# Self-Scaled Cones and Interior-Point Methods in Nonlinear Programming

Yu. E. Nesterov* M. J. Todd†

April 12, 1994

**Abstract:** This paper provides a theoretical foundation for efficient interior-point algorithms for nonlinear programming problems expressed in conic form, when the cone and its associated barrier are *self-scaled.* For such problems we devise long-step and symmetric primal-dual methods. Because of the special properties of these cones and barriers, our algorithms can take steps that go typically a large fraction of the way to the boundary of the feasible region, rather than being confined to a ball of unit radius in the local norm defined by the Hessian of the barrier.

**Key words:** Nonlinear Programming, conical form, interior point algorithms, self-concordant barrier, self-scaled cone, self-scaled barrier, path-following algorithms, potential-reduction algorithms.

**AMS 1980 subject classification.** Primary: 90C05, 90C25, 65Y20.

*CORE, Catholic University of Louvain, Louvain-la-Neuve, Belgium. E-mail: nesterov@core.ucl.ac.be.
Part of this work was done while the author was visiting the Cornell Center for Applied Mathematics, with support from NSF, AFOSR and ONR through NSF Grant DMS-8920550, and from the Center for Applied Mathematics.

†School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853. E-mail: miketodd@cs.cornell.edu.
Research supported in part by NSF, AFOSR and ONR through NSF Grant DMS-8920550.

# 1 Introduction

The aim of this paper is to provide a theoretical foundation for the study of efficient interior-point methods for problems that are extensions of linear programming. Standard form linear programming problems minimize a linear function of a vector of variables subject to linear equality constraints and the requirement that the vector belong to the nonnegative orthant in $n$-dimensional Euclidean space. Here this cone is replaced by a possibly non-polyhedral convex cone. Any convex programming problem can be expressed in this conical form.

Nesterov and Nemirovskii [7] have investigated the essential ingredients necessary to extend several classes of interior-point algorithms for linear programming (inspired by Karmarkar's famous projective-scaling method [5]) to nonlinear settings. The key element is that of a *self-concordant barrier* for the convex feasible region. This is a smooth convex function defined on the interior of the set, tending to $+\infty$ as the boundary is approached, that together with its derivatives satisfies certain Lipschitz continuity properties. The barrier enters directly into functions used in path-following and potential-reduction methods, but, perhaps as importantly, its Hessian at any point defines a local norm whose unit ball, centered at that point, lies completely within the feasible region. Moreover, the Hessian varies in a well-controlled way in the interior of this ball.

This paper is concerned with a special class of convex cones and associated barriers that we term *self-scaled*. While they must satisfy certain apparently restrictive conditions, it seems that this class includes some important instances, for example the cone of positive semidefinite matrices and the second-order cone (to be defined below), as well as the nonnegative orthant in $R^n$. For such cones, the Hessian of the barrier at any interior point maps the cone onto its dual cone, and *vice versa* when we consider the conjugate barrier. In addition, for any pair of points, one in the interior of the original (primal) cone and the other in the interior of the dual cone, there is some point at which the Hessian carries the first into the second. Thus there is a very rich class of scaling transformations, which come from the Hessians evaluated at the points of the cone itself (hence self-scaled).

The consequences of these conditions are quite extensive. For our purposes, the key results are the existence of a symmetric primal-dual scaling and the fact that good approximations of self-scaled barriers and their gradients extend far beyond unit balls defined by the local norm, and in fact are valid up to a constant fraction of the distance to the boundary in any direction. Using these ideas we are able to derive primal long-step potential-reduction and path-following algorithms as well as a symmetric long-step primal-dual potential-reduction method.

The first half of the paper defines self-scaled cones and barriers and then makes a thorough study of their structure. We establish that most relevant properties of the nonnegative orthant extend to this nonlinear setting. Section 2 defines our notions and gives examples. In Section 3, we study the scaling transformations in detail and demonstrate symmetry between the primal and dual cones with respect to the self-scaling property. Section 4 defines several measures of the distance to the boundary of a self-scaled cone and establishes the approximation results, and then in Section 5 we examine the behavior of such barriers on certain two-dimensional cones defined by "orthogonal" directions and derive an important consequence.

