from (32). Note that $\{z_k\}$ and z are unrelated; the later denotes the aggregation of all primal and dual variables.

– *Search-direction and step-size*. Compute an affine direction Δz^{a} as the solution to (11),

$$G(\Delta z^{\mathrm{a}}) = -G(z), \quad W\Delta x^{\mathrm{a}} + W^{-T}\Delta s^{\mathrm{a}} = -v.$$

From Δz^{a} we compute a corrector (16),

$$\eta := -\frac{1}{2}F'''(x)[\Delta x^{\mathrm{a}}, (F''(x))^{-1}\Delta s^{\mathrm{a}}],$$

similar to Mehrotra [13], where details on evaluting the derivatives are given in the appendix, see (34). We define a centering parameter γ as

$$\gamma := (1-\alpha_{\mathrm{a}})\min\{(1-\alpha_{\mathrm{a}})^2, 1/4\},$$

where α_{a} is the stepsize to the boundary, *i.e.*,

$$\alpha_{\mathrm{a}} = \sup\{\alpha \mid (x+\alpha\Delta x^{\mathrm{a}}) \in K,\ (s+\alpha\Delta s^{\mathrm{a}}) \in K^*,\ \alpha \in [0;1]\}$$

which we approximate using a bisection procedure. We then compute a combined centering-corrector search direction Δz as the solution to (18),

$$G(\Delta z) = -(1-\gamma)G(z), \quad W\Delta x + W^{-T}\Delta s = -v + \gamma\mu\tilde{v} - W^{-T}\eta$$

and we update $z := z + \alpha\Delta z$ with the largest step $\alpha \in [0;1)$ inside a neighborhood $\mathcal{N}(\beta)$ of the central path.

– *Checking termination*. Terminate if the updated iterate satisfies the termination criteria (given in Sect. 7.3) or else take a new step.

7 Implementation

MOSEK is software package for solving large scale linear and conic optimization problems. It can solve problems with a mixture of linear, quadratic and semidefinite cones, and the implementation is based on the homogeneous model, the NT search direction and a Mehrotra like predictor-corrector algorithm [2].

Our implementation has been extended to handle the three dimensional exponential cone using the algorithm above. We use the usual NT scaling for the symmetric cones and the Tunçel scalings for the nonsymmetric cones. Except for small differences in the linearization of the complementarity conditions, the symmetric and nonsymmmetric cones are handled completely analogously. Our extension for nonsymmetric cones also includes the three dimensional power cone, but this is not discussed further here.

7.1 Dualization, presolve and scaling

Occasionally it is worthwhile to dualize the problem before solving it, since it will make the linear algebra more efficient. Whether the primal or dual formulation is more efficient is not easily determined in advance. MOSEK makes a heuristic choice between the two forms, and the dualization is transparent to the user.

Furthermore, a presolve step is applied to the problem, which often leads to a significant reduction in computational complexity [1]. The presolve step removes obviously redundant constraints, tries to remove linear dependencies, *etc*. Finally, many optimization problems are badly scaled, so MOSEK rescales the problem before solving it. The rescaling is very simple, essentially normalizing the rows and columns of the A.

7.2 Computing the search direction

Usually the most expensive operation in each iteration of the primal-dual algorithm is to compute the search direction, *i.e.*, solving the linear system

$$\begin{bmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta x \\ \Delta\tau \end{bmatrix} - \begin{bmatrix} 0 \\ \Delta s \\ \Delta\kappa \end{bmatrix} = \begin{bmatrix} r_p \\ r_d \\ r_g \end{bmatrix}$$
$$W\Delta x + W^{-T}\Delta s = r_{xs}, \qquad \tau\Delta\kappa + \kappa\Delta\tau = r_{\tau\kappa},$$

where W is block-diagonal scaling matrix for a product of cones. Eliminating Δs and $\Delta\kappa$ from the linearized centrality conditions results in the reduced bordered system

$$\begin{bmatrix} W^TW & -A^T & c \\ A & 0 & -b \\ -c^T & b^T & \tau^{-1}\kappa \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta\tau \end{bmatrix} = \begin{bmatrix} r_d + W^T r_{xs} \\ r_p \\ r_g + \tau^{-1} r_{\tau\kappa} \end{bmatrix},$$

which can be solved in different ways. Given a (sparse) LDL^T factorization of the symmetric matrix

$$\begin{bmatrix} -W^TW & A^T \\ A & 0 \end{bmatrix}$$

it is computationally cheap to compute the search direction. In the case that the factorization breaks down due to numerical issues we add regularization to the system, *i.e.*, we modify the diagonal, which is common in interior-point methods. If the resulting search direction is inaccurate (*i.e.*, the residuals are not decreased sufficiently) we use iterative refinement, which in most cases improves the accuracy of the search direction. We omit details of computing the LDT^T factorization, since it is fairly conventional and close to the approach discussed in [2].

