CORE DISCUSSION PAPER

2006/34

Nonsymmetric potential-reduction methods for general cones

Yu. Nesterov *

March 28, 2006

Abstract

In this paper we propose two new nonsymmetric primal-dual potential-reduction methods for conic problems. The methods are based on the *primal-dual lifting* [5]. This procedure allows to construct a strictly feasible primal-dual pair related by an exact *scaling* relation even if the cones are not symmetric. It is important that all necessary elements of our methods can be obtained from the standard solvers for *primal* Newton system. The first of the proposed schemes is based on the usual affine-scaling direction. For the second one, we apply a new *first-order* affine-scaling direction, which incorporates in a symmetric way the gradients of primal and dual barriers. For both methods we prove the standard $O(\sqrt{\nu}\ln\frac{1}{\epsilon})$ complexity estimate, where ν is the parameter of the barrier and ϵ is the required accuracy.

Keywords: convex optimization, conic problems, interior-point methods, potential-reduction methods, self-concordant barriers, self-scaled barriers, affine-scaling direction.

*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. In the last years, the main activity in the field of interior-point schemes was related to symmetric cones (see [1], [8], [9]). The general cones did not attract to much attention maybe because of the difficulties related to constructing a good and computable self-concordant barrier and its Fenchel transform (see [2], [10]). However, recently in [6] it was shown that our abilities in constructing good barriers for convex sets and for convex cones are basically the same. For example, it is possible to develop self-concordant barriers for conic hulls of epigraphs of many important functions of one variable. The values of parameters of proposed barriers vary from three to four. This opens a possibility to solve efficiently by conic interior-point schemes many problems of *Separable Optimization.*

On the other hand, very often the computation of the Fenchel transform of the primal barrier is not so easy. That was the main motivation for developing in [5] a framework for *nonsymmetric* interior-point methods for conic problems. The main idea of [5] was to attribute the main part of computations to the primal problem. For that, we treat an auxiliary problem of finding a point in close neighborhood of primal central path as a process of finding a *scaling point.* Using this point w, we can *construct* a feasible primal-dual pair $z = (x, s)$, satisfying the *exact* scaling relation

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

where F is the primal barrier. This operation is called *primal-dual lifting.* In appears that the point z belongs to a neighborhood of the primal-dual central path. Moreover, using the Hessian $F''(w)$ it is possible to define *affine-scaling direction* exactly by the same expression as for self-scaled barrier. Note that now, even for self-scaled barriers, we obtain the standard search direction applying no special machinery. For all cones, we only need to solve a system of linear equations for *primal* Newton direction. In order to perform a prediction step, we need a procedure, which can compute a value of the dual barrier.

The main advantage of this approach is that computationally it is very cheap. In [5] there was analyzed nonsymmetric primal-dual path-following scheme based on the primal-dual lifting. In this paper we show that this lifting can be used for developing nonsymmetric primal-dual potential-reduction methods.

Contents. The paper is organized as follows. In Section 2 we introduce a primal-dual pair of conic problems and the primal-dual central path. We also present the definition and the main properties of primal-dual lifting. In Section 3, using the framework of [3], we introduce the first potential-reduction method based on affine-scaling direction. In this scheme, we apply consequently three stages: the primal correction process, primal-dual lifting, and prediction step. At each stage we can guarantee a certain decrease of a penalty potential. It is shown that even the lifting can cause a significant drop in the value of the potential. We prove also $O(\sqrt{\nu} \ln \frac{1}{\epsilon})$ complexity bound, where ν is the parameter of primal self-concordant barrier and ϵ is the required accuracy. Finally, in Section 4 we propose a *first-order* affine-scaling direction, which incorporates the gradients of the primal and dual barriers computed in the lifted point. We show that our definition possesses a primal-dual symmetry. At the same time, it ensures the same complexity bound as the standard affine-scaling direction.

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

$$\langle s, x\rangle \;\; = \;\; \sum_{i=1}^{n} s^{(i)} x^{(i)}, \quad \|x\| \;\stackrel{\text{def}}{=}\; \langle x, x\rangle^{1/2}, \quad s, x \in R^n.$$

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 consider self-adjoint operators: $A = A^*$.

Let Q be a closed convex set in E. For interior-point methods (IPM), Q must be represented by a self-concordant barrier $F(x)$, $x \in \operatorname{int} Q$, with parameter $\nu \geq 1$ (see Chapter 4 in [4] for definitions, examples and main results). At any $x \in \operatorname{int} Q$ we use the Hessian $F''(x) : E \to E^*$ for defining the following local Euclidean norms:

$$\|h\|_x \;\; = \;\; \langle F''(x)h, h\rangle^{1/2}, \quad h \in E,$$

$$\|s\|_x^* \;\; = \;\; \langle s, [F''(x)]^{-1} s\rangle, \quad s \in E^*.$$

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

$$W(x) \; = \; \{u \in E : \; \|u - x\|_x \leq 1\} \; \subseteq \; Q. \tag{1.1}$$

We often use two important inequalities:

$$F(u) \leq F(x) + \langle F'(x), u - x\rangle + \omega(r), \tag{1.2}$$

$$F''(u) \;\; \preceq \;\; \tfrac{1}{(1-r)^2} F''(x), \tag{1.3}$$

where $x \in \operatorname{int} Q$, $r = \|u - x\|_x < 1$, and $\omega(t) = -t - \ln(1-t)$. Note that a sum of a self-concordant barrier $F(x)$ and a linear function $\langle c, x\rangle$ becomes a self-concordant function. An iterate of Damped Newton method as applied to $f(x) = \langle c, x\rangle + F(x)$ looks as follows:

$$x_+ \;\; = \;\; x - \tfrac{[F''(x)]^{-1} f'(x)}{1 + \|f'(x)\|_x^*}, \quad x \in \operatorname{int} Q. \tag{1.4}$$

It can be shown that

$$f(x) - f(x_+) \;\; \geq \;\; \omega_*(\|f'(x)\|_x), \tag{1.5}$$

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

In many applications, the feasible set Q can be represented as an intersection of an affine subspace and a convex cone $K \subseteq E$. We call K *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.