The second half of the paper applies these results to the derivation of long-step primal and symmetric primal-dual methods. Section 6 states formally the problems with which we are concerned and our assumptions and then investigates the projections that are used in all the algorithms to obtain search directions. Primal potential-reduction methods like those of Karmarkar [5] and Gonzaga [4] are studied in Section 7, while Section 8 develops a symmetric primal-dual algorithm extending that of Kojima, Mizuno, and Yoshise [6]. Finally, Section 9 gives some improvements possible in a primal path-following method like that of Nesterov and Nemirovskii [7] when the cone is self-scaled. In particular, we give a new long-step update for the barrier parameter and a new step-size rule in Newton's method in this approach; this is based on an extension of the quadratic convergence result of Roos and Vial [9] to the case of self-scaled cones.

All these methods require $O(\nu \ln(1/\epsilon))$ or $O(\sqrt{\nu}\ln(1/\epsilon))$ iterations to generate a feasible solution with objective function with $\epsilon$ of the optimal value, where $\nu$ is a parameter of the cone and barrier corresponding to $n$ for the nonnegative orthant in $R^n$. All are variants of methods already known for the standard linear programming case or for the more general conic case, but we stress the improvements possible because the cone is assumed self-scaled. For example, we indicate why Gonzaga's affine-scaling potential-reduction algorithm [4] might be more efficient when the coefficient used in the potential function is chosen as $2\nu$ rather than $\nu + \sqrt{\nu}$ because long steps can be taken while maintaining a guaranteed reduction in the potential function and the reduction is likely to be much larger.

This paper is quite long and detailed. The reader may wish to omit some of the derivations at a first reading. All of Section 2 should be read, with the possible exception of the verification that the second-order cone is self-scaled. The proofs in Sections 3, 4, and 5 can be omitted initially. The most important results are stated as theorems. All of Section 6 should be read, but the succeeding sections are largely independent of each other and can be read separately.

In what follows we often refer to different statements of [7]. The corresponding references we indicate by an upper-case asterisk. Thus, the reference $T^*$ ($C^*$, $D^*$, $P^*$) 1.1.1 corresponds to the first theorem (corollary, definition, proposition) in the first section of Chapter 1 of [7].

# 2 Definition and examples

Let $K$ be a closed convex cone in a finite-dimensional real vector space $E$ (of dimension at least 1) with dual space $E^*$. We denote the corresponding scalar product by $\langle s, x\rangle$ for $x \in E$, $s \in E^*$. In what follows we assume that the interior of the cone $K$ is nonempty and that $K$ is *pointed* (contains no straight line). Let $F$ be a $\nu$-self-concordant logarithmically homogeneous barrier for cone $K$ (see $D^*$ 2.3.2). Recall that by definition, $F$ is a self-concordant barrier for $K$ (see $D^*$ 2.3.1) which for all $x \in \operatorname{int} K$ and $\tau > 0$ satisfies the identity:

$$F(\tau x) \equiv F(x) - \nu \ln \tau. \tag{2.1}$$

Since $K$ is a pointed cone, $\nu \geq 1$ in view of $C^*$ 2.3.3.

We will often use the following straightforward consequences of (2.1): for all $x \in \operatorname{int} K$ and $\tau > 0$,

$$F'(\tau x) = \frac{1}{\tau} F'(x), \quad F''(\tau x) = \frac{1}{\tau^2} F''(x), \tag{2.2}$$

$$F''(x)x = -F'(x), \quad F'''(x)[x] = -2F''(x), \tag{2.3}$$

$$\langle F'(x), x \rangle = -\nu, \tag{2.4}$$

$$\langle F''(x)x, x \rangle = \nu, \quad \langle F'(x), [F''(x)]^{-1} F'(x) \rangle = \nu \tag{2.5}$$

(see $P^*$ 2.3.4).

