CORE DISCUSSION PAPER

2006/28

Towards nonsymmetric conic optimization

Yu. Nesterov *

March 14, 2006

Abstract

In this paper we propose a new interior-point method, which is based on an extension of the ideas of self-scaled optimization to the general cases. We suggest using the primal correction process to find a *scaling point*. This point is used to compute a strictly feasible primal-dual pair by simple projection. Then, we define an affine-scaling direction and perform a prediction step. This is the only moment when the dual barrier is used. Thus, we need only to compute its value, which can even be done approximately. In the second part of the paper we develop a $4n$-self-concordant barrier for n-dimensional p-cone, which can be used for numerical testing of the proposed technique.

Keywords: convex optimization, conic problems, interior-point methods, long-step path-following methods, self-concordant barriers, self-scaled barriers, affine-scaling direction, p-norm minimization

*Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL), 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium; e-mail: nesterov@core.ucl.ac.be.

The research results presented in this paper have been supported by a grant “Action de recherche concertè ARC 04/09-315” from the “Direction de la recherche scientifique - Communautè française de Belgique”. The scientific responsibility rests with its author.

1 Introduction

Motivation. Despite its very powerful theory [7, 10], the practical applications of primal-dual conic programming are mainly restricted to symmetric cones [1, 8, 9]. This situation has remained unchanged for many years for several serious reasons. First of all, conic constraints do not often arise naturally in real-world problems. Even a single non-homogeneous constraint can destroy the pure conic structure. Theoretically, any convex set can be converted into a cone by an appropriate projective transformation. However, such a transformation leads to a multiplication of the complexity parameters of the corresponding barriers by a factor of hundreds (see [2]). This looks like a paradox, but up to now almost no practically interesting example of nonlinear nonsymmetric cone with acceptable value of parameter are known[1]. The only possible exceptions are the hyperbolic cones [3], but their practical importance is questionable.

Further, even if we manage to put the primal problem into a conic form and construct the corresponding barrier function, this is not the end of the story. For the full power of interior-point methods, we need also the dual cone and the dual barrier. Since the latter barrier function must be easily computable, we get one more limitation on the applicability of the primal-dual schemes.

Finally, for general conic problems the primal-dual machinery looks quite heavy [6, 12]. At the same time, from the viewpoint of worst-case complexity analysis, the primal-dual problems have no advantages with respect to the primal ones. In the case of symmetric cones, we can believe in the power of long step methods and try to confirm our hopes by computational practice. For general cones this does not work, since up to now the available computational experiments are extremely limited (see [13]).

Since this situation has already remained unchanged for a long time, it looks natural to sacrifice some elements of the beautiful primal-dual picture, since they may be responsible for keeping the theory so far from implementations. In [4] the authors decided to skip the conic form of the feasible set, but to keep a computable primal barrier function and its Fenchel transform. In this way they managed to get long-step interior point methods, but the cone structure and any hope of a meaningful dual problem was lost.

In this paper we have decided to keep a conic description of the problem. The main reason for our conservatism is that we are now able to confirm the efficiency of new primal-dual methods by numerical experiments with a nontrivial nonlinear cone. This is the epigraph of the n-dimensional p-norm, for which we construct a self-concordant barrier with parameter $4n$. Recall that the previously known barriers for this cone have parameters ranging from $200n$ to $400n + 200$ [13].

From an algorithmic point of view, we work mainly in the primal space and use the dual problem only to justify the long path-following steps. We develop a framework, which can be seen as a shift towards the algorithmic ideas used for symmetric cones. It appears that primal centering can be interpreted as a process for finding a strictly feasible *scaling* point w. Using this point, we can *compute* a strictly feasible primal-dual pair (x, s), which is still well centered, and which satisfies the *exact* scaling condition

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

[1]We mean the cones which cannot be obtained as intersections of symmetric cones by linear subspaces.

where F is the primal barrier function. This computation, which we call *primal-dual lifting*, is carried out by solving the standard primal Newton system. Thus, no special machinery is needed to compute the generalized affine-scaling direction even for self-scaled cones. Note that in the latter case, this direction automatically becomes the standard one [9].

In our approach, the dual barrier is used only to define an appropriate step size. Hence, we only need a procedure for computing the *value* of the dual function. This computation does not need high accuracy. In some situations, the value can even be computed by an auxiliary numerical procedure. For the proposed primal-dual scheme, we prove the standard $O(\sqrt{\nu}\,\ln \frac{1}{\epsilon})$ complexity result keeping the possibility of long-step acceleration with eventual local quadratic convergence.

Contents. In Section 2 we introduce the conic primal-dual problem and describe the generic long-step predictor-corrector path-following scheme [6]. In Section 3 we study the primal correction process. We show that its output, a primal point w, can be interpreted as a scaling point for some special strictly feasible primal-dual pair (x, s). This pair is well centered. Moreover, it can be easily computed by simple projection in the local metric defined by $F''(w)$, the Hessian of the primal barrier at w. In Section 4 we introduce and study a generalized scaling direction $(\Delta x, \Delta s)$ defined by

$$\Delta s + F''(w)\Delta x = s.$$

Formally, this definition is *the same* as that for self-scaled barriers [9]. We show that along this direction the functional proximity measure grows very slowly. Sections 5 and 6 are devoted to complexity analysis of the main and preliminary phases of a new non-symmetric predictor-corrector primal-dual interior-point method. For both phases, we prove the standard $O(\sqrt{\nu}\ln \frac{1}{\epsilon})$ complexity result, where ν is the parameter of the self-concordant barrier and ϵ is the required accuracy of the solution.

The second part of the paper is devoted to the development of background material for the future numerical experiments. In Section 7 we suggest a $(4n)$-self-concordant barrier for the epigraph of the n-dimensional p-norm, $1 \leq p \leq \infty$. It is formed as a combination of 4-self-concordant barriers for three-dimensional p-cones. The whole construction requires $2n+1$ variables linked by one linear equation. In Section 8 we show that the dual barrier for the above three-dimensional cone can be easily computed. For this computation we need to solve a nonlinear equation in one variable. It can be shown that after seven bisection steps we can generate an approximate solution in the region of quadratic convergence of Newton's method.

Notation and generalities. Let E be a finite dimensional real vector space with dual space E^*. We denote the corresponding scalar product by $\langle s, x\rangle$, where $x \in E$ and $s \in E^*$. If $E = R^n$, then $E^* = R^n$ and we use the standard scalar product

$$\langle s, x\rangle \;=\; \sum_{i=1}^{n} s^{(i)} x^{(i)}, \quad s, x \in R^n.$$

For the space of symmetric $n \times n$-matrices $S^n = (S^n)^*$, the scalar product is defined as

$$\langle S, X\rangle \;=\; \sum_{i=1}^{n}\sum_{j=1}^{n} S^{(i,j)} X^{(i,j)}, \quad S, X \in S^n.$$

The actual meaning of the notation $\langle\cdot,\cdot\rangle$ can be always clarified by the space containing the arguments.

For a linear operator $A : E \to E_1^*$ we define its adjoint operator $A^* : E_1 \to E^*$ in a standard way:

$$\langle Ax, y\rangle \;=\; \langle A^* y, x\rangle, \quad x \in E,\; y \in E_1.$$

If $E_1 = E$, we can talk of self-adjoint operators: $A = A^*$.

Let $K \subseteq E$ be a convex cone. We call it *proper* if it is a closed pointed cone with nonempty interior. For a proper cone, its dual cone

$$K^* = \{s \in E^* : \langle s, x\rangle \geq 0 \;\forall x \in K\}$$

is also proper. For interior-point methods (IPM), the cone K is represented by a self-concordant barrier $F(x)$, $x \in \operatorname{int} K$, with parameter $\nu \geq 1$ (see Chapter 4 in [5] for definitions and main results). The important examples of convex cones are the *positive orthant*:

$$R^n_+ \;=\; \{x \in R^n : \; x \geq 0\}, \quad F(x) \;=\; -\sum_{i=1}^n \ln x^{(i)}, \quad \nu \;=\; n,$$

the *Lorentz cone*

$$\mathcal{L}_n \;=\; \{(\tau, x) \in R \times R^n : \; \tau \geq \langle x, x\rangle^{1/2}\}, \quad F(\tau, x) \;=\; -\ln(\tau^2 - \langle x, x\rangle), \quad \nu \;=\; 2.$$

and the *cone of positive semidefinite matrices*

$$S^n_+ \;=\; \{X \in S^n : \; X \succeq 0\}, \quad F(X) \;=\; -\ln\det X, \quad \nu \;=\; n,$$

In all these examples, the cones are *symmetric* and the barriers are *self-scaled* [8]. However, in general this cannot be true. A nontrivial example of a nonsymmetric convex cone in R^n with computable $(4n)$-self-concordant barrier is considered in Sections 7 and 8.

The natural barriers for cones are *logarithmically homogeneous* barriers:

$$F(\tau x) \;\equiv\; F(x) - \nu \ln \tau, \quad x \in \operatorname{int} K,\; \tau > 0. \tag{1.1}$$

Let us point out some straightforward consequences of this property:

$$F'(\tau x) \;=\; \tfrac{1}{\tau} F'(x), \quad F''(\tau x) \;=\; \tfrac{1}{\tau^2} F''(x), \tag{1.2}$$

$$F''(x)x \;=\; -F'(x), \tag{1.3}$$

$$\langle F'(x), x\rangle \;=\; -\nu, \tag{1.4}$$

$$\langle F''(x)x, x\rangle \;=\; \nu, \quad \langle F'(x), [F''(x)]^{-1} F'(x)\rangle \;=\; \nu, \tag{1.5}$$

(for proofs, see Section 2.3 in [7]). In what follows, we always assume that $F(x)$ is logarithmically homogeneous.

It is important that the dual barrier

$$F_*(s) \;=\; \max_x \{-\langle s, x\rangle - F(x) : \; x \in \operatorname{int} K\}, \quad s \in \operatorname{int} K^*,$$

is a ν-self-concordant logarithmically homogeneous barrier for K^*. The pair of primal-dual barriers satisfy the following duality relations:

$$-F'(x) \in \operatorname{int} K, \quad -F'_* \in \operatorname{int} K^*, \tag{1.6}$$

$$\begin{aligned} F_*(-F'(x)) &= \langle F'(x), x\rangle - F(x) \;=\; -\nu - F(x), \\ F(-F_*'(s)) &= -\nu - F(x). \end{aligned} \tag{1.7}$$

$$F_*'(-F'(x)) \;=\; -x, \quad F'(-F_*'(s)) \;=\; -s, \tag{1.8}$$

$$F_*''(-F'(x)) \;=\; [F''(x)]^{-1}, \quad F''(-F_*'(s)) \;=\; [F_*''(s)]^{-1}, \tag{1.9}$$

$$F(x) + F_*(s) \;\;\geq\;\; -\nu - \nu \ln \tfrac{\langle s,x\rangle}{\nu}, \tag{1.10}$$

and the last inequality is satisfied as an equality if and only if $s = -\tau F'(x)$ for some $\tau > 0$ (see Section 2.4 in [7]). In what follows we assume that both primal and dual barriers are computable. In some cases, this assumption must be supported by an auxiliary computational procedure. We give an example of the efficient treatment of such a situation in Section 8.