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

$$F(\tau x) \;\equiv\; F(x) - \nu \ln \tau, \quad x \in \operatorname{int} K,\; \tau > 0. \tag{1.6}$$

Let us point out some straightforward consequences of this identity:

$$F'(\tau x) \;=\; \tfrac{1}{\tau} F'(x), \quad F''(\tau x) \;=\; \tfrac{1}{\tau^2} F''(x), \tag{1.7}$$

$$F''(x)x \;=\; -F'(x), \tag{1.8}$$

$$\langle F'(x), x\rangle \;=\; -\nu, \tag{1.9}$$

$$\langle F''(x)x, x\rangle \;=\; \nu, \quad \langle F'(x), [F''(x)]^{-1} F'(x)\rangle \;=\; \nu, \tag{1.10}$$

(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:

$$-F'(x) \in \operatorname{int} K, \quad -F_*' \in \operatorname{int} K^*, \tag{1.11}$$

$$\begin{aligned} F_*(-F'(x)) \;&=\; \langle F'(x), x\rangle - F(x) \;=\; -\nu - F(x), \\ F(-F_*'(s)) \;&=\; -\nu - F(x). \end{aligned} \tag{1.12}$$

$$F_*'(-F'(x)) \;=\; -x, \quad F'(-F_*'(s)) \;=\; -s, \tag{1.13}$$

$$F_*''(-F'(x)) \;=\; [F''(x)]^{-1}, \quad F''(-F_*'(s)) \;=\; [F_*''(s)]^{-1}, \tag{1.14}$$

$$F(x) + F_*(s) \;\geq\; -\nu - \nu \ln \tfrac{\langle s, x\rangle}{\nu}, \tag{1.15}$$

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]).

2 Primal-dual pair of conic problems

Consider the standard conic optimization problem

$$\begin{aligned} f^* \;=\; & \min_x \langle c, x\rangle, \\ \text{s.t.} \;\; & x \in \mathcal{F}_P \overset{\text{def}}{=} \{x \in K :\; Ax = b\}, \end{aligned} \tag{2.1}$$

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

$$\begin{aligned} & \max_{s,y} \langle b, y\rangle, \\ \text{s.t.} \;\; & \left.\begin{aligned} & s + A^* y = c, \\ & s \in K^*,\; y \in R^m. \end{aligned}\right\} \overset{\text{def}}{=} \mathcal{F}_D \end{aligned} \tag{2.2}$$

For a feasible primal-dual point $z = (x, s, y)$ the following relation holds

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

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, for problems (2.1), (2.2) the strong duality holds.

We assume that the primal cone is endowed with a ν-logarithmically homogeneous self-concordant barrier $F(x)$. Then, the barriers $F(x)$ and $F_*(s)$ define the *primal-dual central path* (see, for example, [3]).

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

$$\left.\begin{array}{rcl} x(t) & = & \arg\min\limits_x \{ t\langle c, x\rangle + F(x) :\; Ax = b\} \\[2ex] y(t) & = & \arg\max\limits_y \{ t\langle b, y\rangle - F_*(c - A^* y)\} \\[2ex] s(t) & = & c - A^* y(t) \end{array}\right\}, \quad t > 0, \tag{2.5}$$

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.

Modern IPM usually work directly with primal-dual problem

$$\begin{array}{c} \min\limits_{x,s,y} \{ \langle c, x\rangle - \langle b, y\rangle :\; (x, s, y) \in \mathcal{F}\}, \\[2ex] \mathcal{F} \;=\; \{(x, s, y) :\; Ax = b,\; s + A^* y = c,\; x \in K,\; s \in K^*\}. \end{array} \tag{2.9}$$

Note that (2.6) justifies unboundedness of $\mathcal{F}$ (consider $t \to 0$).

The main advantage of (2.9) 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 [3], [8])

$$\begin{array}{rcl} \Omega(x, s, y) & \stackrel{(1.15)}{=} & F(x) + F_*(s) + \nu \ln \frac{\langle s, x\rangle}{\nu} + \nu \\[2ex] & \stackrel{(2.3)}{=} & F(x) + F_*(s) + \nu \ln \frac{\langle c, x\rangle - \langle b, y\rangle}{\nu} + \nu \;\geq\; 0. \end{array} \tag{2.10}$$

This function vanishes only at points of the primal-dual central path.