Define the cone $K^*$ dual to $K$ as follows:

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

Note that $K^*$ is also a pointed cone with nonempty interior. Let the function $F_*$ on $\operatorname{int} K^*$ be conjugate to $F$, namely:

$$F_*(s) := \max\{-\langle s, x \rangle - F(x) : x \in K\}. \tag{2.6}$$

In accordance with $T^*$ 2.4.4, $F_*$ is a $\nu$-self-concordant logarithmically homogeneous barrier for $K^*$. We will often use the following properties of conjugate self-concordant barriers for dual cones: for any $x \in \operatorname{int} K$ and $s \in \operatorname{int} K^*$,

$$-F'(x) \in \operatorname{int} K^*, \quad -F_*'(s) \in \operatorname{int} K, \tag{2.7}$$

$$F_*(-F'(x)) = \langle F'(x), x \rangle - F(x) = -\nu - F(x) \tag{2.8}$$

(using (2.4)),

$$F(-F_*'(s)) = -\nu - F_*(s), \tag{2.9}$$

$$F_*'(-F'(x)) = -x, \quad F'(-F_*'(s)) = -s, \tag{2.10}$$

$$F''(-F_*'(s)) = [F_*''(s)]^{-1}, \quad F_*''(-F'(x)) = [F''(x)]^{-1}, \tag{2.11}$$

$$F(x) + F_*(s) \geq -\nu + \nu \ln \nu - \nu \ln \langle s, x \rangle, \tag{2.12}$$

and the last inequality is satisfied as an equality if and only if $s = -\alpha F'(x)$ for some $\alpha > 0$ (see $P^*$ 2.4.1).

In this paper we consider cones and barriers of rather special type. As we will see later, the properties of these cones are very close to those of the positive orthant. Let us give our main definition.

**Definition 2.1** *Let $K$ be a pointed cone with nonempty interior. We call this cone* self-scaled *if there exists a $\nu$-self-concordant logarithmically homogeneous barrier $F$ for cone $K$ such that for any $w$ and $x$ from* $\operatorname{int} K$*,*

$$F''(w)x \in \operatorname{int} K^* \tag{2.13}$$

*and*

$$F_*(F''(w)x) = F(x) - 2F(w) - \nu. \tag{2.14}$$

*We call the barrier $F$ a $\nu$-*self-scaled *barrier for cone $K$.*

In what follows we always assume that cone $K$ is self-scaled.

We will study properties of self-scaled cones and barriers in detail in the following sections. Now let us give three important examples of such cones.

1. **Positive orthant**. For

$$K := \{x \in R^n : x^{(i)} \geq 0,\ i = 1 \dots n\}$$

($x^{(i)}$ denotes the $i$th component of $x$), we have

$$K^* = \{s \in R^n : s^{(i)} \geq 0,\ i = 1 \dots n\}.$$

Let us take

$$F(x) := -\sum_{i=1}^{n} \ln x^{(i)}$$

(this barrier is $n$-logarithmically homogeneous, see [7], Example 3, p. 40). Then

$$F'(x)^{(i)} = -\frac{1}{x^{(i)}},\ i = 1 \dots n, \quad F''(x) = \operatorname{diag}\left\{\frac{1}{(x^{(i)})^2},\ i = 1 \dots n\right\},$$

and $\nu = n$. Note that $F_*$ is given by

$$F_*(s) = -\sum_{i=1}^{n} \ln s^{(i)} - n$$

and therefore relations (2.13) and (2.14) are evidently satisfied.