At any $x \in \operatorname{int} K$ we use the Hessian $F''(x) : E \to E^*$ to define the following local Euclidean norms:

$$\begin{aligned} \|h\|_x &= \langle F''(x)h, h\rangle^{1/2}, \quad h \in E, \\ \|s\|_x^* &= \langle s, [F''(x)]^{-1}s\rangle, \quad s \in E^*. \end{aligned}$$

It is well known that for any $x \in \operatorname{int} K$ the corresponding *Dikin ellipsoid* is feasible:

$$W(x) \;=\; \{u \in E : \; \|u - x\|_x \leq 1\} \;\subseteq\; K. \tag{1.11}$$

We often use two important inequalities:

$$F(u) \leq F(x) + \langle F'(x), u - x\rangle + \omega(r), \tag{1.12}$$

$$F''(u) \;\;\preceq\;\; \tfrac{1}{(1-r)^2} F''(x), \tag{1.13}$$

where $x \in \operatorname{int} K$, $r = \|u - x\|_x < 1$, and $\omega(t) = -t - \ln(1-t)$.

For a self-adjoint operator $B : E \to E^*$ we define also a point-dependent operator norm:

$$\|B\|_x \;=\; \max_h \{\|Bh\|_x^* : \; \|h\|_x \leq 1\}.$$

In the sequel, we need the following simple result

Lemma 1 *Let F be a self-concordant barrier for Q and $x, u \in \operatorname{int} K$, $r \stackrel{\rm def}{=} \|x - u\|_u < 1$. Then*

$$\|F'(x) - F'(u) - F''(u)(x-u)\|_u^* \;\;\leq\;\; \tfrac{r^2}{1-r}. \tag{1.14}$$

Proof:
Indeed,

$$F'(x) - F'(u) \;\;=\;\; \left(\int\limits_0^1 F''(u + \tau(x-u)) d\tau \right) \cdot (x - u) \;\stackrel{\rm def}{=}\; G \cdot (x-u).$$

In view of Corollary 4.1.4 in [5], we have

$$-(r - \tfrac{1}{3}r^2) F''(u) \preceq G - F''(u) \preceq \tfrac{r}{1-r} F''(u).$$

Therefore

$$\begin{aligned}\|F'(x)-F'(u)-F''(u)(x-u)\|_u^* &= \|(G-F''(u))(x-u)\|_u^* \\ &\le \|G-F''(u)\|_u\cdot r \;\le \tfrac{r^2}{1-r}.\end{aligned}$$

□

2 Primal-dual predictor-corrector IPM

Consider the standard conic optimization problem

$$\begin{aligned}&\min_x \langle c,x\rangle,\\ &\text{s.t.}\quad x\in\mathcal{F}_P \stackrel{\mathrm{def}}{=} \{x\in K:\ Ax=b\},\end{aligned}\tag{2.1}$$

where $K\subset E$ is a proper cone, $c\in E^*$, $b\in R^m$, and the linear operator A maps E to R^m. Then, we can write down the dual problem

$$\begin{aligned}&\max_{s,y} \langle b,y\rangle,\\ &\text{s.t.}\quad s+A^*y=c,\\ &\qquad\ s\in K^*,\ y\in R^m.\end{aligned}\tag{2.2}$$

Note that the dual cone K^* is also proper. For a feasible primal-dual point $z=(x,s,y)$ the following relations hold

$$0 \;\le\; \langle s,x\rangle \;=\; \langle c-A^*y,x\rangle \;=\; \langle c,x\rangle-\langle Ax,y\rangle \;=\; \langle c,x\rangle-\langle b,y\rangle.\tag{2.3}$$

In what follows, we always assume existence of a strictly feasible primal-dual point

$$(x_0,s_0,y_0):\quad Ax_0=b,\ x_0\in\operatorname{int}K,\quad s_0+A^*y_0=c,\ s_0\in\operatorname{int}K^*.\tag{2.4}$$

In this case, strong duality holds for problems (2.1), (2.2).

We assume that the primal cone is endowed with a ν-logarithmically homogeneous self-concordant barrier $F(x)$. Then, the conjugate barrier

$$F_*(s) \;=\; \max_x\{-\langle s,x\rangle - F(x)\},\ s\in\operatorname{int}K^*,$$

is also ν-logarithmically homogeneous and self-concordant. These barriers define the *primal-dual central path* (see, for example, [6]).

Theorem 1 *Under assumption (2.4), the primal-dual central path,*

$$\left.\begin{aligned}x(t) &= \arg\min_x\{t\langle c,x\rangle+F(x):\ Ax=b\}\\ y(t) &= \arg\max_y\{t\langle b,y\rangle-F_*(c-A^*y)\}\\ s(t) &= c-A^*y(t)\end{aligned}\right\},\quad t>0,\tag{2.5}$$

is well defined. Moreover, for any $t > 0$, the following identities hold:

$$\langle s(t), x(t)\rangle \;=\; \langle c, x(t)\rangle - \langle b, y(t)\rangle \;=\; \tfrac{\nu}{t}, \tag{2.6}$$

$$F(x(t)) + F_*(s(t)) \;=\; -\nu + \nu \ln t, \tag{2.7}$$

$$s(t) \;=\; -\tfrac{1}{t}F'(x(t)), \quad x(t) \;=\; -\tfrac{1}{t}F'_*(s(t)). \tag{2.8}$$

Hence, the optimal values of problems (2.1), (2.2) coincide and their optimal sets are bounded.

Note that the central path $z(t) = (x(t), s(t), y(t))$ is differentiable. Its derivatives can be found from the following linear system:

$$\begin{aligned} s'(t) + B(t)\cdot x'(t) \;&=\; -\tfrac{1}{t}s(t), \\ Ax'(t) \;=\; 0, \quad s'(t) + A^*y'(t) \;&=\; 0, \end{aligned} \tag{2.9}$$

where $B(t) \stackrel{\mathrm{def}}{=} \frac{1}{t}F''(x(t)) \stackrel{(2.8)}{=} \frac{1}{t}F''(-\frac{1}{t}F'_*(s(t))) \stackrel{(1.2)}{=} tF''(-F'_*(s(t))) \stackrel{(1.9)}{=} [\frac{1}{t}F''_*(s(t))]^{-1}$. Since

$$s(t) \stackrel{(2.8),(1.3)}{=} B(t)\cdot x(t), \quad t > 0, \tag{2.10}$$

the first equation in (2.9) can be written in an equivalent symmetric form:

$$x'(t) + B^{-1}(t)\cdot s'(t) \;=\; -\tfrac{1}{t}x(t). \tag{2.11}$$

Modern IPM's are often based on different path-following strategies for the primal-dual problem

$$\begin{aligned} &\min_{x,s,y}\{\langle c, x\rangle - \langle b, y\rangle :\; (x,s,y) \in \mathcal{F}\}, \\ &\mathcal{F} \;=\; \{(x,s,y) :\; Ax = b,\; s + A^*y = c,\; x \in K,\; s \in K^*\}. \end{aligned} \tag{2.12}$$

The main advantage of this formulation lies in the very useful relations (2.6) - (2.8), which allow one to define different *global* proximity measures for the primal-dual central path. One of the most natural is the *functional* measure (see [6], [9])

$$\begin{aligned} \Omega(x,s,y) &\stackrel{(1.10)}{=} F(x) + F_*(s) + \nu \ln \tfrac{\langle s,x\rangle}{\nu} + \nu \\ &\stackrel{(2.3)}{=} F(x) + F_*(s) + \nu \ln \tfrac{\langle c,x\rangle - \langle b,y\rangle}{\nu} + \nu. \end{aligned}$$

The corresponding generic predictor-corrector path-following scheme looks as follows.

$$
\begin{array}{|ll|}
\hline
1. & \text{Choose parameters } \beta_1 > \beta_0 > 0. \\
2. & \text{Find } z_0 = (x_0, s_0, y_0) \in \mathcal{F} \text{ with } \Omega(z_0) \le \beta_0. \text{ Define } t_0 = \frac{\nu}{\langle s_0, x_0 \rangle}. \\
3. & \text{For } k \ge 0 \text{ iterate:} \\
 & \text{a) Compute } \Delta z_k \approx z'(t_k). \text{ Find } \alpha_k : \ \Omega(z_k + \alpha_k \Delta z_k) = \beta_1. \\
 & \text{b) Set } \tilde{z}_k = z_k + \alpha_k \Delta z_k, \text{ and } t_{k+1} = \frac{\nu}{\langle \tilde{s}_{k+1}, \tilde{x}_{k+1} \rangle}. \\
 & \text{c) Using } \tilde{z}_k \text{ as a starting point, find} \\
 & \quad z_{k+1} \in \mathcal{F} \bigcap \{z : \ \langle c, x \rangle - \langle b, y \rangle = \frac{\nu}{t_{k+1}}\} : \ \Omega(z_{k+1}) \le \beta_0. \\
\hline
\end{array}
\tag{2.13}
$$

Let us discuss the above scheme. First of all, note that the restriction of the proximity measure $\Omega(z)$ onto the hyperplane $\{z : \ \langle c, x \rangle - \langle b, y \rangle = \frac{\nu}{t_{k+1}}\}$ is a self-concordant barrier. It can be minimized by a damped Newton method, which at each iteration reduces the value of the objective function by an absolute constant (see, for example, Section 4.2.5 in [5]). Hence, the duration of the correction phase 3c) in (2.13) is bounded by $O(\beta_1)$ Newton steps. Note that the optimization problem in the correction phase is solved over the *full* primal-dual set $\mathcal{F}$.

A reasonable strategy for implementing the predictor step 3a) is not so evident. If z_k belongs to the primal-dual central path, then the expressions for the derivatives (2.9) and (2.11) are symmetric. However, outside of the path the symmetry is lost. In a straightforward implementation, an approximate tangent direction to the central path can be computed using the Hessians of the primal and dual barriers (see Section 9.1 in [6]). But this can double the computational cost of the predictor step. For self-scaled barriers [9], the symmetric *affine-scaling* direction can be found by a *scaling point* $w \in \operatorname{int} K$, which links two arbitrary points $x \in \operatorname{int} K$ and $s \in \operatorname{int} K^*$ through the equation $s = F''(w)x$. However, this technique works only for symmetric cones. Moreover, the computation of the point w is usually rather difficult (see [11]).

In the next sections we will show that an approximate symmetric tangent direction to the primal-dual central path can be found even for general cones using a non-symmetric correction phase. Moreover, the computational cost of this direction is exactly the same as that of the Newton step in the correction phase.