Recently, in [5] there was proposed a *primal-dual lifting* procedure, which transforms a well centered point $u \in \operatorname{rint} \mathcal{F}_P$ into a well centered primal-dual point $z_t(u)$. Namely, let us fix a penalty parameter $t > 0$ and point $u \in \mathrm{rint}\, \mathcal{F}_P$. Consider directions $\delta = \delta_t(u)$, and $y = y_t(u)$, which provide a unique solution to the following linear system:

$$\begin{array}{rcl} tc + F'(u) + F''(u)\delta & = & tA^* y, \\ A\delta & = & 0. \end{array} \tag{2.11}$$

Note that this system corresponds to the standard Newton step:

$$\delta_t(u) \quad = \quad \arg\min_{A\delta=0} \left\{ \langle tc + F'(u), \delta \rangle + \tfrac{1}{2} \langle F''(u)\delta, \delta \rangle \right\}.$$

Define primal-dual lifting $z_t(u) = (x_t(u), s_t(u), y_t(u))$ of point u as follows:

$$x_t(u) = u - \delta_t(u), \quad s_t(u) = c - A^* y_t(u). \tag{2.12}$$

Note that $x_t(u)$ is formed by a shift from u along a direction pointing away from the central path. Denote $\lambda_t(u) = \|\delta_t(u)\|_u$.

Theorem 2 **[5]** *If $\lambda_t(u) \leq \beta < 1$, then $z_t(u) \in \mathrm{rint}\, \mathcal{F}$ satisfies the scaling relations*

$$\begin{array}{rcl} s_t(u) & = & \frac{1}{t} F''(u) \cdot x_t(u), \\ \|F'(x_t(u)) - \frac{1}{t} F''(u) \cdot F'_*(s_t(u))\|^*_u & \leq & \frac{2\beta^2}{1-\beta}. \end{array} \tag{2.13}$$

Moreover, it is well centered:

$$\Omega(z_t(u)) \quad \leq \quad 2\omega(\beta) + \beta^2, \tag{2.14}$$

and its duality gap is bounded as follows:

$$\ln\left(1 - \tfrac{\beta}{\sqrt{\nu}}\right) \quad \leq \quad \tfrac{1}{2} \ln\left(\tfrac{t}{\nu} \cdot \langle s_t(u), x_t(u) \rangle\right) \quad \leq \quad \ln\left(1 + \tfrac{\beta}{\sqrt{\nu}}\right). \tag{2.15}$$

Finally, we have the following bounds on the Hessians:

$$F''(x_t(u)) \preceq \tfrac{1}{(1-\beta)^2} F''(u), \quad F''_*(s_t(u)) \preceq \tfrac{t^2}{(1-\beta)^2} [F''(u)]^{-1}. \tag{2.16}$$

In [5] it was shown how to use the primal-dual lifting in predictor-corrector path-following scheme. In the next section we apply this procedure in the potential-reduction framework.

3 Potential-reduction IPM

One of the most useful functional characteristics of point $z = (x, s, y) \in \mathrm{rint}\, \mathcal{F}$ is the *primal-dual potential*

$$\begin{array}{rcl} \Phi(x, s, y) & = & F(x) + F_*(s) + 2\nu \ln\langle s, x \rangle \\ & \overset{(2.3)}{=} & F(x) + F_*(s) + 2\nu \ln[\langle c, x \rangle - \langle b, y \rangle]. \end{array} \tag{3.1}$$

This potential provides us with an easily computable upper bound for the *duality gap*:

$$\langle c, x\rangle - \langle b, y\rangle \quad \leq \quad \tfrac{1}{\nu} \exp\left(1 + \tfrac{1}{\nu}\Phi(z)\right). \tag{3.2}$$

It can be proved [3], that the region of the fastest decrease of Φ is located in a small neighborhood of the primal-dual central path. Therefore, *potential reduction IPM* tempt to decrease the *penalty potential*:

$$\begin{aligned} P_\gamma(z) \quad &= \quad \Phi(z) + \gamma\Omega(z) \stackrel{\mathrm{def}}{=} (1+\gamma)P^0_\gamma(z), \quad \gamma > 0, \\ P^0_\gamma(z) \quad &\stackrel{\mathrm{def}}{=} \quad F(x) + F_*(s) + (\nu+\rho)\ln\langle s, x\rangle + \tfrac{\gamma}{1+\gamma}\cdot \nu \cdot (1 - \ln \nu), \\ &= \quad \Omega(z) + \rho \ln\langle s, x\rangle - \tfrac{\nu}{1+\gamma}(1-\ln\nu), \quad z \in \operatorname{rint} \mathcal{F}, \end{aligned} \tag{3.3}$$

where $\rho \stackrel{\mathrm{def}}{=} \frac{\nu}{1+\gamma} < \nu$. Thus, inequality (3.2) implies

$$\langle c, x\rangle - \langle b, y\rangle \quad \leq \quad \tfrac{1}{\nu} \exp\left(1 + \tfrac{1}{\rho}P^0_\gamma(z)\right). \tag{3.4}$$

Let us discuss different strategies for decreasing the *normalized penalty potential* $P^0_\gamma(z)$. We are going to present two nonsymmetric primal-dual methods empoying mainly a primal barrier $F(x)$, which is assumed to be easily available. 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 (2.9). 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:

$$\min_{\kappa, x} \{-\kappa : \ \langle c, x\rangle + \kappa = \langle c, x_0\rangle + 1, \ Ax = b, \ x \in K, \ \kappa \geq 0\}. \tag{3.5}$$

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

Nonsymmetric primal-dual potential-reduction IPM consists of four stages.