2. **Cone of positive semidefinite matrices**. Let $K$ be the cone of positive semidefinite matrices:

$$K := \{X \in M_n : \langle Xy, y\rangle \geq 0, \forall y \in R^n\},$$

where $M_n$ is the space of symmetric $n \times n$-matrices. Then $M_n^*$ can be identified with $M_n$ and we can use the usual inner product

$$\langle S, X\rangle := \sum_{i=1}^{n}\sum_{j=1}^{n} S^{(i,j)} X^{(i,j)} \equiv \operatorname{Trace} SX,$$

where Trace denotes the trace. Then

$$K^* = \{S \in M_n : \langle Sy, y\rangle \geq 0, \forall y \in R^n\}.$$

Let us take

$$F(X) := -\ln \det X.$$

Note that

$$F'(X) = -X^{-1}, \quad F''(X)H = X^{-1}HX^{-1},\ \forall H \in M_n,$$

and $F$ is a logarithmically homogeneous $\nu$-self-concordant barrier with $\nu = n$ (see [7], $P^*$ 5.4.5). The structure of the inverse Hessian for this barrier is also very simple:

$$[F''(X)]^{-1}H = XHX,\ \forall H \in M_n.$$

The corresponding conjugate barrier is given by

$$F_*(S) = -\ln\det S - n$$

and we see that conditions (2.13) and (2.14) are satisfied.

Note that in some applications the cone of positive semidefinite matrices is represented in a slightly different form (see, for example, paper [2] devoted to the Truss Topology Design Problem). Namely, consider the cone

$$\bar{K} := \mathrm{cl}\{(\tau, x, W) \in R \times R^n \times M_n : \tau > \langle W^{-1}x, x\rangle,\ W \in \mathrm{int}\, K\}.$$

Note that the point $(\tau, x, W)$ lies in the interior of cone $\bar{K}$ if and only if the matrix

$$\bar{X}(\tau, x, W) := \begin{pmatrix} \tau & x^T \\ x & W \end{pmatrix}$$

is positive definite. This implies that cone $\bar{K}$ is self-scaled and the corresponding $(n+1)$-self-scaled barrier is given by

$$\bar{F}(\tau, x, W) = -\ln\det \bar{X}(\tau, x, W) = -\ln(\tau - \langle W^{-1}x, x\rangle) - \ln\det W.$$

3. **Second-order cone**. Let

$$K := \{(\tau, x) \in R^{n+1} : \tau \geq \| x \|\},$$

where $\| \cdot \|$ is the Euclidean norm on $R^n$. In this case

$$K^* = \{(\rho, s) \in R^{n+1} : \rho \geq \| s \|\}.$$

Let us take

$$F(\tau, x) := -\ln(\tau^2 - \| x \|^2).$$

In accordance with [7], $P^*$ 5.4.3, $F$ is a 2-self-concordant logarithmically homogeneous barrier for $K$. Note that

$$F'(\tau, x) = \frac{2}{\tau^2 - \| x \|^2}\begin{pmatrix} -\tau \\ x \end{pmatrix},$$

$$F''(\tau, x) = \frac{2}{\tau^2 - \| x \|^2}\begin{pmatrix} -1 & 0 \\ 0 & I_n \end{pmatrix} + \frac{4}{(\tau^2 - \| x \|^2)^2}\begin{pmatrix} \tau^2 & -\tau x^T \\ -\tau x & x x^T \end{pmatrix}.$$

It can easily be seen that the conjugate barrier is given by

$$F_*(\rho, s) = -\ln(\rho^2 - \| s \|^2) - 2 + 2\ln 2.$$

Therefore for any $(\eta, w)$ and $(\tau, x)$ from $\mathrm{int}\, K$ we have:

$$F''(\eta, w)\begin{pmatrix} \tau \\ x \end{pmatrix} = \begin{pmatrix} \rho \\ s \end{pmatrix},$$

where

$$\rho = \frac{2(\tau(\eta^2 + \| w \|^2) - 2\eta\langle w, x\rangle)}{(\eta^2 - \| w \|^2)^2},$$

$$s = \frac{2((\eta^2 - \| w \|^2)x + 2(\langle w, x\rangle - \tau\eta)w)}{(\eta^2 - \| w \|^2)^2}.$$