3 Pure primal correction phase

For a feasible primal-dual point $z \in \operatorname{rint} \mathcal{F}$ we define its *penalty value* as

$$
t(z) \overset{(2.3)}{=} \frac{\nu}{\langle s, x \rangle} = \frac{\nu}{\langle c, x \rangle - \langle b, y \rangle}. \tag{3.1}
$$

Consider $z_0 = (x_0, s_0, y_0) \in \operatorname{rint} \mathcal{F}$. Our goal is to find a close approximation to the point $z(t_0)$ with $t_0 \stackrel{\mathrm{def}}{=} t(z_0)$. This can be done by solving the problem

$$\min_u \left\{ f_{t_0}(u) \stackrel{\mathrm{def}}{=} t_0\langle c, u\rangle + F(u) : \ Au = b \right\}. \tag{3.2}$$

Note that the objective function in (3.2) is self-concordant. Moreover, in view of Theorem 1, the unique minimum of this function is attained at $x(t_0)$.

Lemma 2 *We have* $0 \le f_{t_0}(x_0) - f_{t_0}(x(t_0)) \le \Omega(z_0)$.

Proof:
Note that the point y_0 is feasible for the maximization problem in (2.5). Therefore

$$t_0\langle b, y_0\rangle - F_*(s_0) \ \le \ t_0\langle b, y(t_0)\rangle - F_*(s(t_0)).$$

Hence,

$$\begin{aligned}
f_{t_0}(x_0) - f_{t_0}(x(t_0)) \quad & \le \quad f_{t_0}(x_0) - f_{t_0}(x(t_0)) + t_0\langle b, y(t_0)\rangle - F_*(s(t_0)) \\
& \qquad -t_0\langle b, y_0\rangle + F_*(s_0) \\
& = \quad t_0[\langle c, x_0\rangle - \langle b, y_0\rangle] + F(x_0) + F_*(s_0) \\
& \qquad -t_0[\langle c, x(t_0)\rangle - \langle b, y(t_0)\rangle] - F(x(t_0)) - F_*(s(t_0)) \\
& \stackrel{(2.6),(2.7)}{=} \quad t_0\langle s_0, x_0\rangle + F(x_0) + F_*(s_0) - \nu - [-\nu + \nu \ln t_0] \\
& = \quad F(x_0) + F_*(s_0) + \nu + \nu \ln \tfrac{1}{t_0} \ = \ \Omega(z_0).
\end{aligned}$$

□

Thus, we can treat the problem (3.2) by a damped Newton method. Each iteration of this scheme requires computation of the Newton step

$$\delta_{t_0}(u) \quad = \quad \arg\min_\delta \left\{ \langle t_0 c + F'(u), \delta\rangle + \tfrac{1}{2}\langle F''(u)\delta, \delta\rangle : \ A\delta = 0 \right\} \tag{3.3}$$

at some point $u \in \operatorname{int} K$, $Au = b$. Then the value $\lambda_{t_0}(u) \stackrel{\mathrm{def}}{=} \|\delta_{t_0}(u)\|_u$ can be seen as the local norm of the gradient of function f_{t_0} restricted to the affine subspace $\{u : \ Au = b\}$. If $\lambda_{t_0}(u)$ is small, then u is close to the solution $x(t_0)$. By the general theory of self-concordant functions and Lemma 2, this can be achieved in $O(\Omega(z_0))$ steps of the damped Newton method started at $u = x_0$.

Let us write down the optimality conditions for problem (3.3):

$$\begin{aligned}
t_0 c + F'(u) + F''(u)\delta \quad & = \quad A^* y, \\
A\delta \quad & = \quad 0.
\end{aligned} \tag{3.4}$$

Let y be the optimal dual multipliers of problem (3.3). Denote

$$x_{t_0}(u) \;=\; u - \delta_{t_0}(u), \quad s_{t_0}(u) \;=\; c - \tfrac{1}{t_0} A^* y, \quad y_{t_0}(u) \;=\; \tfrac{1}{t_0} y. \tag{3.5}$$

Note that $x_{t_0}(u)$ is formed by a shift from u along a direction pointing *away* from the primal central path. We call the procedure (3.5) the *primal-dual lifting* of point $u \in \mathcal{F}_P$.

Theorem 2 *If $\lambda_{t_0}(u) \leq \beta < 1$, then the point $z_{t_0}(u) = (x_{t_0}(u), s_{t_0}(u), y_{t_0}(u))$ is strictly feasible and satisfies the following scaling relations*

$$s_{t_0}(u) \;=\; \tfrac{1}{t_0} F''(u) \cdot x_{t_0}(u), \tag{3.6}$$

$$\| F'(x_{t_0}(u)) - \tfrac{1}{t_0} F''(u) \cdot F'_*(s_{t_0}(u)) \|^*_u \;\leq\; \tfrac{2\beta^2}{1-\beta}. \tag{3.7}$$

Moreover, the point $z_{t_0}(u)$ is well centered:

$$\Omega(z_{t_0}(u)) \;\leq\; 2\omega(\beta) + \beta^2, \tag{3.8}$$

and for its penalty value the following bounds hold:

$$\begin{aligned} t(z_{t_0}(u)) \;\equiv\; \tfrac{\nu}{\langle s_{t_0}(u), x_{t_0}(u)\rangle} \;&\geq\; \tfrac{t_0}{(1+\beta/\sqrt{\nu})^2} \;\geq\; t_0 \cdot e^{-\frac{2\beta}{\sqrt{\nu}}}, \\ t(z_{t_0}(u)) \;&\leq\; \tfrac{t_0}{(1-\beta/\sqrt{\nu})^2} \;\leq\; t_0 \cdot e^{\frac{2\beta}{\sqrt{\nu}-\beta}}. \end{aligned} \tag{3.9}$$

Proof:
Indeed, $\|x_{t_0}(u) - u\|_u = \|\delta_{t_0}(u)\|_u \leq \beta < 1$. Therefore, $x_{t_0}(u) \overset{(1.11)}{\in} \operatorname{int} K$. Moreover,

$$A x_{t_0}(u) \;=\; Au - A\delta_{t_0}(u) \;=\; Au \;=\; b.$$

Further, note that

$$t_0 \cdot s_{t_0}(u) \;\overset{(3.4),(3.5)}{=}\; -F'(u) - F''(u) \cdot \delta_{t_0}(u). \tag{3.10}$$

Therefore,

$$\|t_0 s_{t_0}(u) + F'(u)\|^*_u \;=\; \|F''(u)\delta_{t_0}(u)\|^*_u \;=\; \|\delta_{t_0}(u)\|_u \;\leq\; \beta \;<\; 1.$$

Thus, $s_{t_0}(u) \overset{(1.11)}{\in} \operatorname{int} K^*$, and we conclude that $z_{t_0}(u)$ is strictly feasible. Moreover,

$$t_0 \cdot s_{t_0}(u) \;\overset{(1.3)}{=}\; F''(u) \cdot u - F''(u) \cdot \delta_{t_0}(u) \;\overset{(3.5)}{=}\; F''(u) \cdot x_{t_0}(u).$$

Let us justify inequality (3.7). Denote

$$r_d \;=\; F'(x_{t_0}(u)) - F'(u) + F''(u) \cdot \delta_{t_0}(u).$$

In view of inequality (1.14) and (3.5), we have $\|r_d\|_u^* \leq \frac{\beta^2}{1-\beta}$. Consider now

$$
\begin{aligned}
r_p &\stackrel{\mathrm{def}}{=} F_*'(t_0 s_{t_0}(u)) + u + \delta_{t_0}(u) \\
&\stackrel{(1.8)}{=} F_*'(t_0 s_{t_0}(u)) - F_*'(-F'(u)) + \delta_{t_0}(u) \\
&\stackrel{(1.9)}{=} F_*'(t_0 s_{t_0}(u)) - [F_*'(-F'(u)) + F_*''(-F'(u))(-F''(u)\delta_{t_0}(u))] \\
&\stackrel{(3.10)}{=} F_*'(-F'(u) - F''(u)\cdot\delta_{t_0}(u)) - [F_*'(-F'(u)) + F_*''(-F'(u))(-F''(u)\delta_{t_0}(u))]\,.
\end{aligned}
$$

Since $\|F''(u)\delta_{t_0}(u)\|_u^* = \|\delta_{t_0}(u)\|_u \leq \beta$, by a dual variant of Lemma 1 we conclude that $\|r_p\|_u \leq \frac{\beta^2}{1-\beta}$. Therefore,

$$
F'(x_{t_0}(u)) - \tfrac{1}{t_0}F''(u)\cdot F_*'(s_{t_0}(u)) \stackrel{(1.2)}{=} F'(x_{t_0}(u)) - F''(u)\cdot F_*'(t_0 s_{t_0}(u))
$$

$$
= r_d + F'(u) - F''(u)\cdot\delta_{t_0}(u) - F''(u)\cdot(r_p - u - \delta_{t_0}(u)) \stackrel{(1.3)}{=} r_d - F''(u)\cdot r_p.
$$

Hence,

$$
\|F'(x_{t_0}(u)) - \tfrac{1}{t_0}F''(u)\cdot F_*'(s_{t_0}(u)\|_u^* = \|r_d - F''(u)\cdot r_p\|_u^*
$$

$$
\leq \|r_d\|_u^* + \|F''(u)\cdot r_p\|_u^* = \|r_d\|_u^* + \|r_p\|_u \leq \tfrac{2\beta^2}{1-\beta}.
$$

Further, in order to prove (3.9), note that

$$
t_0\langle s_{t_0}(u), x_{t_0}(u)\rangle \stackrel{(3.6)}{=} \|x_{t_0}(u)\|_u^2 \leq (\|u\|_u + \|\delta_{t_0}(u)\|_u)^2 \stackrel{(1.5)}{\leq} (\sqrt{\nu}+\beta)^2.
$$

Therefore,

$$
t(z_{t_0}(u)) = \tfrac{\nu}{\langle s_{t_0}(u), x_{t_0}(u)\rangle} \geq \tfrac{t_0}{(1+\beta/\sqrt{\nu})^2} \geq t_0\cdot e^{-\frac{2\beta}{\sqrt{\nu}}}.
$$

For the second part of the inequality note that $t_0\langle s_{t_0}(u), x_{t_0}(u)\rangle \geq (\sqrt{\nu}-\beta)^2$.