0. Initialization. Choose arbitrary $x_0 \in \operatorname{rint} \mathcal{F}_P$ and an estimate $f_0 < f^*$. Since $\mathcal{F}_P$ is bounded, the dual feasible set must be unbounded. Therefore, for any $f_0 < f^*$ there exists a point $(s_0, y_0) \in \operatorname{rint} \mathcal{F}_D$ such that $f_0 = \langle b, y_0\rangle$. This point is used only for interpreting our lower bound. Denote

$$\psi_k(x) \quad = \quad (\nu+\rho)\ln(\langle c, x\rangle - f_k) + F(x), \quad k \geq 0.$$

We need to choose a tolerance parameter $\beta \in (0, 1)$.

1. Primal kth stage ($k \geq 0$). Set $u_0 = x_k$, $i = 0$.

Repeat

a) Find solution (δ_i, y_i) of the following linear system:

$$\begin{aligned} \tfrac{\nu+\rho}{\langle c, u_i \rangle - f_k} \cdot c + F'(u_i) + F''(u_i)\delta_i &= A^* y_i, \\ A\delta_i &= 0. \end{aligned} \tag{3.6}$$

b) Compute $\lambda_i = \|\delta_i\|_{u_i}$. If $\lambda_i > \beta$, then $u_{i+1} = u_i + \frac{\delta_i}{1+\lambda_i}$.

until $\lambda_i \leq \beta$**.**

Denote $g_i = \frac{\nu+\rho}{\langle c, u_i \rangle - f_k} \cdot c + F'(u_i)$. Note that

$$\begin{aligned} \psi_k(u_i + \delta) - \psi_k(u_i) &= (\nu + \rho) \ln \left(1 + \tfrac{\langle c, \delta \rangle}{\langle c, u_i \rangle - f_k} \right) + F(u_i + \delta) - F(u_i) \\ &\leq \langle g_i, \delta \rangle + F(u_i + \delta) - F(u_i). \end{aligned}$$

Therefore the Step b) in (3.6) can be interpreted as a Damped Newton step (1.4) with λ_i being the local norm of the gradient g_i. Hence, in view of (1.5), if $\lambda_i \geq \beta$, then

$$\psi_k(u_i) - \psi_k(u_{i+1}) \;\geq\; \omega_*(\beta).$$

Hence, if the termination criterion is satisfied at step i_k, then for any i, $0 \leq i \leq i_k$, we have

$$P^0_\gamma(u_i, s_k) \;\leq\; P^0_\gamma(x_k, s_k) - i \cdot \omega_*(\beta). \tag{3.7}$$

2. Primal-dual lifting. We come at this stage after termination of the kth primal process with $\lambda_{i_k} \leq \beta < 1$. Denote

$$\begin{aligned} t_k &\stackrel{\text{def}}{=} \tfrac{\nu+\rho}{\langle c, u_{i_k} \rangle - f_k} \stackrel{(2.3)}{=} \tfrac{\nu+\rho}{\langle s_k, u_{i_k} \rangle}, \\ \hat{z}_k &\stackrel{\text{def}}{=} \left(\hat{x}_k = x_{t_k}(u_{i_k}), \quad \hat{s}_k = s_{t_k}(u_{i_k}), \quad \hat{y}_k = y_{t_k}(u_{i_k}) \right). \end{aligned} \tag{3.8}$$

Then,

$$\begin{aligned} P^0_\gamma(\hat{z}_k) &\stackrel{(3.3)}{=} \Omega(\hat{z}_k) + \rho \ln \langle \hat{s}_k, \hat{x}_k \rangle - \tfrac{\nu}{1+\gamma}(1 - \ln \nu) \\ &\stackrel{(2.10)}{\leq} P^0_\gamma(u_{i_k}, s_k) + \Omega(\hat{z}_k) + \rho \ln \tfrac{\langle \hat{s}_k, \hat{x}_k \rangle}{\langle s_k, u_{i_k} \rangle} \\ &\stackrel{(3.8)}{=} P^0_\gamma(u_{i_k}, s_k) + \Omega(\hat{z}_k) + \rho \ln \tfrac{t_k \langle \hat{s}_k, \hat{x}_k \rangle}{\nu} + \rho \ln \tfrac{\nu}{\nu+\rho} \\ &\stackrel{(2.14),(2.15)}{\leq} P^0_\gamma(u_{i_k}, s_k) + 2\omega(\beta) + \beta^2 + 2\beta \tfrac{\rho}{\sqrt{\nu}} - \tfrac{\rho^2}{\nu+\rho}. \end{aligned}$$

Thus, we have proved the following inequality

$$\begin{aligned} P^0_\gamma(\hat z_k) &\leq P^0_\gamma(u_{i_k}, s_k) - \Delta_1, \\ \Delta_1 &\stackrel{\mathrm{def}}{=} \tfrac{\rho^2}{\nu+\rho} - 2\beta \tfrac{\rho}{\sqrt{\nu}} - 2\omega(\beta) - \beta^2. \end{aligned} \tag{3.9}$$

Note that we do not need to keep Δ_1 positive.