This implies that

$$\begin{aligned}
&\frac{1}{4}(\eta^2 - \| w \|^2)^4(\rho^2 - \| s \|^2) \\
&= (\tau(\eta^2 + \| w \|^2) - 2\eta\langle w, x\rangle)^2 - \| (\eta^2 - \| w \|^2)x + 2(\langle w, x\rangle - \tau\eta)w \|^2 \\
&= \tau^2(\eta^2 + \| w \|^2)^2 - (\eta^2 - \| w \|^2)^2 \| x \|^2 - 4\tau^2\eta^2 \| w \|^2 \\
&= (\tau^2 - \| x \|^2)(\eta^2 - \| w \|^2)^2.
\end{aligned}$$

Thus,

$$\rho^2 - \| s \|^2 = \frac{4(\tau^2 - \| x \|^2)}{(\eta^2 - \| w \|^2)^2} > 0.$$

Since

$$\rho \geq \frac{2(\eta^2\tau^{-1} \| x \|^2 + \tau \| w \|^2 - 2\eta\langle w, x\rangle)}{(\eta^2 - \| w \|^2)^2} = \frac{2 \| \eta\tau^{-1/2}x - \tau^{1/2}w \|^2}{(\eta^2 - \| w \|^2)^2} \geq 0,$$

we conclude that $(\rho, s) \in \operatorname{int} K^*$. Moreover,

$$F_*(\rho, s) = -\ln(\rho^2 - \| s \|^2) - 2 + 2\ln 2 = F(\tau, x) - 2F(\eta, w) - 2$$

and thus cone $K$ is self-scaled.

In some applications the second-order cone arises in a "hyperbolic" form:

$$\bar{K} := \{(\tau, \eta, x) \in R \times R \times R^n : \tau\eta \geq \| x \|^2,\ \tau \geq 0,\ \eta \geq 0\}.$$

Note that this cone is also self-scaled since

$$\bar{K} = \{(\tau, \eta, x) \in R \times R \times R^n : \tfrac{1}{2}(\tau + \eta) \geq \sqrt{\tfrac{1}{4}(\tau - \eta)^2 + \| x \|^2}\}$$

(see Theorem 2.1(ii)). The corresponding 2-self-scaled barrier for cone $\bar{K}$ is given by

$$\bar{F}(\tau, \eta, x) = -\ln(\tau\eta - \| x \|^2).$$

In the next sections we will see that self-scaled cones and barriers have many interesting properties. Let us present to conclude this section the combination rules which preserve the self-scaling property.

**Theorem 2.1** *(i) Let $E_1$ and $E_2$ be finite-dimensional linear spaces. If cone $K_i \subseteq E_i$ is self-scaled with $\nu_i$-self-scaled barrier $F_i$, $i = 1, 2$, then the cone*

$$K := K_1 \times K_2 \subseteq E_1 \times E_2$$

*is also self-scaled, with $\nu$-self-scaled barrier given by $F(x_1, x_2) := F_1(x_1) + F_2(x_2)$, where $\nu = \nu_1 + \nu_2$. The self-scaled barrier for the conjugate cone $K^* = K_1^* \times K_2^*$ is the function defined by $F_*(s_1, s_2) := (F_1)_*(s_1) + (F_2)_*(s_2)$ with the same value of the parameter $\nu$.*

(ii) *Let $A$ be an automorphism of $E$. If a cone $K \subseteq E$ is self-scaled with $\nu$-self-scaled barrier $F$, then the cone*

$$\bar{K} := \{x \in E : Ax \in K\}$$

*is also self-scaled, with $\nu$-self-scaled barrier given by $\bar{F}(x) := F(Ax)$. The $\nu$-self-scaled barrier for the dual cone*

$$\bar{K}^* = \{s \in R^n : A^{-*}s \in K^*\}$$

*is given by $\bar{F}_*(s) := F_*(A^{-*}s)$. ($A^*$ denotes the adjoint of $A$, and $A^{-*}$ the inverse of $A^*$.)*

# 3 Self-dual transformations and scaling

This section begins our investigation of the special properties of self-scaled cones and barriers. In particular, we show the existence of a *scaling point*, which relates any point $x$ in $\operatorname{int} K$ and any point $s$ in $\operatorname{int} K^*$. Let us start from the following result.