In order to prove (3.8), we need a more accurate estimate for the new penalty value:

$$
\begin{aligned}
t_0\langle s_{t_0}(u), x_{t_0}(u)\rangle &= \|x_{t_0}(u)\|_u^2 = \|u\|_u^2 - 2\langle F''(u)u, \delta_{t_0}(u)\rangle + \|\delta_{t_0}(u)\|_u^2 \\
&\stackrel{(1.3),(1.5)}{\leq} \nu + 2\langle F'(u), \delta_{t_0}(u)\rangle + \beta^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\Omega(z_{t_0}(u)) &\stackrel{(3.10)}{=} F(x_{t_0}(u)) + F_*\left(\tfrac{1}{t_0}(-F'(u) - F''(u)\delta_{t_0}(u))\right) + \nu + \nu\ln\tfrac{\langle s_{t_0}(u), x_{t_0}(u)\rangle}{\nu} \\
&\stackrel{(1.1)}{=} F(x_{t_0}(u)) + F_*\left(-F'(u) - F''(u)\delta_{t_0}(u)\right) + \nu + \nu\ln\tfrac{t_0\langle s_{t_0}(u), x_{t_0}(u)\rangle}{\nu}
\end{aligned}
$$

Note that

$$
\begin{aligned}
F_*(-F'(u) - F''(u)\delta_{t_0}(u)) &\stackrel{(1.12)}{\leq} F_*(-F'(u)) + \langle -F''(u)\delta_{t_0}(u), F_*'(-F'(u))\rangle + \omega(\beta) \\
&\stackrel{(1.8)}{=} -\nu - F(u) + \langle F''(u)\delta_{t_0}(u), u\rangle + \omega(\beta) \\
&\stackrel{(1.3)}{=} -\nu - F(u) - \langle F'(u), \delta_{t_0}(u)\rangle + \omega(\beta).
\end{aligned}
$$

Hence, we conclude that

$$\begin{array}{rcl}\Omega(z_{t_0}(u)) & \leq & F(x_{t_0}(u)) - F(u) - \langle F'(u), \delta_{t_0}(u)\rangle + \omega(\beta) \\ & & +\nu \ln\left(1 + \frac{2}{\nu}\langle F'(u), \delta_{t_0}(u)\rangle + \frac{1}{\nu}\beta^2\right) \\ & \overset{(3.5)}{\leq} & F(u - \delta_{t_0}(u)) - F(u) + \langle F'(u), \delta_{t_0}(u)\rangle + \omega(\beta) + \beta^2 \\ & \overset{(1.12)}{\leq} & 2\omega(\beta) + \beta^2.\end{array}$$

□

Remark 1 *Using a finer assumption on the properties of the barrier function $F(x)$, it is possible to guarantee a smaller right-hand side in inequality (3.7). For example, for self-scaled cones the right-hand side is zero. However, this does not improve the worst-case complexity bounds of the corresponding IPM.*

Thus, we have found a new interpretation of the auxiliary problem (3.2). It can be used to compute the *scaling point* u in a neighborhood of the primal central path. Using this point, we can form a strictly feasible primal-dual point $z_{t_0}(u)$, which is well centered and satisfies the scaling relations (3.6), (3.7). In the next section we will show that these relations are crucial for defining a symmetric approximation of a tangent direction to the primal-dual central path.

4 Affine-scaling direction for general cones

In Section 3 we have proved that for any value of the centering parameter $\beta \in (0, 1)$ it is possible to compute a scaling point $w = \sqrt{t_0} \cdot u \in \operatorname{int} K$ and a strictly feasible primal-dual point $z = z_{t_0}(u)$, which belongs to a small neighborhood (3.8) of the central path and satisfies the following scaling relations

$$\begin{array}{rcl} s & \overset{(1.2)}{=} & F''(w) \cdot x, \\ \|F'(x) - F''(w) \cdot F'_*(s)\|^*_w & \leq & \frac{2\beta^2}{1-\beta} \cdot \sqrt{t_0}. \end{array} \tag{4.1}$$

Now we can define an approximate tangent direction $\Delta z = (\Delta x, \Delta s, \Delta y)$ to the central path in the following symmetric way (compare with (2.9):

$$\begin{array}{rcl} \Delta s + F''(w) \cdot \Delta x & = & s, \\ A\Delta x & = & 0, \\ \Delta s + A^* \Delta y & = & 0. \end{array} \tag{4.2}$$

We call Δz the *affine scaling direction*. Our definition coincides with the definition of the affine-scaling direction for self-scaled barriers (see Section 5.1 in [9]). Note that

$$\langle \Delta s, \Delta x \rangle \;=\; -\langle A^* \Delta y, \Delta x \rangle \;=\; -\langle A \Delta x, \Delta y \rangle \;=\; 0. \tag{4.3}$$

Let us present the main properties of the affine-scaling direction.

Lemma 3 *The following relations hold:*

$$\langle s, \Delta x \rangle + \langle \Delta s, x \rangle \;=\; \langle s, x \rangle, \tag{4.4}$$

$$\langle c, x - \Delta x \rangle - \langle b, y - \Delta y \rangle \;=\; 0, \tag{4.5}$$

$$\| \Delta x \|_w^2 + (\| \Delta s \|_w^*)^2 \;=\; \langle s, x \rangle, \tag{4.6}$$

$$\begin{aligned} |\nu + \langle F'(x), \Delta x \rangle + \langle \Delta s, F_*'(s) \rangle| \;&\leq\; \tfrac{1}{2} \langle s, x \rangle^{1/2} \cdot \| F'(x) - F''(w) F_*'(s) \|_w^* \\ &\overset{(4.1)}{\leq}\; \langle t_0 s, x \rangle^{1/2} \cdot \tfrac{\beta^2}{1-\beta} \;\overset{(3.9)}{\leq}\; \tfrac{\beta^2}{1-\beta} \cdot (\beta + \sqrt{\nu}). \end{aligned} \tag{4.7}$$

Proof:
Indeed, in view of definition (4.2) and relation (4.1), we have

$$\langle s, \Delta x \rangle + \langle \Delta s, x \rangle \;=\; \langle s, \Delta x \rangle + \langle s - F''(w) \Delta x, x \rangle \;=\; \langle s, x \rangle.$$

Therefore

$$\begin{aligned} \langle c, x - \Delta x \rangle - \langle b, y - \Delta y \rangle \;&=\; \langle s + A^* y, x - \Delta x \rangle - \langle b, y - \Delta y \rangle \\ &=\; \langle s, x - \Delta x \rangle + \langle A x, \Delta y \rangle \\ &=\; \langle s, x \rangle - \langle s, \Delta x \rangle - \langle \Delta s, x \rangle \;=\; 0. \end{aligned}$$

Further, multiplying (4.2) by Δx, we get $\| \Delta x \|_w^2 = \langle s, \Delta x \rangle$. Multiplying the same inequality by $[F''(w)]^{-1} \Delta s$, we get $(\| \Delta s \|_w^*)^2 = \langle \Delta s, x \rangle$. Adding these equalities and using (4.4), we obtain (4.6). Let us prove the remaining inequality.

Multiplying the first equation in (4.2) by $F_*'(s)$, we get

$$\langle \Delta s, F_*'(s) \rangle + \langle F''(w) F_*'(s), \Delta x \rangle \;=\; \langle s, F_*'(s) \rangle \;\overset{(1.4)}{=}\; -\nu.$$

Multiplying the same equation by $[F''(w)]^{-1} F'(x)$, we obtain

$$\langle \Delta s, [F''(w)]^{-1} F'(x) \rangle + \langle F'(x), \Delta x \rangle \;=\; \langle s, [F''(w)]^{-1} F'(x) \rangle \;\overset{(4.1)}{=}\; \langle F'(x), x \rangle \;\overset{(1.4)}{=}\; -\nu.$$

Adding these equalities, we have

$$\begin{aligned} -2\nu \;&=\; \langle F'(x), \Delta x \rangle + \langle \Delta s, F_*'(s) \rangle + \langle F''(w) \Delta x, F_*'(s) \rangle + \langle \Delta s, [F''(w)]^{-1} F'(x) \rangle \\ &=\; 2 \langle F'(x), \Delta x \rangle + 2 \langle \Delta s, F_*'(s) \rangle + \langle F''(w) \Delta x - \Delta s, F_*'(s) - [F''(w)]^{-1} F'(x) \rangle. \end{aligned}$$

Therefore

$$|\nu + \langle F'(x), \Delta x \rangle + \langle \Delta s, F_*'(s) \rangle| \leq \tfrac{1}{2} \| \Delta s - F''(w) \Delta x \|_w^* \cdot \| F_*'(s) - [F''(w)]^{-1} F'(x) \|_w.$$

It remains to note that $(\|\Delta s - F''(w)\Delta x\|_w^*)^2 \stackrel{(4.3)}{=} \|\Delta x\|_w^2 + (\|\Delta s\|_w^*)^2 \stackrel{(4.6)}{=} \langle s, x\rangle$. $\square$

Thus, a unit step along the affine-scaling direction results in zero duality gap. Let us show that along this direction the functional proximity measure grows very slowly. Indeed,

$$
\begin{array}{rcl}
\Omega(z \pm \alpha\Delta z) - \Omega(z) & = & F(x \pm \alpha\Delta x) + F_*(s \pm \alpha\Delta s) + \nu \ln \frac{\langle s \pm \alpha\Delta s, x \pm \alpha\Delta s\rangle}{\nu} \\
& & -F(x) - F_*(s) - \nu \ln \frac{\langle s,x\rangle}{\nu} \\
& \stackrel{(4.4)}{=} & F(x \pm \alpha\Delta x) + F_*(s \pm \alpha\Delta s) - F(x) - F_*(s) + \nu \ln(1 \pm \alpha) \\
& \stackrel{(1.12)}{\leq} & \pm\alpha[\langle F'(x), \Delta x\rangle + \langle \Delta s, F_*'(s)\rangle] + \nu \ln(1 \pm \alpha) \\
& & +\omega(\alpha\|\Delta x\|_x) + \omega(\alpha\|\Delta s\|_s) \\
& \stackrel{(4.7)}{\leq} & \alpha\beta^2 \cdot \frac{\beta + \sqrt{\nu}}{1-\beta} + \omega(\alpha\|\Delta x\|_x) + \omega(\alpha\|\Delta s\|_s).
\end{array}
$$

Let us estimate the sum of the last two terms. Note that function $\psi(t) \stackrel{\text{def}}{=} \omega(\sqrt{t})$ is convex:

$$
\begin{array}{rcl}
\psi'(t) & = & \omega'(\sqrt{t}) \cdot \frac{1}{2\sqrt{t}}, \quad t > 0, \\
\psi''(t) & = & \omega''(\sqrt{t}) \cdot \frac{1}{4t} - \omega'(\sqrt{t}) \cdot \frac{1}{4t\sqrt{t}} \; = \; \frac{1}{4t\sqrt{t}} \cdot \left(\sqrt{t} \cdot \omega''(\sqrt{t}) - \omega'(\sqrt{t}) \right) \\
& = & \frac{1}{4t\sqrt{t}} \cdot \left(\frac{\sqrt{t}}{(1-\sqrt{t})^2} - \frac{\sqrt{t}}{1-\sqrt{t}} \right) \; > \; 0.
\end{array}
$$