The algorithm is implemented in the C programming language, and the Intel MKL BLAS library is used for small dense matrix operations; the remaining portions of

the code are developed internally, and the most computationally expensive parts have been parallelized.

7.3 The termination criteria

We next discuss the termination criteria employed in MOSEK. Let $(\varepsilon_p, \varepsilon_d, \varepsilon_g, \varepsilon_i) > 0$ be given tolerance levels for the algorithm, and denote by $(\hat{x}^k, \hat{s}^k, y^k, \tau^k, \kappa^k) \in \mathbf{int}(\mathcal{D})$ the kth interior-point iterate. Consider the metrics

$$\rho_p^k := \min\left\{\rho \mid \left\| A\frac{\hat{x}^k}{\tau^k} - b\right\|_\infty \le \rho\varepsilon_p(1 + \|b\|_\infty)\right\},$$
$$\rho_d^k := \min\left\{\rho \mid \left\| A^T\frac{y^k}{\tau^k} + \frac{\hat{s}^k}{\tau^k} - c\right\|_\infty \le \rho\varepsilon_d(1 + \|c\|_\infty)\right\},$$

and

$$\rho_g^k := \min\left\{\rho \mid \min\left(\frac{\langle\hat{x}^k, \hat{s}^k\rangle}{(\tau^k)^2}, \left|\frac{\langle c, \hat{x}^k\rangle}{\tau^k} - \frac{\langle b, y^k\rangle}{\tau^k}\right|\right) \le \rho\varepsilon_g \max\left(1, \frac{\min\left(\left|\langle c, \hat{x}^k\rangle\right|, \left|\langle b, y^k\rangle\right|\right)}{\tau^k}\right)\right\}.$$

If

$$\max(\rho_p^k, \rho_d^k, \rho_g^k) \le 1$$

then

$$\begin{aligned}
\left\| A\tfrac{\hat{x}^k}{\tau^k} - b\right\|_\infty &\le \varepsilon_p(1 + \|b\|_\infty),\\
\left\| A^T\tfrac{y^k}{\tau^k} + \tfrac{\hat{s}^k}{\tau^k} - c\right\|_\infty &\le \varepsilon_d(1 + \|c\|_\infty),\\
\min\left(\tfrac{\langle\hat{x}^k,\hat{s}^k\rangle}{(\tau^k)^2}, \left|\tfrac{\langle c,\hat{x}^k\rangle}{\tau^k} - \tfrac{\langle b,y^k\rangle}{\tau^k}\right|\right) &\le \varepsilon_g \max\left(1, \tfrac{\min\left(\left|\langle c,\hat{x}^k\rangle\right|, \left|\langle b,y^k\rangle\right|\right)}{\tau^k}\right),
\end{aligned}$$

and hence $(\hat{x}^k, y^k, \hat{s}^k)/\tau^k$ is an almost primal and dual feasible solution with small duality gap. Clearly, the quality of the approximation is depends on the problem and the specified tolerances $(\varepsilon_p, \varepsilon_d, \varepsilon_g, \varepsilon_i)$. Therefore, ρ_p^k and ρ_d^k measure how far the kth iterate is from being approximately primal and dual feasible, respectively. Furthermore, ρ_g^k measure how far the kth iterate is from having a zero duality gap.