3. Affine-scaling prediction. In accordance to (4.2) in [5], define *affine-scaling direction* $\Delta z_k = (\Delta x_k, \Delta s_k, \Delta y_k)$ as a unique solution to the following system of linear equations:

$$\begin{aligned} \Delta s_k + \tfrac{1}{t_k} F''(u_{i_k}) \cdot \Delta x_k &= \hat s_k, \\ A \Delta x_k &= 0, \\ \Delta s_k + A^* \Delta y_k &= 0. \end{aligned} \tag{3.10}$$

Note that

$$\langle \hat s_k, \Delta x_k \rangle + \langle \Delta s_k, \hat s_k \rangle = \langle \hat s_k, \hat x_k \rangle, \tag{3.11}$$

and for any $\alpha \in \left[0, \frac{1-\beta}{\beta+\sqrt{\nu}}\right)$ we have (see Theorem 3 in [5])

$$\Omega(\hat z_k - \alpha \Delta z_k) - \Omega(\hat z_k) \leq \alpha \beta^2 \cdot \tfrac{\beta+\sqrt{\nu}}{1-\beta} + \omega\left(\alpha \cdot \tfrac{\beta+\sqrt{\nu}}{1-\beta}\right). \tag{3.12}$$

Therefore,

$$\begin{aligned} P^0_\gamma(\hat z_k - \alpha \Delta z_k) &\stackrel{(3.11),(3.12)}{\leq} P^0_\gamma(\hat z_k) + \rho \ln(1-\alpha) + \alpha\beta^2 \cdot \tfrac{\beta+\sqrt{\nu}}{1-\beta} + \omega\left(\alpha \cdot \tfrac{\beta+\sqrt{\nu}}{1-\beta}\right) \\ &\leq P^0_\gamma(\hat z_k) - \alpha \cdot \left(\rho - \beta^2 \cdot \tfrac{\beta+\sqrt{\nu}}{1-\beta}\right) + \omega\left(\alpha \cdot \tfrac{\beta+\sqrt{\nu}}{1-\beta}\right) \end{aligned}$$

Denoting $\tau = \alpha \cdot \frac{\beta+\sqrt{\nu}}{1-\beta}$ we can see that for finding the optimal step size we need to minimize $-\tau\Delta_2 + \omega(\tau)$, where

$$\Delta_2 = \rho \cdot \tfrac{1-\beta}{\beta+\sqrt{\nu}} - \beta^2. \tag{3.13}$$

Of course, we need to assume that Δ_2 is positive. This gives the optimal $\tau^* = \frac{\Delta_2}{1+\Delta_2}$.

Thus, we conclude that the affine-scaling prediction step ensures at least the following decrease of the normalized penalty potential:

$$\begin{aligned} P^0_\gamma(\hat z_k - \alpha^* \Delta z_k) &\leq P^0_\gamma(\hat z_k) - \omega_*(\Delta_2), \\ \alpha^* &= \tfrac{1-\beta}{\beta+\sqrt{\nu}} \cdot \tfrac{\Delta_2}{1+\Delta_2}. \end{aligned} \tag{3.14}$$

Hence, we can form

$$\begin{aligned} z_{k+1} &= \hat z_k - \alpha_k \Delta z_k, \\ f_{k+1} &= \langle b, y_{k+1} \rangle, \end{aligned} \tag{3.15}$$

where α_k is equal to α^* or to any other positive value ensuring sufficient decrease of the normalized penalty potential.

Since we do not know in advance how many steps of the process (3.6) we are going to perform between the affine scaling prediction steps, it is necessary to establish a sufficiently large uniform lower bound on the decrease of the potential at any iteration. That is

$$\Delta \stackrel{\rm def}{=} \min\{\omega_*(\beta), \Delta_1 + \omega_*(\Delta_2)\}.$$

Taking $\rho = \sqrt{\nu} \geq 1$, we obtain

$$\Delta_1 \stackrel{(3.9)}{=} \tfrac{\nu}{\nu+\sqrt{\nu}} - 2\beta - \beta^2 - 2\omega(\beta) \;\geq\; \tfrac{1}{2} - 2\beta - \beta^2 - 2\omega(\beta),$$

$$\Delta_2 \stackrel{(3.13)}{=} \sqrt{\nu} \cdot \tfrac{1-\beta}{\beta+\sqrt{\nu}} - \beta^2 \;\geq\; \tfrac{1-\beta}{1+\beta} - \beta^2.$$

Clearly, for β small enough we can make Δ_2 positive and ensure

$$\Delta_1 + \omega_*(\Delta_2) \geq \omega_*(\beta) > 0. \tag{3.16}$$

Thus, we have proved the following statement.

Theorem 3 *Let $\rho = \sqrt{\nu}$. For any $\beta \in (0,1)$ satisfying condition (3.16) and ensuring $\Delta_2 > 0$, each step (3.6) or (3.8) with (3.15) ensures a decrease of the normalized penalty potential at least by the value $\omega_*(\beta)$. We get ϵ-solution of problem (2.9) at most in*

$$N \;\leq\; \tfrac{1}{\omega_*(\beta)} \left[\Omega(z_0) + \sqrt{\nu} \ln \tfrac{\langle c, x_0 \rangle - f_0}{\epsilon} \right]. \tag{3.17}$$

iterations.

Proof:
Indeed, we have seen that any step of the potential-reduction method decreases the value of penalty potential by $\omega(\beta)$. Denoting by z_j a current primal-dual point after jth iteration of the scheme, we have

$$\begin{array}{rcl} \sqrt{\nu} \ln \langle s_j, x_j \rangle & \stackrel{(3.3)}{\leq} & P^0_\gamma(z_j) + \tfrac{\nu}{1+\gamma}(1 - \ln \nu) \\[1em] & \leq & P^0_\gamma(z_0) + \tfrac{\nu}{1+\gamma}(1 - \ln \nu) - \omega_*(\beta) \cdot j \\[1em] & \stackrel{(3.3)}{\leq} & \Omega(z_0) + \sqrt{\nu} \ln \langle s_0, x_0 \rangle - \omega_*(\beta) \cdot j. \end{array}$$

□

4 First-order prediction step

Note that in the definition of affine-scaling direction (3.10) we do not use any new information computed at points $\hat{x}_k$ and $\hat{s}_k$. Let us show, that this can be done in a natural way.