Denote $r \stackrel{\text{def}}{=} [\|\Delta x\|_x^2 + \|\Delta s\|_s^2]^{1/2}$. Then

$$
\begin{array}{rcl}
\omega(\alpha\|\Delta x\|_x) + \omega(\alpha\|\Delta s\|_s) & = & \psi(\alpha^2\|\Delta x\|_x^2) + \psi(\alpha^2\|\Delta s\|_s^2) \\
& \leq & \psi(\alpha^2 r^2) \; = \; \omega(\alpha r).
\end{array} \tag{4.8}
$$

In view of (1.13), we have

$$
\begin{array}{rcl}
F''(x) & \preceq & \frac{1}{(1-\beta)^2} F''(u) \stackrel{(1.2)}{=} \frac{t_0}{(1-\beta)^2} F''(w), \\
\frac{1}{t_0^2} F_*''(s) & \stackrel{(1.2)}{=} & F_*''(t_0 s) \stackrel{(3.10)}{\preceq} \frac{1}{(1-\beta)^2} F_*''(-F'(u)) \\
& \stackrel{(1.9)}{=} & \frac{1}{(1-\beta)^2} [F''(u)]^{-1} \stackrel{(1.2)}{=} \frac{1}{t_0(1-\beta)^2} [F''(w)]^{-1}.
\end{array}
$$

Thus,

$$
\begin{array}{rcl}
r^2 & \leq & \frac{t_0}{(1-\beta)^2} \left[\langle F''(w)\Delta x, \Delta x\rangle + \langle \Delta s, [F''(w)]^{-1}\Delta s\rangle \right] \\
& \stackrel{(4.6)}{=} & \frac{t_0 \langle s,x\rangle}{(1-\beta)^2} \stackrel{(3.9)}{\leq} \frac{(\beta+\sqrt{\nu})^2}{(1-\beta)^2}.
\end{array}
$$

We have proved the following statement.

Theorem 3 *For all* $\alpha \in \left[0, \frac{1-\beta}{\beta+\sqrt{\nu}}\right)$ *we have*

$$\Omega(z \pm \alpha \Delta z) - \Omega(z) \quad \leq \quad \alpha\beta^2 \cdot \tfrac{\beta+\sqrt{\nu}}{1-\beta} + \omega\left(\alpha \cdot \tfrac{\beta+\sqrt{\nu}}{1-\beta}\right). \tag{4.9}$$

If the barrier F *is self-scaled, then*

$$\Omega(z \pm \alpha \Delta z) - \Omega(z) \quad \leq \quad \omega\left(\alpha \cdot \tfrac{\beta+\sqrt{\nu}}{1-\beta}\right). \tag{4.10}$$

Proof:
We need to justify only the second statement. Indeed, if the barrier F is self-scaled, then the first inequality in (4.1) implies $F'(x) = F''(w)F'_*(s)$ (see Theorem 3.2 in [8]). Taking into account this simplification, we come to the estimate (4.10). □

Theorem 3 justifies the high quality of the tangent direction Δz. To the best of our knowledge, in all existing general approaches the coefficient of α in the right-hand side of estimate (4.9) depends linearly on the centering parameter β (see, for example, Lemma 10 in [6]). In our case, this dependence is quadratic. Another advantage of definition (4.2) is that it automatically results in the standard search directions when our barrier appears to be self-scaled.

5 Non-symmetric primal-dual IPM

We are ready now to analyze a new predictor-corrector path-following method. It is controlled by two parameters $\beta, \gamma \in (0, 1)$. At each iteration of the method we perform two operations.

Prediction
Input: Point $u \in \mathrm{rint}\, \mathcal{F}_P$ and penalty value $t > 0$ such that $$\lambda_t(u) \leq \beta.$$ **1.** Compute $z = z_t(u)$ by (3.5). Set $w = \sqrt{t} \cdot u$. **2.** Compute Δz by (4.2) and find $\alpha > 0$ such that $$\Omega(z - \alpha\Delta z) = \Omega(z) + \beta^2\gamma + \omega(\gamma).$$ **Output:** $p_t(u) \stackrel{\mathrm{def}}{=} z - \alpha\Delta z.$

(5.1)

In order to avoid unnecessary complications, we assume that it is possible to find an exact solution to the equation of Step 2 in (5.1).

Lemma 4 *The output of the predictor step (5.1) satisfies the following inequalities:*

$$t(p_t(u)) \quad \geq \quad t \cdot \exp\left(\gamma \cdot \tfrac{1-\beta}{\beta+\sqrt{\nu}} - \tfrac{2\beta}{\sqrt{\nu}}\right), \tag{5.2}$$

$$\Omega(p_t(u)) \quad \leq \quad 2\omega(\beta) + \beta^2(1+\gamma) + \omega(\gamma). \tag{5.3}$$

Proof:
Denote $\tau = \alpha \cdot \frac{\beta+\sqrt{\nu}}{1-\beta}$. In view of the rule of Step 2 in (5.1), we have

$$\beta^2\gamma + \omega(\gamma) \quad = \quad \Omega(z - \alpha\Delta z) - \Omega(z) \overset{(4.9)}{\leq} \beta^2\tau + \omega(\tau).$$

Thus, $\tau \geq \gamma$, and we obtain

$$\begin{aligned} t(z - \alpha\Delta z) \quad & \overset{(4.3)}{=} \quad \tfrac{\nu}{\langle s,x\rangle - \alpha(\langle s,\Delta x\rangle + \langle \Delta x, s\rangle)} \overset{(4.4)}{=} \tfrac{1}{1-\alpha} \cdot \tfrac{\nu}{\langle s,x\rangle} \\ & \overset{(3.9)}{\geq} \quad \tfrac{1}{1-\alpha} \cdot \tfrac{t}{(1+\beta/\sqrt{\nu})^2} \geq t \cdot \exp\left(\alpha - \tfrac{2\beta}{\sqrt{\nu}}\right) \geq t \cdot \exp\left(\gamma \cdot \tfrac{1-\beta}{\beta+\sqrt{\nu}} - \tfrac{2\beta}{\sqrt{\nu}}\right). \end{aligned}$$

The remaining inequality (5.3) follows from (3.8). □

The second operation of our method is the correction process.

Correction

Input: Point $z \in \operatorname{rint} \mathcal{F}$ with penalty value $t = \langle s, x\rangle$.

1. Set $u_0 = x$.

2. while $\lambda_t(u_k) > \beta$ **do** $u_{k+1} := u_k + \frac{\delta_t(u_k)}{1+\lambda_t(u_k)}$.

Output: $\sigma_t(z) \overset{\text{def}}{=} u_N$, where u_N is the last point in the sequence.

(5.4)

Lemma 5 *The number of points generated in the correction process (5.4) is bounded as follows:*

$$N \quad \leq \quad \tfrac{\omega(z)}{\omega_*(\beta)}, \tag{5.5}$$

where $\omega_*(\tau) = \tau - \ln(1+\tau)$.

Proof:
Indeed, in Step 2 of (5.4) we apply the Damped Newton Method for the minimization of a self-concordant objective function $f_t(x)$ subject to the linear constraints $Ax = b$. In view of Theorem 4.1.12 in [5], at each iteration of the process the objective function is decreased at least by the value $\omega_*(\lambda_t(u_k))$. Since the process is running while $\lambda_t(u_k) > \beta$, we get the estimate (5.5) from Lemma 2. □

Now we can put the two pieces together.

Non-symmetric primal-dual IPM

Input: Point $u_0 \in \mathrm{rint}\, \mathcal{F}_P$ and penalty value $t_0 > 0$ such that $\lambda_{t_0}(u_0) \leq \beta$.

Iteration $k \geq 0$**:**

$$\begin{array}{rcl} \tilde{z}_k & = & p_{t_k}(u_k), \\ t_{k+1} & = & t(\tilde{z}_k), \\ u_{k+1} & = & \sigma_{t_{k+1}}(\tilde{z}_k). \end{array} \tag{5.6}$$

Theorem 4 *Let parameters* $\beta, \gamma \in (0, 1)$ *satisfy inequality*

$$\rho \stackrel{\mathrm{def}}{=} \gamma \cdot \tfrac{1-\beta}{1+\beta} - 2\beta \; > 0. \tag{5.7}$$

Then the rate of convergence of process (5.6) is given by

$$\langle s_k, x_k \rangle \;\leq\; \langle s_0, x_0 \rangle \cdot e^{-\rho k/\sqrt{\nu}}. \tag{5.8}$$

At the same time, the number of iterations in the correction process never exceeds

$$N(\beta, \gamma) \;\leq\; \tfrac{1}{\omega_*(\beta)} (2\omega(\beta) + \beta^2(1+\gamma) + \omega(\gamma)). \tag{5.9}$$

Proof:
Indeed, in view of inequality (5.2) we have

$$\begin{array}{rcl} t_{k+1} & \geq & t_k \cdot \exp\left(\gamma \cdot \tfrac{1-\beta}{\beta+\sqrt{\nu}} - \tfrac{2\beta}{\sqrt{\nu}}\right) \;=\; t_k \cdot \exp\left(\tfrac{1}{\sqrt{\nu}} \cdot \left[\gamma \cdot \tfrac{(1-\beta)\sqrt{\nu}}{\beta+\sqrt{\nu}} - 2\beta\right]\right) \\ & \geq & t_k \cdot \exp\left(\tfrac{1}{\sqrt{\nu}} \cdot \left[\gamma \cdot \tfrac{1-\beta}{1+\beta} - 2\beta\right]\right) \;=\; t_k \cdot \exp\left(\tfrac{\rho}{\sqrt{\nu}}\right). \end{array}$$

This proves (5.8). Inequality (5.9) follows from (5.3) and (5.5). □

Remark 2 *The main goal of Theorem 4 is to establish the polynomial-time complexity bounds for method (5.6). Therefore, we did not try to get the best possible value in the right-hand side of inequality (5.9). An immediate improvement can be obtained by dividing the correction process (5.4) into two stages:*

$$\{k:\; \lambda_t(u_k) \geq \tfrac{1}{3}\}, \quad \{k:\; \lambda_t(u_k) \in (\beta, \tfrac{1}{3})\}.$$

Then the duration of the first stage cannot exceed $\frac{2\omega(\beta)+\beta^2(1+\gamma)+\omega(\gamma)}{\omega_*(\frac{1}{3})}$ *iterations. The duration of the second stage is very short since it corresponds already to the region of quadratic convergence of the Damped Newton Method. We leave to the reader a possibility to play with the numbers and improve the estimate (5.9).*

6 Non-symmetric primal-dual IPM for Phase I

In order to start method (5.6), it is necessary to have a scaling point in a small neighborhood of the primal central path. This point can be found by an auxiliary procedure, which is called Phase I of the method. Let us show that such a procedure can be implemented in the same vein as the main scheme (5.6).

We need the following non-restrictive assumption.