Similarly, define infeasibility metrics

$$\rho_{pi}^k := \min\left\{\rho \mid \left\| A^T y^k + \hat{s}^k\right\|_\infty \le \rho\varepsilon_i\langle b, y^k\rangle,\ \langle b, y^k\rangle > 0\right\},$$
$$\rho_{di}^k := \min\left\{\rho \mid \left\| A\hat{x}^k\right\|_\infty \le -\rho\varepsilon_i\langle c, \hat{x}^k\rangle,\ \langle c, \hat{x}^k\rangle < 0\right\},$$

and

$$\rho_{ip}^k := \min \left\{ \rho \mid \left\| \begin{matrix} A^T y^k + \hat{s}^k \\ A\hat{x}^k \end{matrix} \right\|_\infty \leq \rho \varepsilon_i \left\| \begin{matrix} y^k \\ \hat{s}^k \\ \hat{x}^k \end{matrix} \right\|_\infty, \left\| \begin{matrix} y^k \\ \hat{s}^k \\ \hat{x}^k \end{matrix} \right\|_\infty > 0 \right\}.$$

If $\rho_{pi} \leq 1$ then

$$\left\| A^T y^k + \hat{s}^k \right\|_\infty \leq \varepsilon_i \langle b, y^k \rangle, \quad \langle b, y^k \rangle > 0.$$

Thus, for

$$\bar{y} := \frac{y^k}{\langle b, y^k \rangle}, \quad \bar{s} := \frac{\hat{s}^k}{\langle b, y^k \rangle}$$

we have

$$b^T \bar{y} \geq 1, \quad \left\| A^T \bar{y} + \bar{s} \right\| \leq \varepsilon_i, \quad \bar{s} \in K^*$$

i.e., $(\bar{y}, \bar{s})$ is an approximate certificate of primal infeasibility. Similarly, if $\rho_{di} \leq 1$ we have an approximate certificate of dual infeasibility. Finally, assume that $\rho_{ip} \leq 1$. Then

$$\left\| \begin{matrix} A^T y^k + \hat{s}^k \\ A\hat{x}^k \end{matrix} \right\|_\infty \leq \varepsilon_i \left\| \begin{matrix} y^k \\ \hat{s}^k \\ \hat{x}^k \end{matrix} \right\|_\infty, \left\| \begin{matrix} y^k \\ \hat{s}^k \\ \hat{x}^k \end{matrix} \right\|_\infty > 0$$

is an approximate certificate of ill-posedness. For example, if $\left\| y^k \right\|_\infty \gg 0$ then a tiny perturbation in b will make the problem infeasible. Hence, the problem is by definition unstable.

8 Numerical results

We investigate the numerical performance of our implementation on a selection of exponential cone problems from the Conic Benchmark Library (CBLIB) [29], as well as a selection of customer provided problems. Some of those problems have integer variables, in which case we solve their continuous relaxations, *i.e.*, we ignore the integrality constraints. In the study we compare the performance of MOSEK 9.2, both with and without the proposed corrector. In the case without a corrector we also disable the standard Mehrotra corrector effecting linear and quadratic cones; otherwise the residuals will not decrease at the same rate. We also compare our implementation with the open-source solver ECOS [6], which implements the algorithm by Serrano [26]. Since ECOS only supports linear, quadratic and exponential cones we limit the test-set to examples with combinations of those cones; there are also instances in CBLIB with both exponential and semidefinite cones.

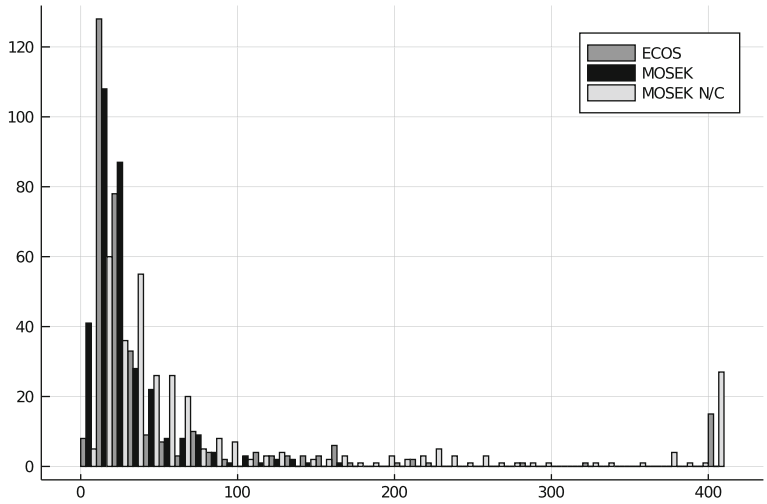

Fig. 2 Histograms of solver iterations for 326 test problems with exponential cones. MOSEK (including corrector) succesfully solved all instances and generally required the fewest iterations. MOSEK N/C (no corrector) solved 300 instances, and ECOS solved 201 instances. The number of iterations was limited to 400 in all solvers

Fig. 2 shows histograms of the number of iterations required to the solve the problems for ECOS and MOSEK (with and without the proposed corrector). The figure shows a substantial advantage of the proposed corrector over a wide selection of test problems, both in terms of stability and required number of iterations.