Consider a point $u \in \mathrm{rint}\, \mathcal{F}$ with $\lambda_t(u) \leq \beta < 1$. Denote $x = x_t(u)$ and $s = s_t(u)$. Assume that we are able to compute the gradients of the primal and dual barriers $F'(x)$ and $F'_*(s)$. Our goal is to ensure a better decrease the normalized penalty potential $P^0_\gamma(z)$ using this additional information. However, we would like to keep the algebraic complexity of the iteration. This means that we agree to solve only some variants of the linear system (3.10) with different right-hand side $g \in E^*$:

$$\begin{aligned} \delta s + \tfrac{1}{t} F''(u) \cdot \delta x &= g, \\ A \delta x &= 0, \\ \delta s + A^* \delta y &= 0. \end{aligned} \tag{4.1}$$

Denote $B = \frac{1}{t} F''(u)$ and $P = B^{-1/2} A^* [A B^{-1} A^*]^{-1} A B^{-1/2}$. Note that the operator P is a projector: $P = P^2$. It can be easily checked that the solution $\delta z(g) = (\delta x(g), \delta s(g), \delta y(g))$ of system (4.1) is given by the following expressions:

$$\begin{aligned} \delta x(g) &= B^{-1/2}(I - P) B^{-1/2} g, \\ \delta s(g) &= B^{1/2} P B^{-1/2} g, \\ \delta y(g) &= -[A B^{-1} A^*]^{-1} A B^{-1} g. \end{aligned} \tag{4.2}$$

Note that for a feasible displacement $\delta z = (\delta x, \delta s, \delta y)$ we have

$$\begin{aligned} P_\gamma(z - \delta z) - P_\gamma(z) \leq & - \tfrac{\nu + \rho}{\langle s, x \rangle} [\langle s, \delta x \rangle + \langle \delta s, x \rangle] \\ & + F(x - \delta x) - F(x) + F_*(s - \delta s) - F_*(s). \end{aligned}$$

At the same time,

$$F(x - \delta x) - F(x) + \langle F'(x), \delta x \rangle \overset{(1.2)}{\leq} \omega(\|\delta x\|_x) \overset{(2.16)}{\leq} \omega\left(\tfrac{\sqrt{t}}{1-\beta} \langle B \delta x, \delta x \rangle^{1/2} \right),$$

$$F_*(s - \delta s) + F_*(s) - \langle \delta s, F'_*(s) \rangle \overset{(1.2)}{\leq} \omega(\|\delta s\|_s) \overset{(2.16)}{\leq} \omega\left(\tfrac{\sqrt{t}}{1-\beta} \langle \delta s, B^{-1} \delta s \rangle^{1/2} \right).$$

Since $\omega(\xi) + \omega(\tau) \leq \omega([\xi^2 + \tau^2]^{1/2})$ for any $\xi, \tau \geq 0$ with $\xi^2 + \tau^2 < 1$, (see inequality (4.8) in [5]), we conclude that

$$\begin{aligned} P_\gamma(z - \delta z) - P_\gamma(z) \leq & -\langle \tfrac{\nu+\rho}{\langle s, x \rangle} s + F'(x), \delta x \rangle - \langle \delta s, \tfrac{\nu+\rho}{\langle s, x \rangle} x + F'_*(s) \rangle \\ & + \omega\left(\tfrac{\sqrt{t}}{1-\beta} \left[\langle B \delta x, \delta x \rangle^2 + \langle \delta s, B^{-1} \delta s \rangle^2 \right]^{1/2} \right). \end{aligned} \tag{4.3}$$

Hence, it seems reasonable to compute the prediction directions as solutions to the following minimization problem:

$$\begin{aligned} \min_g \{ & -\langle \tfrac{\nu+\rho}{\langle s, x \rangle} s + F'(x), \delta x(g) \rangle - \langle \delta s(g), \tfrac{\nu+\rho}{\langle s, x \rangle} x + F'_*(s) \rangle \\ & + \tfrac{1}{2} \langle B \delta x(g), \delta x(g) \rangle^2 + \tfrac{1}{2} \langle \delta s(g), B^{-1} \delta s(g) \rangle^2 \quad \}. \end{aligned} \tag{4.4}$$

It appears that the solution of this problem can be obtained by solving the linear system (4.1) twice with different right-hand sides.

Theorem 4 *The solution $\delta z_* = (\delta x_*, \delta s_*, \delta y_*)$ of problem (4.4) can be obtained in the following way:*

$$\begin{array}{llcl} 1. & \textit{Compute} \quad g_0 & = & \delta s(F'(x) - B^{-1}F'_*(s)), \\ 2. & \textit{Define} \quad g_* & = & \frac{\nu+\rho}{\langle s,x\rangle} s + F'(x) - g_0, \\ 3. & \textit{Compute} \quad \delta z_* & = & \delta z(g_*). \end{array} \tag{4.5}$$

Moreover, for the optimal predictor direction δz_* we have

$$\begin{aligned} \langle \tfrac{\nu+\rho}{\langle s,x\rangle} s + F'(x), \delta x_* \rangle + \langle \delta s_*, \tfrac{\nu+\rho}{\langle s,x\rangle} x + F'_*(s)\rangle &= \langle B\delta x_*, \delta x_*\rangle^2 + \langle \delta s_*, B^{-1}\delta s_*\rangle^2 \\ &= \langle g_*, B^{-1} g_*\rangle \\ &\geq \tfrac{1}{\langle s,x\rangle} \left[\rho - \tfrac{\beta^2}{1-\beta}(\beta + \sqrt{\nu})\right]^2, \end{aligned} \tag{4.6}$$

where the last inequality is valid for $\rho \geq \frac{\beta^2}{1-\beta}(\beta + \sqrt{\nu})$.