**Theorem 3.1** (i) *For any $w$ and $x$ from $\operatorname{int} K$ we have:*

$$F_*'(F''(w)x) = [F''(w)]^{-1}F'(x), \tag{3.1}$$

$$F_*''(F''(w)x) = [F''(w)]^{-1}F''(x)[F''(w)]^{-1}. \tag{3.2}$$

(ii) *Let $w_1$ and $w_2$ belong to $\operatorname{int} K$. If there exists $x \in \operatorname{int} K$ such that*

$$F''(w_1)x = F''(w_2)x$$

*then $w_1 = w_2$.*

(iii) *Let us fix any $w \in \operatorname{int} K$. Then*

$$K^* = F''(w)K.$$

**Proof:**
Inequalities (3.1), (3.2) are obtained by differentiating the identity (2.14) with respect to $x$. Let us prove part (ii) of the theorem. For this, we choose any dual bases of $E$ and $E^*$ and represent the Hessians $F''(\cdot)$ as symmetric positive definite matrices with respect to these bases. Let

$$s := F''(w_1)x = F''(w_2)x.$$

Denote $Q := [F''(x)]^{1/2}$. From (3.2) we have:

$$QF_*''(s)Q = (Q[F''(w_1)]^{-1}Q)^2 = (Q[F''(w_2)]^{-1}Q)^2.$$

This implies that $F''(w_1) = F''(w_2)$ since the symmetric positive definite square root of a symmetric positive definite matrix is unique. Denote $G := F''(w_1) = F''(w_2)$. From (2.3) we have:

$$F'(w_1) = -Gw_1, \quad F'(w_2) = -Gw_2.$$

Therefore (since $F$ is convex),

$$0 \leq \langle F'(w_1) - F'(w_2), w_1 - w_2 \rangle = -\langle G(w_1 - w_2), w_1 - w_2 \rangle$$

and we conclude that $w_1 = w_2$ since $G$ is positive definite.

Let us now prove part (iii). Of course, it is enough to prove that

$$\operatorname{int} K^* = F''(w) \operatorname{int} K.$$

For any $\hat{s} \in \operatorname{int} K^*$ let $\hat{x} := [F''(w)]^{-1}\hat{s}$. Then, in view of (2.14), (3.1) and (2.4), for any $v \in \operatorname{int} K$ we have:

$$F_*(\hat{s}) \geq F_*(F''(w)v) + \langle \hat{s} - F''(w)v, F_*'(F''(w)v) \rangle$$

$$= F(v) - 2F(w) - \nu + \langle F'(v), \hat{x} - v \rangle = F(v) - 2F(w) + \langle F'(v), \hat{x} \rangle.$$

Further, let $s = -F'(v)$. Then, by (2.9) and (2.10) we obtain:

$$F_*(\hat{s}) \geq F(-F_*'(s)) - 2F(w) - \langle s, \hat{x} \rangle = -F_*(s) - \nu - 2F(w) - \langle s, \hat{x} \rangle.$$

Thus, we have proved that for any $s \in \operatorname{int} K^*$,

$$F_*(s) + \langle s, \hat{x} \rangle \geq -F_*(\hat{s}) - 2F(w) - \nu.$$

However, the function (of $s$) on the left hand side of this inequality is bounded below if and only if $\hat{x} \in \operatorname{int} K$ (see [7], $T^*$ 2.4.2, $T^*$ 2.4.4). This proves part (iii). □

Let us now prove that Definition 2.1 leads to symmetric relations for the dual cone.

**Proposition 3.1** *Let $K$ be a self-scaled cone with $\nu$-self-scaled barrier $F(x)$. Then, for any $u$ and $s$ from* $\operatorname{int} K^*$, *the point $F_*''(u)s$ belongs to* $\operatorname{int} K$ *and*

$$F(F_*''(u)s) = F_*(s) - 2F_*(u) - \nu. \tag{3.3}$$

**Proof:**

Let us fix $u$, $s \in \operatorname{int} K^*$ and let $w := -F_*'(u)$. Then by part (iii) of Theorem 3.1 there exists $x \in \operatorname{int} K$ such that $s = F''(w)x$. Therefore

$$F_*''(u)s = [F''(w)]^{-1}s = x \in \operatorname{int} K$$

(see (2.11)). Further, since $F_*(s) = F_*(F''(w)x)$, in view of (2.14) we have:

$$F_*(s) = F(x) - 2F(w) - \nu = F(F_*''(u)s) - 2F(-F_*'(u)) - \nu = F(F_*''(u)s) + 2F_*(u) + \nu$$