Assumption 1 *The primal feasible set $\mathcal{F}_P$ is bounded.*

Note that assumption (2.4) and Theorem 1 guarantee only the boundedness of optimal set in the primal-dual problem (2.12). Hence, $\mathcal{F}_P$ may be unbounded. However, if we know a point $x_0 \in \operatorname{rint} \mathcal{F}_P$, then we can modify the initial problem (2.1) as follows:

$$\min_{\kappa,x} \{-\kappa : \ \langle c, x\rangle + \kappa = \langle c, x_0\rangle + 1, \ Ax = b, \ x \in K, \ \kappa \geq 0\}. \tag{6.1}$$

This problem has the same form as (2.1), but now its feasible set is bounded.

From now on, we assume that Assumption 1 holds and we know a feasible point $x_1 \in \operatorname{rint} \mathcal{F}_P$. Let us define an auxiliary central path

$$\hat{x}(\tau) \quad = \quad \arg\min_x \{-\tau\langle F'(x_1), x\rangle + F(x) : \ Ax = b\}, \quad \tau \geq 0. \tag{6.2}$$

Since $\mathcal{F}_P$ is bounded, this trajectory is well defined for all $\tau \geq 0$. Moreover, $\hat{x}(1) = x_1$.

Our goal is to trace $\hat{x}(\tau)$ as $\tau \to 0$. For that, we need to modify the sense of some notation. Up to the end of this section, we assume that definitions (3.3), (3.4) and (3.5) correspond to $c = -F'(x_1)$. The definition of the prediction step has also changed.

Prediction at Phase I (6.3)

Input: Point $v \in \operatorname{rint} \mathcal{F}_P$ and penalty value $\tau > 0$ such that

$$\lambda_t(v) \leq \beta.$$

1. Compute $z = z_\tau(v)$ by (3.5). Set $w = \sqrt{\tau} \cdot v$.

2. Compute Δz by (4.2) and find $\alpha > 0$ such that

$$\Omega(z + \alpha\Delta z) = \Omega(z) + \beta^2\gamma + \omega(\gamma).$$

Output: $p_\tau^+(v) \stackrel{\text{def}}{=} z + \alpha\Delta z.$

Lemma 6 *The output of the predictor step (6.3) satisfies the following inequalities:*

$$t(p_\tau^+(v)) \quad \leq \quad \tau \cdot \exp\left(-\frac{\gamma\cdot(1-\beta)}{\gamma\cdot(1-\beta)+\beta+\sqrt{\nu}} + \frac{2\beta}{\sqrt{\nu}-\beta}\right), \tag{6.4}$$

$$\Omega(p_\tau^+(v)) \quad \leq \quad 2\omega(\beta) + \beta^2(1+\gamma) + \omega(\gamma). \tag{6.5}$$

Proof:
As in the proof of Lemma 4, we argue that $\alpha \cdot \frac{\beta+\sqrt{\nu}}{1-\beta} \geq \gamma$. Therefore

$$\begin{array}{rcl} t(z+\alpha\Delta z) & \stackrel{(4.3)}{=} & \frac{\nu}{\langle s,x\rangle+\alpha(\langle s,\Delta x\rangle+\langle \Delta x,s\rangle)} \stackrel{(4.4)}{=} \frac{1}{1+\alpha}\cdot\frac{\nu}{\langle s,x\rangle} \\ & \stackrel{(3.9)}{\leq} & \frac{1}{1+\alpha}\cdot\frac{t}{(1-\beta/\sqrt{\nu})^2} \leq t\cdot\exp\left(-\frac{\alpha}{1+\alpha}+\frac{2\beta}{\sqrt{\nu}-\beta}\right) \\ & \leq & \tau\cdot\exp\left(-\frac{\gamma\cdot(1-\beta)}{\gamma\cdot(1-\beta)+\beta+\sqrt{\nu}}+\frac{2\beta}{\sqrt{\nu}-\beta}\right). \end{array}$$

The remaining inequality (6.5) follows from (3.8). □

In Phase I, we can use the correction process (5.4) without any change. Thus, we come to the following scheme.

Non-symmetric primal-dual IPM for Phase I
Input: Point $v_0 = x_1 \in \operatorname{rint} \mathcal{F}_P$ and penalty $\tau_0 = 1$. **Iteration** $k \geq 0$**:** $\tilde{z}_k = p^+_{\tau_k}(v_k)$, $\tau_{k+1} = t(\tilde{z}_k)$, $v_{k+1} = \sigma_{\tau_{k+1}}(\tilde{z}_k)$. **Stopping criterion:** $\lambda_0(v_k) \leq 2\beta$.

(6.6)

Theorem 5 *Let parameters $\beta, \gamma \in (0,1)$ satisfy inequality*

$$\rho_1 \stackrel{\text{def}}{=} \frac{\gamma\cdot(1-\beta)^2}{\gamma\cdot(1-\beta)+\beta+1} - 2\beta > 0. \tag{6.7}$$

Then the rate of convergence of process (6.6) is given by

$$\tau_k \leq e^{\frac{-k\rho_1}{\sqrt{\nu}-\beta}}. \tag{6.8}$$

At the same time, the number of iterations in the correction process never exceeds

$$N(\beta,\gamma) \leq \frac{1}{\omega_*(\beta)}(2\omega(\beta)+\beta^2(1+\gamma)+\omega(\gamma)). \tag{6.9}$$

The proof of this theorem is very similar to that of Theorem 4.

For a direction $g \in E^*$ and primal point $u \in \operatorname{rint} \mathcal{F}_P$ define now the local norm

$$|g|^*_v \stackrel{\text{def}}{=} \left[\max_{\delta}\{2\langle g,\delta\rangle - \|\delta\|^2_v : A\delta = 0\}\right]^{1/2}.$$

Note that (6.6) keeps $\lambda_{\tau_k}(v_k) = |F'(v_k) - \tau_k F'(x_1)|^*_{v_k} \leq \beta$. Therefore

$$\lambda_0(v_k) \quad = \quad |F'(u_k)|^*_{v_k} \; \leq \beta + \tau_k |F'(x_1)|^*_{v_k}.$$

From this inequality, it is easy to derive that the process (6.6) terminates in

$$O(\sqrt{\nu} \ln |F'(x_1)|^*_{\hat{x}(0)})$$

iterations (see Section 4.2.5 in [5] for details).

Further, let l be the last iteration of process (6.6). For our actual objective vector $c \in E^*$, we can choose coefficient $t_0 = \beta / |c|^*_{v_l}$. Then, choosing $\tilde{u} = v_l$, we get

$$|t_0 c + F'(\tilde{u})|^*_{\tilde{u}} \quad \leq \quad 3\beta.$$

If β is small enough, one iteration of the Damped Newton Method from $\tilde{u}$ results in the point u_0, which satisfies the input conditions of (5.6).

7 Primal barrier for epigraph of power function

Let us fix some $\alpha \in (0,1)$. Consider the following homogeneous function of two variables:

$$\xi(x,y) = x^{\alpha} \cdot y^{1-\alpha}, \quad (x,y) \in R^2_+.$$

We are going to find a barrier description of the cone

$$Q_{\alpha} = \{(x,y,z) \in R^3 : \; \xi(x,y) \geq |z|\}.$$

Since function $\xi(\cdot)$ is concave, the cone Q_{α} is convex. In this section we construct a self-concordant barrier for the set Q_{α} using a well-known 2-self-concordant barrier for the cone $\{(\tau,z) \in R^2 : \; \tau \geq |z|\}$, that is

$$f_1(\tau,z) \quad = \quad -\ln(\tau^2 - z^2).$$

First of all, let us mention some properties of the function ξ. Let us fix a point $(x,y) > 0$. Consider an arbitrary direction $d = (x',y') \in R^2$. Denote by

$$\xi_0 \; = \; \xi(x,y), \quad \xi_k \; = \; D^k \xi(x,y)[\underbrace{d,\ldots,d}_{k \text{ times}}], \quad k = 1 \ldots 3,$$

the directional derivatives of function ξ. Let $\delta_x = \frac{x'}{x}$, and $\delta_y = \frac{y'}{y}$.

Lemma 7 *We have the following relations:*

$$\begin{aligned} \xi_1 &= \xi_0 \cdot [\alpha \delta_x + (1-\alpha)\delta_y], \\ \xi_2 &= -\xi_0 \cdot \alpha(1-\alpha)(\delta_x - \delta_y)^2, \qquad (7.1) \\ \xi_3 &= -\xi_2 \cdot [(2-\alpha)\delta_x + (1+\alpha)\delta_y]. \end{aligned}$$

Proof:
Indeed,

$$D\xi(x,y)[d] = \alpha \cdot x^{\alpha-1} \cdot y^{1-\alpha} \cdot x' + (1-\alpha) \cdot x^{\alpha} \cdot y^{-\alpha} \cdot y' = \xi_0 \cdot (\alpha\delta_x + (1-\alpha)\delta_y),$$

$$\begin{aligned} D^2\xi(x,y)[d,d] &= \alpha(\alpha-1) \cdot x^{\alpha-2} \cdot y^{1-\alpha} \cdot (x')^2 + 2\alpha(1-\alpha) \cdot x^{\alpha-1} \cdot y^{-\alpha} \cdot x' \cdot y' \\ &\quad -\alpha(1-\alpha) \cdot x^{\alpha} \cdot y^{-\alpha-1} \cdot (y')^2 = -\xi_0 \cdot \alpha(1-\alpha)(\delta_x - \delta_y)^2. \end{aligned}$$

Therefore, from the last equation we have

$$\begin{aligned} D^3\xi(x,y)[d,d,d] &= -\xi_1 \cdot \alpha(1-\alpha)(\delta_x-\delta_y)^2 - \xi_0 \cdot 2\alpha(1-\alpha)(\delta_x-\delta_y)(-\delta_x^2+\delta_y^2) \\ &= \alpha(1-\alpha)(\delta_x-\delta_y)^2 \cdot [-\xi_1 + \xi_0 \cdot 2(\delta_x+\delta_y)] \\ &= \xi_0 \cdot \alpha(1-\alpha)(\delta_x-\delta_y)^2 \cdot [(2-\alpha)\delta_x + (1+\alpha)\delta_y]. \qquad \Box \end{aligned}$$

In this and the next section it is convenient to use ∇-notation for partial derivatives of different order. Let us prove the main statement of this section.