Proof:
First of all, note that the quadratic term in the objective function of problem (4.4) can be written as follows:

$$\begin{aligned} \langle B\delta x(g), \delta x(g)\rangle^2 + \langle \delta s(g), B^{-1}\delta s(g)\rangle^2 &= \langle \delta s(g) + B\delta x(g), B^{-1}(\delta s(g) + B\delta x(g))\rangle \\ &= \langle g, B^{-1} g\rangle. \end{aligned}$$

Denoting now $w = B^{-1/2} g$, $\hat{s} = \frac{\nu+\rho}{\langle s,x\rangle} s$, and $\hat{x} = \frac{\nu+\rho}{\langle s,x\rangle} x$, we get the following minimization problem:

$$\min_w \left\{ -\langle \hat{s} + F'(x), B^{-1/2}(I-P)w\rangle - \langle B^{1/2}Pw, \hat{x} + F'_*(s)\rangle + \tfrac{1}{2}\|w\|^2 \right\}. \tag{4.7}$$

Since $\hat{s} = B\hat{x}$, its solution can be represented as follows:

$$\begin{aligned} w &= (I-P)B^{-1/2}(\hat{s} + F'(x)) + PB^{1/2}(\hat{x} + F'_*(s)) \\ &= B^{-1/2}\left[\hat{s} + B^{1/2}(I-P)B^{-1/2}F'(x) + B^{1/2}PB^{1/2}F'_*(s)\right] \\ &= B^{-1/2}\left[\hat{s} + F'(x) - B^{1/2}PB^{-1/2}(F'(x) - BF'_*(s))\right]. \end{aligned}$$

Hence, the optimal $g = g_* \stackrel{\text{def}}{=} \hat{s} + F'(x) - g_0$ with $g_0 \stackrel{(4.2)}{=} \delta s(F'(x) - BF'_*(s))$.

Further, the first two equalities in (4.6) follow from the form of objective function in minimization (4.7). Let us prove the remaining inequality. Note that

$$\begin{aligned} g_* &= \hat{s} + \tfrac{1}{2}\left[F'(x) + BF'_*(s)\right] + \tfrac{1}{2}\left[F'(x) - BF'_*(s)\right] - B^{1/2}PB^{-1/2}\left[F'(x) - BF'_*(s)\right] \\ &\stackrel{\text{def}}{=} \hat{s} + g_1 + g_2, \\ g_1 &= \tfrac{1}{2}\left[F'(x) + BF'_*(s)\right], \quad g_2 = \tfrac{1}{2}B^{1/2}(I-2P)B^{-1/2}\left[F'(x) - BF'_*(s)\right], \end{aligned}$$

and the vector g_2 is not too big:

$$\begin{aligned}\|g_2\|_u^* &= \tfrac{1}{2}\|[F''(u)]^{1/2}(I-2P)[F''(u)]^{1/2}[F'(x)-BF'_*(s)]\|_u^* \\ &= \tfrac{1}{2}\|(I-2P)[F''(u)]^{-1/2}[F'(x)-BF'_*(s)]\| \\ &= \tfrac{1}{2}\|[F''(u)]^{-1/2}[F'(x)-BF'_*(s)]\| \;=\; \tfrac{1}{2}\|F'(x)-BF'_*(s)\|_u^* \;\overset{(2.13)}{\leq}\; \tfrac{\beta^2}{1-\beta}.\end{aligned}$$

Hence, $\|g_*\|_u^* \geq \|\hat{s}+g_1\|_u^* - \frac{\beta^2}{1-\beta}$. At the same time,

$$\begin{aligned}t\cdot(\|\hat{s}+g_1\|_u^*)^2 &= \tfrac{t(\nu+\rho)^2}{\langle s,x\rangle^2}\langle s,[F''(u)]^{-1}s\rangle + 2t\tfrac{\nu+\rho}{\langle s,x\rangle}\langle g_1,[F''(u)]^{-1}s\rangle + t(\|g_1\|_u^*)^2 \\ &\overset{(2.13)}{=} \tfrac{(\nu+\rho)^2}{\langle s,x\rangle} + 2\tfrac{\nu+\rho}{\langle s,x\rangle}\langle g_1,x\rangle + t(\|g_1\|_u^*)^2.\end{aligned}$$

Since $2\langle g_1,x\rangle = \langle F'(x)+BF'_*(s),x\rangle \overset{(2.13)}{=} \langle F'(x),x\rangle + \langle s,F'_*(s)\rangle \overset{(1.9)}{=} -2\nu$, we get inequality

$$t\cdot(\|\hat{s}+g_1\|_u^*)^2 \;\geq\; \tfrac{\rho^2-\nu^2}{\langle s,x\rangle} + t(\|g_1\|_u^*)^2.$$