9 Conclusions

Based on previous work by Tunçel we have presented a generalization of the Nesterov-Todd algorithm for symmetric conic optimization to handle the nonsymmetric exponential cone. Our main contribution is a new Mehrotra-like corrector search direction for the to nonsymmetric case, which improves practical performance significantly. Moreover, we presented a practical implementation with extensive computational results documenting the efficiency of proposed algorithm. Indeed the suggested algorithm is significantly more robust and faster than the current state of the art software for nonsymmetric conic optimization ECOS.

Possible future work includes establishing the complexity of the algorithm and applying it to other nonsymmetric cone types, possibly of larger dimensions. One such is example is the nonsymmetric cone of semidefinite matrices with sparse chordal structure [32], which could extend primal-dual solvers like MOSEK with the ability to solve large sparse semidefinite programs.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included

in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

A Barrier function and derivatives

We consider derivatives up to third order of the exponential-cone barrier (2).

A.1 First-order derivatives

Let $\psi(x) = x_2 \log(x_1/x_2) - x_3$, $g(x) = -\log(\psi(x))$ and $h(x) = -\log x_1 - \log x_2$, *i.e.*, $F(x) = g(x) + h(x)$. Then $F'(x) = g'(x) + h'(x)$ with

$$g'(x) = -\frac{\psi'(x)}{\psi(x)}$$

and $\psi'(x) = (x_2/x_1, \log(x_1/x_2) - 1, -1)$, $h'(x) = -(1/x_1, 1/x_2, 0)$. Sometimes we will omit the arguments for these functions and their derivatives when it is implicitly given.

A.2 Second-order derivatives

$$F''(x) = -\frac{1}{\psi(x)}\left(\psi''(x) - \frac{\psi'(x)\psi'(x)^T}{\psi(x)}\right) + h''(x)$$

with $h''(x) = \mathbf{diag}(1/x_1^2, 1/x_2^2, 0)$ and

$$\psi''(x) = \begin{bmatrix} -\frac{x_2}{x_1^2} & \frac{1}{x_1} & 0 \\ \frac{1}{x_1} & -\frac{1}{x_2} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Let $\hat{\psi}'(x)$ and $\hat{\psi}''(x)$ denote the leading parts of $\psi'(x)$ and $\psi''(x)$, respectively, *i.e.*,

$$\hat{\psi}'(x) := \begin{bmatrix} x_2/x_1 \\ \log(x_1/x_2) - 1 \end{bmatrix},$$

$$\hat{\psi}''(x) := \begin{bmatrix} -\frac{x_2}{x_1^2} & \frac{1}{x_1} \\ \frac{1}{x_1} & -\frac{1}{x_2} \end{bmatrix} = -\begin{bmatrix} \frac{\sqrt{x_2}}{x_1} \\ -\frac{1}{\sqrt{x_2}} \end{bmatrix}\begin{bmatrix} \frac{\sqrt{x_2}}{x_1} \\ -\frac{1}{\sqrt{x_2}} \end{bmatrix}^T = -v(x)v(x)^T$$

and similarly for $\hat{h}'$ and $\hat{h}''$. We can then write $F''(x)$ as

$$F''(x) = \frac{1}{\psi(x)^2}\begin{bmatrix} A(x) + \hat{\psi}'(x)\hat{\psi}'(x)^T & -\hat{\psi}'(x) \\ -\hat{\psi}'(x)^T & 1 \end{bmatrix}$$

where

$$A(x) := \psi(x)^2 \hat{h}''(x) + \psi(x) v(x) v(x)^T.$$

We can factor $A(x) = V(x)V(x)^T$ with

$$V(x) = \psi(x) \begin{bmatrix} \frac{1-\sqrt{1+2x_2/\psi}}{2x_1} & \frac{1+\sqrt{1+2x_2/\psi}}{2x_1} \\ \frac{1+\sqrt{1+2x_2/\psi}}{2x_2} & \frac{1-\sqrt{1+2x_2/\psi}}{2x_2} \end{bmatrix},$$