Theorem 6 *For any $\alpha \in [0,1]$ the function*

$$F_\alpha(x,y,z) = -\ln\left(x^{2\alpha} \cdot y^{2(1-\alpha)} - z^2\right) - \ln x - \ln y$$

is a 4-self-concordant barrier for the cone

$$Q_\alpha = \left\{(x,y,z) \in R^2_+ \times R: \; x^\alpha \cdot y^{1-\alpha} \geq |z|\right\}.$$

Proof:
Assume $\alpha \in (0,1)$. Consider $f_2(x,y,z) = f_1(\xi(x,y),z)$. Let us write down the derivatives of this function along the direction $\hat{d} = (x',y',z') \equiv (d,z')$. Denote $l = (\xi_1, z')$. Then

$$\begin{aligned} \Delta_1 &\stackrel{\text{def}}{=} Df_2(x,y,z)[\hat{d}] = \nabla_\tau f_1(\xi_0,z) \cdot \xi_1 + \nabla_z f_1(\xi_0,z) \cdot z' = \langle \nabla f_1(\xi_0,z), l\rangle, \\ \Delta_2 &\stackrel{\text{def}}{=} D^2 f_2(x,y,z)[\hat{d},\hat{d}] \\ &= \nabla^2_{\tau\tau} f_1(\xi_0,z) \cdot \xi_1^2 + \nabla_\tau f_1(\xi_0,z) \cdot \xi_2 + 2\nabla^2_{\tau z} f_1(\xi_0,z) \cdot \xi_1 \cdot z' + \nabla^2_{zz} f_1(\xi_0,z) \cdot (z')^2 \\ &= \langle \nabla^2 f_1(\xi_0,z) l, l\rangle + \nabla_\tau f_1(\xi_0,z) \cdot \xi_2 \stackrel{\text{def}}{=} \sigma_1 + \sigma_2. \end{aligned}$$

Note that $\sigma_1, \sigma_2 \geq 0$. Finally, denoting $e_1 = (1,0) \in R^2$, we get

$$\begin{aligned} \Delta_3 &\stackrel{\text{def}}{=} D^3 f_2(x,y,z)[\hat{d},\hat{d},\hat{d}] \\ &= \nabla^3_{\tau\tau\tau} f_1(\xi_0,z) \cdot \xi_1^3 + 3\nabla^3_{\tau\tau z} f_1(\xi_0,z) \cdot \xi_1^2 z' + 3\nabla^2_{\tau\tau} f_1(\xi_0,z) \cdot \xi_1 \cdot \xi_2 \\ &\quad +3\nabla^2_{\tau z} f_1(\xi_0,z) \cdot \xi_2 z' + \nabla_\tau f_1(\xi_0,z) \cdot \xi_3 + 3\nabla^2_{\tau zz} f_1(\xi_0,z) \cdot \xi_1 \cdot (z')^2 \\ &\quad +\nabla^3_{zzz} f_1(\xi_0,z) \cdot (z')^3 \\ &= D^3 f_1(\xi_0,z)[l,l,l] + 3\xi_2 \langle \nabla^2 f_1(\xi_0,z) l, e_1\rangle + \nabla_\tau f_1(\xi_0,z) \cdot \xi_3. \end{aligned}$$

Since f_1 is a self-concordant barrier for a cone with recession direction e_1, we have

$$\begin{aligned}
\xi_2 \cdot \langle \nabla^2 f_1(\xi_0, z) l, e_1 \rangle &\le \langle \nabla^2 f_1(\xi_0, z) l, l \rangle^{1/2} \cdot (-\xi_2) \cdot \langle \nabla^2 f_1(\xi_0, z) e_1, e_1 \rangle^{1/2} \\
&\le \langle \nabla^2 f_1(\xi_0, z) l, l \rangle^{1/2} \cdot (\xi_2 \cdot \langle \nabla f_1(\xi_0, z), e_1 \rangle) \\
&= \sigma_1^{1/2} \cdot \sigma_2.
\end{aligned}$$

Denote $\sigma_3 = \delta_x^2 + \delta_y^2$. Then, using self-concordance of f_1 again, by (7.1) we obtain

$$\begin{aligned}
\Delta_3 &\le 2\sigma_1^{3/2} + 3\sigma_1^{1/2} \cdot \sigma_2 - \sigma_2 \cdot [(2-\alpha)\delta_x + (1+\alpha)\delta_y] \\
&\le 2\sigma_1^{3/2} + 3\sigma_1^{1/2} \cdot \sigma_2 + \sigma_2 \cdot \sigma_3^{1/2} \cdot \left[(2-\alpha)^2 + (1+\alpha)^2\right]^{1/2} \qquad (7.2) \\
&\le 2\sigma_1^{3/2} + 3\sigma_2 \cdot \left(\sigma_1^{1/2} + \sigma_3^{1/2}\right).
\end{aligned}$$

Finally, consider the function $f_3(x, y) = -\ln x - \ln y$. Clearly, this is a 2-self-concordant barrier for the positive orthant in R^2. Note that

$$F_\alpha(x, y, z) = f_2(x, y, z) + f_3(x, y).$$

Let us estimate its derivatives along direction $\hat{d}$.

$$\begin{aligned}
D_2 &\stackrel{\mathrm{def}}{=} D^2 F_\alpha(x, y, z)[\hat{d}, \hat{d}] = \Delta_2 + D^2 f_3(x, y)[d, d] = \sigma_1 + \sigma_2 + \sigma_3, \\
D_3 &\stackrel{\mathrm{def}}{=} D^3 F_\alpha(x, y, z)[\hat{d}, \hat{d}, \hat{d}] = \Delta_3 + D^3 f_3(x, y)[d, d, d] \\
&\stackrel{(7.2)}{\le} 2\sigma_1^{3/2} + 3\sigma_2 \cdot \left(\sigma_1^{1/2} + \sigma_3^{1/2}\right) + 2\sigma_3^{3/2} \\
&= \left(\sigma_1^{1/2} + \sigma_3^{1/2}\right) \cdot \left(3\sigma_2 + 2\sigma_1 + 2\sigma_2 - 2\sigma_1^{1/2}\sigma_3^{1/2}\right) \\
&= \left(\sigma_1^{1/2} + \sigma_3^{1/2}\right) \cdot \left(3D_2 - \left(\sigma_1^{1/2} + \sigma_3^{1/2}\right)^2\right) \le 2D_2^{3/2}.
\end{aligned}$$

Thus, we have proved that function $F_\alpha(x, y, z)$ is convex and satisfies the characteristic condition for derivatives of self-concordant functions, namely, $D_3 \le 2D_2^{3/2}$. It remains to note that this function is logarithmically homogeneous of degree four.

For $\alpha = 0$ and $\alpha = 1$ the statement of the theorem is trivial. □

Note that for $\alpha \to 0$, the set Q_α approaches the direct product $\{x \ge 0\} \times \{y \ge |z|\}$, for which the parameter of self-concordant barrier cannot be less than three.

Let us present now a self-concordant barrier for p-cone

$$K_p = \left\{(\tau, z) \in R \times R^n : \tau \ge \|z\|_{(p)}\right\}, \quad p \ge 1,$$

where $\|z\|_{(p)} = \left[\sum\limits_{i=1}^n |z^{(i)}|^p\right]^{1/p}$. Without loss of generality, assume $\alpha \stackrel{\mathrm{def}}{=} \frac{1}{p} \in (0, 1)$.

Theorem 7 *Point (τ, z) belongs to K_p if and only if there exist $x \in R^n_+$ satisfying conditions*

$$\begin{aligned} (x^{(i)})^\alpha \cdot \tau^{1-\alpha} &\geq |z^{(i)}|, \quad i = 1, \dots, n, \\ \sum_{i=1}^n x^{(i)} &= \tau. \end{aligned} \tag{7.3}$$

Thus, a self-concordant barrier for the cone K_p can be implemented by restricting the $(4n)$-self-concordant barrier

$$\Phi_p(\tau, x, z) = -\sum_{i=1}^n \left[\ln \left((x^{(i)})^{2\alpha} \cdot \tau^{2(1-\alpha)} - (z^{(i)})^2 \right) + \ln x^{(i)} + \ln \tau \right]$$

onto the linear hyperplane $\sum_{i=1}^n x^{(i)} = \tau$.

Proof:
Indeed, let the triple (τ, x, z) satisfy relations (7.3). Then

$$\tau = \sum_{i=1}^n x^{(i)}, \qquad x^{(i)} \cdot \tau^{p-1} \geq |z^{(i)}|^p, \quad i = 1, \dots, n,$$

Therefore, $\tau^p \geq \sum_{i=1}^n |z^{(i)}|^p$.

Consider now an arbitrary point $(\tau, z) \in K_p$. Without loss of generality assume $\tau > 0$. Then, we can choose

$$\begin{aligned} \epsilon &= \tau^p - \sum_{i=1}^n |z^{(i)}|^p \geq 0, \\ x^{(i)} &= \tfrac{1}{\tau^{p-1}} \left(\tfrac{\epsilon}{n} + |z^{(i)}|^p \right), \quad i = 1, \dots, n. \end{aligned}$$

Clearly, the triple (τ, x, z) satisfies relations (7.3). The remaining statements follow from Theorem 6. □

Alternatively, the cone K_p can be seen as an intersection of $Q^n_{1/p}$ by a linear subspace. Indeed, it is easy to see that $(\tau, z) \in K_p$ if and only if there exist non-negative vectors x and y from R^n such that

$$\begin{aligned} (x^{(i)}, y^{(i)}, z^{(i)}) &\in Q_{1/p}, \quad i = 1, \dots, n, \\ y^{(i)} &= \tau, \qquad i = 1, \dots, n, \\ \sum_{i=1}^n x^{(i)} &= \tau. \end{aligned} \tag{7.4}$$

The main advantage of such a representation consists in separability of the corresponding barrier function

$$\Psi_p(x, y, x) = \sum_{i=1}^n F_{1/p}(x^{(i)}, y^{(i)}, z^{(i)}).$$

In particular, this opens the possibility of employing the dual barriers.

8 Dual barrier

It is well known that the cone Q_α is self-dual. For completeness of presentation, let us provide this statement with a simple proof. Recall, that for a primal cone $K \subset E$, the dual cone is defined as follows:

$$K^* \;=\; \{s \in E^* : \; \langle s, x \rangle \geq 0 \; \forall x \in K\}.$$

The cone K is called *self-dual* if there exists a self-adjoint positive-definite operator $B : E \to E^*$ such that $K^* = BK$. In our situation, $E = R^3$, and $E^* = R^3$. For $x \in E$ and $s \in E^*$, we define the scalar product in the standard way:

$$\langle s, x \rangle \;=\; \sum_{i=1}^{3} s^{(i)} \cdot x^{(i)}.$$

Lemma 8 *For $\alpha \in (0,1)$, define*

$$B_\alpha \;=\; \begin{pmatrix} \alpha & 0 & 0 \\ 0 & 1-\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then $Q_\alpha^* = B_\alpha Q_\alpha$.