Note that

$$\begin{aligned}t\cdot\langle s,x\rangle\cdot(\|g_1\|_u^*)^2 &\overset{(2.13)}{=} \|x\|_u^2\cdot(\|g_1\|_u^*)^2 \;\geq\; \langle g_1,x\rangle^2 \;=\; \tfrac{1}{4}\langle F'(x)+BF'_*(s),x\rangle^2 \\ &\overset{(2.13)}{=} \tfrac{1}{4}[\langle F'(x),x\rangle + \langle s,F'_*(s)\rangle]^2 \;\overset{(1.9)}{=}\; \nu^2.\end{aligned}$$

Therefore,

$$t^{1/2}\cdot\|g_*\|_u^* \;\geq\; \tfrac{\rho}{\langle s,x\rangle^{1/2}} - \tfrac{\beta^2 t^{1/2}}{1-\beta} \;\overset{(2.15)}{\geq}\; \tfrac{1}{\langle s,x\rangle^{1/2}}\left[\rho - \tfrac{\beta^2}{1-\beta}(\beta+\sqrt{\nu})\right].$$

Since $t\cdot(\|g_*\|_u^*)^2 = \langle g_*,B^{-1}g_*\rangle$, we obtain (4.6). □

Note that for self-scaled barriers we have in (4.5) $g_0 = 0$. Let us investigate now the efficiency of the step along direction δz_*. In view of inequality (4.3) and relation (4.6), we have

$$\begin{aligned}P_\gamma(z-\alpha\delta z_*) - P_\gamma(z) &\leq -\alpha\langle g_*,B^{-1}g_*\rangle + \omega\left(\tfrac{\alpha\, t^{1/2}}{1-\beta}\langle g_*,B^{-1}g_*\rangle^{1/2}\right) \\ \left(\tau \overset{\text{def}}{=} \tfrac{\alpha\, t^{1/2}}{1-\beta}\langle g_*,B^{-1}g_*\rangle^{1/2}\right) &= -\tau\cdot\tfrac{1-\beta}{t^{1/2}}\langle g_*,B^{-1}g_*\rangle^{1/2} + \omega(\tau).\end{aligned}$$

Thus, the optimal step is

$$\tau^* = \tfrac{(1-\beta)\langle g_*,B^{-1}g_*\rangle^{1/2}}{t^{1/2}+(1-\beta)\langle g_*,B^{-1}g_*\rangle^{1/2}}, \quad \alpha^* = \tau^*\cdot\tfrac{1-\beta}{t^{1/2}}\langle g_*,B^{-1}g_*\rangle^{-1/2},$$

and we prove the following statement.

Theorem 5 *Assume that* $(1-\beta)\rho \geq \beta^2(\beta+\sqrt{\nu})$*. Then, with the step size* α^**, the decrease of normalized penalty potential along direction* δz_* *can be estimated as folows:*

$$P_\gamma(z) - P_\gamma(z-\alpha^*\delta z_*) \quad \geq \quad \omega_*(\tau^*) \;\geq\; \omega_*\left(\frac{(1-\beta)\rho-\beta^2(\beta+\sqrt{\nu})}{(1-\beta^2)(\beta+\sqrt{\nu})+(1-\beta)\rho}\right). \qquad (4.8)$$

Proof:
Indeed, in view of inequality (4.6) we have

$$\begin{aligned}\tau^* \quad &= \quad \frac{(1-\beta)\langle g_*, B^{-1}g_*\rangle^{1/2}}{t^{1/2}+(1-\beta)\langle g_*, B^{-1}g_*\rangle^{1/2}} \;\geq\; \frac{(1-\beta)\rho-\beta^2(\beta+\sqrt{\nu})}{t^{1/2}\langle s,x\rangle^{1/2}+(1-\beta)\rho-\beta^2(\beta+\sqrt{\nu})}\\ &\overset{(2.15)}{\geq} \quad \frac{(1-\beta)\rho-\beta^2(\beta+\sqrt{\nu})}{(1-\beta^2)(\beta+\sqrt{\nu})+(1-\beta)\rho}.\end{aligned}$$

□

Thus, we can see that the prediction direction δz_* can be used in Step 3 of potential-reduction scheme. Its complexity estimate is similar to (3.17). However, this direction can be better than the standard affine-scaling direction (3.10) since it takes into account the gradients of barrier functions. Of course, the final conclusion about the quality of these directions can be derived only from intensive computational testing.

To conclude, let us shortly discuss reasonable strategies for choosing parameters in the new potential-reduction schemes. The most important parameter is, of course, ρ. From viewpoint of the worst case complexity analysis, we need to choose $\rho = \kappa\cdot\sqrt{\nu}$, where $\kappa \geq 1$ is an absolute constant. Then the impact of all three stages of the scheme is balanced and we can guarantee a constant decrease of the normalized penalty potential at any step of any stage. However, note that it is possible to choose, for example, $\rho = \frac{1}{2}\nu$. Then the primal-dual lifting decreases the potential by $O(\rho)$ (see (3.9)). This means, that for problems with an easy correction phase, we can gain a lot by increasing ρ. Unfortunately, up to now we cannot convert this reasoning in a complexity bound.

References

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

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

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

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

[5] Yu. Nesterov. Towards nonsymmetric conic optimization. CORE Discussion Paper #2006/28, (2006)

[6] Yu. Nesterov. Constructing self-concordant barriers for convex cones. CORE Discussion Paper #2006/30, (2006)

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

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

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

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