which gives a factored expression of $F''(x) = R(x)R(x)^T$ where

$$R(x) = \frac{1}{\psi(x)} \begin{bmatrix} V(x) & \hat{\psi}'(x) \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} \frac{1-\sqrt{1+2x_2/\psi}}{2x_1} & \frac{1+\sqrt{1+2x_2/\psi}}{2x_1} & \frac{1}{\psi}\frac{x_2}{x_1} \\ \frac{1+\sqrt{1+2x_2/\psi}}{2x_2} & \frac{1-\sqrt{1+2x_2/\psi}}{2x_2} & \frac{\log(x_1/x_2)-1}{\psi} \\ 0 & 0 & -\frac{1}{\psi} \end{bmatrix}.$$

A.3 Third-order directional derivatives

We have $F'''(x)[u] = \frac{d}{dt} F''(x+tu)\Big|_{t=0}$ with $F''(x) = g''(x) + h''(x)$. Then

$$\begin{aligned} F'''(x)[u] = &-2\frac{\langle \psi'(x), u\rangle}{\psi(x)} g''(x) - \frac{\langle \psi'(x), u\rangle}{\psi(x)^2} \psi''(x) - \frac{1}{\psi(x)} \psi'''(x)[u] \\ &+ \frac{1}{\psi(x)^2} \left(\psi'(x) u^T \psi''(x) + \psi''(x) u \psi'(x)^T \right) + h'''(x)[u] \end{aligned}$$

with

$$\begin{aligned} \hat{h}'''(x)[u] &= \begin{bmatrix} \frac{d\hat{h}''(x)}{dx_1}\hat{u} & \frac{d\hat{h}''(x)}{dx_2}\hat{u} \end{bmatrix} = -2 \begin{bmatrix} \frac{u_1}{x_1^3} & 0 \\ 0 & \frac{u_2}{x_2^3} \end{bmatrix} \\ \hat{\psi}'''(x)[u] &= \begin{bmatrix} \frac{d\hat{\psi}''(x)}{dx_1}\hat{u} & \frac{d\hat{\psi}''(x)}{dx_2}\hat{u} \end{bmatrix} = \begin{bmatrix} \frac{2x_2u_1}{x_1^3} - \frac{u_2}{x_1^2} & -\frac{u_1}{x_1^2} \\ -\frac{u_1}{x_1^2} & \frac{u_2}{x_2^2} \end{bmatrix}. \end{aligned} \tag{33}$$

To evalutate the corrector (16) we compute

$$\eta = -(1/2) F'''(x)[u, v] \tag{34}$$

using (33) where $u := \Delta x^{\mathrm{a}}$ and $F''(x)v = \Delta s^{\mathrm{a}}$. For stability we solve for v using the factored expression of $F''(x)$.