Proof:
Indeed, $s \in Q_\alpha^*$ if and only if

$$s^{(1)}x^{(1)} + s^{(2)}x^{(2)} + s^{(3)}x^{(3)} \geq 0 \quad \forall x : \; (x^{(1)})^\alpha (x^{(2)})^{1-\alpha} \geq |x^{(3)}|.$$

Hence,

$$s^{(1)}x^{(1)} + s^{(2)}x^{(2)} - |s^{(3)}| \cdot (x^{(1)})^\alpha (x^{(2)})^{1-\alpha} \geq 0 \quad \forall x^{(1)}, x^{(2)} \geq 0$$

Clearly, the only nontrivial case is $s^{(3)} \neq 0$ and $x^{(2)} > 0$. Then the minimum of the last expression in $x^{(1)}$ is achieved for $x^{(1)} = x^{(2)} \cdot \left(\frac{s^{(1)}}{\alpha |s^{(3)}|} \right)^{-\frac{1}{1-\alpha}}$. Substituting this value, we come to the following characteristic inequality of the dual cone:

$$\left(\frac{s^{(1)}}{\alpha} \right)^\alpha \left(\frac{s^{(2)}}{1-\alpha} \right)^{1-\alpha} \;\geq\; |s^{(3)}|.$$

This proves the required statement. □

Note that in the primal-dual setting we need to work with a pair of conjugate barriers for the primal and the dual cones. In our situation, for the primal cone Q_α we can use the barrier

$$F_\alpha(x) \;=\; -\ln\left((x^{(1)})^{2\alpha} (x^{(2)})^{2(1-\alpha)} - (x^{(3)})^2 \right) - \ln x^{(1)} - \ln x^{(2)}, \quad x \in \operatorname{rint} Q_\alpha.$$

Unfortunately, the conjugate barrier

$$F_\alpha^*(s) \;=\; \max_x [-\langle s, x \rangle - F_\alpha(x)], \quad s \in \operatorname{rint} Q_\alpha^*, \tag{8.1}$$

cannot be written in a closed form. However, there exists an efficient strategy for computing the value and the derivatives of this barrier .

Denote by $x(s)$ the solution of the optimization problem in (8.1). It is well known that

$$\begin{array}{rcl} \nabla F_\alpha^*(s) & = & -x(s), \\ \nabla^2 F_\alpha^*(s) & = & [\nabla^2 F(x(s))]^{-1}. \end{array} \tag{8.2}$$

Note that the point $x = x(s)$ can be found from the following system of nonlinear equations:

$$\begin{array}{rcl} s^{(1)} & = & \frac{2\alpha}{\omega} \cdot \frac{\beta}{x^{(1)}} + \frac{1}{x^{(1)}}, \\ s^{(2)} & = & \frac{2(1-\alpha)}{\omega} \cdot \frac{\beta}{x^{(2)}} + \frac{1}{x^{(2)}}, \\ s^{(3)} & = & -\frac{2x^{(3)}}{\omega}, \end{array} \tag{8.3}$$

where $\omega = (x^{(1)})^{2\alpha}(x^{(2)})^{2(1-\alpha)} - (x^{(3)})^2$, and $\beta = (x^{(1)})^{2\alpha}(x^{(2)})^{2(1-\alpha)} = \omega + (x^{(3)})^2$. Thus,

$$\begin{array}{rcl} x^{(1)} & = & \frac{1}{s^{(1)}} \left[1 + \frac{2\alpha\beta}{\omega}\right], \\ x^{(2)} & = & \frac{1}{s^{(2)}} \left[1 + \frac{2(1-\alpha)\beta}{\omega}\right], \\ x^{(3)} & = & -\frac{\omega}{2} s^{(3)}. \end{array} \tag{8.4}$$

Therefore,

$$\begin{array}{rcl} s^{(1)}x^{(1)} + s^{(2)}x^{(2)} + s^{(3)}x^{(3)} & = & 2 + \frac{2\beta}{\omega} - \frac{\omega}{2}(s^{(3)})^2 \\ & = & 2 + 2\,\frac{\beta - (x^{(3)})^2}{\omega} \;=\; 4, \\ (1-\alpha)s^{(1)}x^{(1)} - \alpha s^{(2)}x^{(2)} & = & 1 - 2\alpha. \end{array} \tag{8.5}$$

Consequently, we can express $x^{(1)}$ and $x^{(2)}$ as linear functions of $x^{(3)}$:

$$\begin{array}{rcl} s^{(1)}x^{(1)} & = & 1 + 2\alpha - \alpha s^{(3)}x^{(3)}, \\ s^{(2)}x^{(2)} & = & 3 - 2\alpha - (1-\alpha)s^{(3)}x^{(3)}. \end{array} \tag{8.6}$$

Then, substituting the above expressions in (8.1), we come to the following representation:

$$\begin{array}{rcl} F_\alpha^*(s) & = & \max\limits_{x^{(3)}} \Big[\ln\left(\left(\frac{1+2\alpha-\alpha s^{(3)}x^{(3)}}{s^{(1)}}\right)^{2\alpha} \left(\frac{3-2\alpha-(1-\alpha)s^{(3)}x^{(3)}}{s^{(2)}}\right)^{2(1-\alpha)} - (x^{(3)})^2 \right) \\ & & + \ln \frac{1+2\alpha-\alpha s^{(3)}x^{(3)}}{s^{(1)}} + \ln \frac{3-2\alpha-(1-\alpha)s^{(3)}x^{(3)}}{s^{(2)}} \Big] - 4. \end{array} \tag{8.7}$$

Without loss of generality, we can assume $s^{(3)} \neq 0$. Therefore, denoting $\tau = s^{(3)}x^{(3)}$ and ignoring the additive constant terms, we get the following univariate minimization problem:

$$f^* = \min_\tau [\, f(\tau) \stackrel{\mathrm{def}}{=} -\ln\left(q^2 \cdot (2+\tfrac{1}{\alpha}-\tau)^{2\alpha}(2+\tfrac{1}{1-\alpha}-\tau)^{2(1-\alpha)} - \tau^2\right) - \ln(2+\tfrac{1}{\alpha}-\tau) - \ln(2+\tfrac{1}{1-\alpha}-\tau)\,], \tag{8.8}$$

where $q = |s^{(3)}| \left(\frac{\alpha}{s^{(1)}}\right)^{\alpha}\left(\frac{1-\alpha}{s^{(2)}}\right)^{1-\alpha} < 1$. Let us show that this minimization problem can be easily solved up to any desired accuracy by a quadratically convergent procedure. Using its solution, the point $x(s)$ can be formed by (8.6), and then, applied in (8.1), (8.2) to compute the value, the gradient, and the Hessian of the dual barrier.

In view of Theorem 6, the objective function of problem (8.8) is a 4-self-concordant barrier. Note that $f'(0) < 0$. On the other hand, for any feasible τ we have

$$\begin{aligned} |\tau| &\le q \exp\left[\alpha \ln\left(2+\tfrac{1}{\alpha}-\tau\right) + (1-\alpha)\ln\left(2+\tfrac{1}{1-\alpha}-\tau\right)\right] \\ &\le q \exp[\,\ln(4-\tau)\,] = q\cdot(4-\tau). \end{aligned}$$

Hence, $\tau \ge -4r$, where $r \stackrel{\mathrm{def}}{=} \frac{q}{1-q}$. At the same time, for $\hat\tau = -2r$ we have

$$\begin{aligned} q^2\cdot(2+\tfrac{1}{\alpha}-\hat\tau)^{2\alpha}(2+\tfrac{1}{1-\alpha}-\hat\tau)^{2(1-\alpha)} &= q^2\cdot(2+\tfrac{1}{\alpha}+\tfrac{2q}{1-q})^{2\alpha}(2+\tfrac{1}{1-\alpha}+\tfrac{2q}{1-q})^{2(1-\alpha)} \\ &= q^2\cdot(\tfrac{1}{\alpha}+\tfrac{2}{1-q})^{2\alpha}(\tfrac{1}{1-\alpha}+\tfrac{2}{1-q})^{2(1-\alpha)} > \hat\tau^2. \end{aligned}$$

Thus, $\hat\tau \in \mathrm{dom}\, f$.

Denote by τ^* the optimal solution of problem (8.8) and let $\rho = \frac{1}{[f''(\tau^*)]^{1/2}}$. Since f is a self-concordant barrier with parameter $\nu = 4$, we have

$$\tau^* + (\nu + 2\sqrt{\nu})\rho\cdot[-1,1] \supseteq \mathrm{dom}\, f \supseteq [0,\hat\tau] \quad\Rightarrow\quad 16\rho \ge 2r.$$

Note, that for problem (8.8), the Newton method converges quadratically from any point τ_0, $|\tau_0 - \tau^*| \le \frac{1}{4}\rho$.[2] We can find such a point by a bisection procedure based on computations of the derivative $f'(\cdot)$ starting from the middle point of the initial interval $[0,-4r]$. Clearly, we need at most $k = 7$ bisection steps:

$$4r\cdot 2^{-k} \le 32\rho\cdot 2^{-k} = \tfrac{1}{4}\rho.$$

Acknowledgements. The author would like to thank Laurence Wolsey for useful comments on the text and the anonymous referees for their suggestions.

[2]This can be justified, for example, by the results of Section 4.1.5 [5]. Indeed, the Damped Newton Method converges quadratically from any point τ with $\lambda(\tau) \stackrel{\mathrm{def}}{=} \frac{|f'(\tau)|}{[f''(\tau)]^{1/2}} < \frac{1}{2}$. In view of Theorem 4.1.13 [5], the latter inequality follows from $|\tau - \tau^*| \le \frac{1}{4}\rho$.

References

[1] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization . *SIAM Journal of Optimization*, 5(1995), 13 – 51.

[2] R. W. Freund, F. Jarre, and S. Schaible. On self-concordant barrier functions for conic hulls and fractional programming. *Mathematical Programming*, 74 (1996), 237 – 246.

[3] O. Guler. Hyperbolic Polynomials and Interior Point Methods for Convex Programming. *Mathematics of Operations Research*, 22(1997), 350 – 377.

[4] A. Nemirovski, L. Tuncel. Cone-free path-following and potential reduction polynomial time interior-point algorithms. *Mathematical Programming*, On line First issue, DOI: 10.1007/s10107-004-0545-4 (2004).

[5] Yu. Nesterov. *Introductory Lectures on Convex Optimization*. Kluwer, Boston, 2004.

[6] Yu. Nesterov. Long-Step Strategies in Interior-Point Primal-Dual Methods. *Mathematical Programming*, 76(1996) 1, 47 – 94.

[7] Yu. Nesterov and A. Nemirovsky, *Interior Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, 1994.

[8] Yu. Nesterov, M. J. Todd. Self-scaled Barriers and Interior-Point Methods for Convex Programming. *Mathematics of Operation Research*, 22(1997) 1, 1 – 42.

[9] Yu. Nesterov, M. J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM Journal of Optimization*, 8(1998), 324 – 364.

[10] J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization*. MPS/SIAM Series on Optimization 3. SIAM Publications, Philadelphia, 2001.

[11] M. J. Todd, K. C. Toh and R. H. Tutuncu. On the Nesterov-Todd direction in semidefinite programming. *SIAM Journal on Optimization*, 8(1998), 769 – 796.

[12] L. Tuncel. Generalization of primal-dual interior-point methods to convex optimization problems in conic form. *Foundations of computational mathematics*, 1 (2001): 229 – 254.

[13] G. Xue, Y. Ye. An efficient algorithm for minimizing a sum of p-norms. *SIAM Journal on Optimization*, 10(1998) 2, 551 – 579.