References

1. Andersen, E.D., Andersen, K.D.: Presolving in linear programming. Math. Program. **71**(2), 221–245 (1995)
2. Andersen, E.D., Roos, C., Terlaky, T.: On implementing a primal-dual interior-point method for conic quadratic optimization. Math. Program. **95**(2), 249–277 (2003)
3. Andersen, E.D., Ye, Y.: On a homogeneous algorithm for the monotone complementarity problem. Math. Program. **84**(2), 375–399 (1999)
4. Anstreicher, K., Vial, J.P.: On the convergence of an infeasible primal-dual interior-point method for convex programming. Optim. Methods Softw. **3**, 273–283 (1994)
5. Chares, P.R.: Cones and interior-point algorithms for structured convex optimization involving powers and exponentials. Ph.D. thesis, Université Catholique de Louvain, Louvain-la-Neuve (2009)
6. Domahidi, A., Chu, E., Boyd, S.: ECOS: An SOCP solver for embedded systems. In: European Control Conference (ECC), pp. 3071–3076 (2013)
7. El-Bakry, A.S., Tapia, R.A., Tsuchiya, T., Zhang, Y.: On the formulation and theory of the primal-dual Newton interior-point method for nonlinear programming. J. Optim. Theory Appl. **89**(3), 507–541 (1996)
8. Güler, O.: Barrier functions in interior point methods. Math. Oper. Res. **21**(4), 860–885 (1996)
9. Huang, K.L., Mehrotra, S.: A Modified Potential Reduction Algorithm for Monotone Complementarity and Convex Programming Problems and Its Performance. Northwestern University, Technical repot (2012)
10. Karimi, M., Tunçel, L.: Primal-dual interior-point methods for domain-driven formulations: Algorithms. arXiv preprint arXiv:1804.06925 (2018)
11. Karmarkar, N.: A polynomial-time algorithm for linear programming. Combinatorica **4**, 373–395 (1984)
12. Lubin, M., Yamangil, E., Bent, R., Vielma, J.P.: Extended Formulations in Mixed-integer Convex Programming. In: Q. Louveaux, M. Skutella (eds.) Integer Programming and Combinatorial Optimization. IPCO 2016. Lecture Notes in Computer Science, Volume 9682, pp. 102–113. Springer, Cham (2016)
13. Mehrotra, S.: On the implementation of a primal-dual interior point method. SIAM J. Optim. **2**(4), 575–601 (1992)
14. Myklebust, T.G.J.: On primal-dual interior-point algorithms for convex optimisation. Ph.D. thesis, University of Waterloo (2015)
15. Myklebust, T.G.J., Tunçel, L.: Interior-point algorithms for convex optimization based on primal-dual metrics. arXiv preprint arXiv:1411.2129 (2014)
16. Nemirovski, A.: Advances in convex optimization: conic programming. Int. Congress Math. **1**, 413–444 (2007)
17. Nemirovski, A., Tunçel, L.: "cone-free" primal-dual path-following and potential-reduction polynomial time interior-point methods. Mathematical Programming **102**(2), 261–294 (2005)
18. Nesterov, Y.: Towards non-symmetric conic optimization. Optim. Methods Softw. **27**(4–5), 893–917 (2012)
19. Nesterov, Y., Nemirovskii, A.: Interior-Point Polynomial Algorithms in Convex Programming, 1st edn. SIAM, Philadelphia (1994)
20. Nesterov, Y., Todd, M.J.: Self-scaled barriers and interior-point methods for convex programming. Math. Oper. Res. **22**(1), 1–42 (1997)
21. Nesterov, Y., Todd, M.J.: Primal-dual interior-point methods for self-scaled cones. SIAM J. Optim. **8**(2), 324–364 (1998)
22. Nesterov, Y., Todd, M.J., Ye, Y.: Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. Math. Program. **84**(2), 227–267 (1999)
23. Nesterov, Y., Tuncel, L.: Local superlnear convergence of polynomial-time interior-point methods for hyperbolicity cone optimization problems. Tech. rep., CORE, Lovain-la-Neuve (2009). Revised August 2015
24. Øbro, M.: Conic optimization with exponential cones. Master's thesis, Technical University of Denmark (2019)
25. Schnabel, R.B.: Quasi-newton methods using multiple secant equations. Colorado University at Boulder, Technical Report (1983)
26. Serrano, S.A.: Algorithms for unsymmetric cone optimization and an implementation for problems with the exponential cone. Ph.D. thesis, Stanford University (2015)

27. Skajaa, A., Ye, Y.: A homogeneous interior-point algorithm for nonsymmetric convex conic optimization. Mathe. Program. **150**(2), 391–422 (2015)
28. Sturm, J.F.: SeDuMi 1.02, a MATLAB toolbox for optimizing over symmetric cones. Optim. Methods Softw. **11–12**, 625–653 (1999)
29. The Conic Benchmark Library. http://cblib.zib.de/ (2018). [Online; accessed 01-December-2018]
30. Tunçel, L.: Generalization of primal-dual interior-point methods to convex optimization problems in conic form. Found. Comput. Math. **1**(3), 229–254 (2001)
31. Vandenberghe, L.: The cvxopt linear and quadratic cone program solvers. Online: http://cvxopt.org/documentation/coneprog.pdf (2010)
32. Vandenberghe, L., Andersen, M.S.: Chordal graphs and semidefinite optimization. Found. Trends Optim. **1**(4), 241–433 (2015)
33. Wright, S.J.: Primal-dual interior-point methods, vol. 54. SIAM, Philadelphia (1997)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.