(see (2.8)). □

From now on we will present only the properties of the primal self-scaled cone and barrier. The corresponding statements for the dual case have an absolutely symmetric form.

Thus, we have proved that any Hessian $F''(x)$ of a self-scaled barrier defines a one-to-one linear mapping from the primal to the dual cone. Let us now prove a converse statement.

**Theorem 3.2** *For each pair $x \in \operatorname{int} K$, $s \in \operatorname{int} K^*$, there exists a unique* scaling point *$z \in \operatorname{int} K$ such that*

$$s = F''(z)x.$$

*Moreover, $F'(x) = F''(z)F'_*(s)$ and $F''(x) = F''(z)F''_*(s)F''(z)$.*

**Proof:**
Consider the function

$$\phi(w) := \langle s, w\rangle - \langle F'(w), x\rangle.$$

In view of (2.6), (2.3), and (2.14), for any $w \in \operatorname{int} K$ we have:

$$\phi(w) = \langle s, w\rangle + \langle F''(w)x, w\rangle \geq \langle s, w\rangle - F(w) - F_*(F''(w)x) = \langle s, w\rangle + F(w) - F(x) + \nu.$$

Note that all level sets of the right hand side of this inequality (as a function of $w$) are closed and bounded. Therefore the function $\phi$ attains its minimum. Its minimizers are exactly the solutions of the equation

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

in $z$. But this equation has a unique solution by part (ii) of Theorem 3.1.

The expressions for the derivatives of the barrier $F$ follow from (3.1) and (3.2). $\square$

Note that, if we set $t := -F'(z)$, then (2.11) implies that the theorem remains true with $x$ and $s$, $K$ and $K^*$, $z$ and $t$, and $F$ and $F_*$ interchanged. Thus the existence of a scaling point is completely symmetric between $K$ and $K^*$.

In most applications, it is quite straightforward to compute $z$ given $x$ and $s$. For example, if $K = R^n_+$, then $z = [\operatorname{diag}(x)]^{1/2}[\operatorname{diag}(s)]^{-1/2}e$, with $e$ the vector of ones in $R^n$. If $K$ is the cone of positive semidefinite matrices, $X \in K$, and $S \in K^* = K$, then $Z = X^{1/2}(X^{1/2}SX^{1/2})^{-1/2}X^{1/2}$. The case of the second-order cone can also be worked out easily if tediously.

**Corollary 3.1** *For any $v, w \in \operatorname{int} K$ there exists $z \in \operatorname{int} K$ such that*

$$F'(v) = -F''(z)w.$$

*Moreover, $F'(w) = -F''(z)v$ and therefore*

$$F'(v) - F'(w) = F''(z)(v - w).$$

**Proof:**
This statement is a straightforward consequence of the above theorem in view of Theorem 3.1 (iii). $\square$

Let us describe now some interesting properties of the derivatives of a self-scaled barrier. For fixed $w \in K$, denote

$$g(x) := -\langle F'(x), w\rangle.$$

**Lemma 3.1** *The function $g$ is convex.*

**Proof:**
Assume first that $w \in \operatorname{int} K$. Let us fix $z \in \operatorname{int} K$. We shall prove that for any $x \in \operatorname{int} K$ the following inequality holds:

$$g(x) = -\langle F'(x), w\rangle \geq -\langle F'(z), w\rangle - \langle F''(z)w, x - z\rangle = -2\langle F'(z), w\rangle - \langle F''(z)w, x\rangle.$$

Note that in the proof of Theorem 3.2 we have already established that the function

$$-\langle F'(x), w\rangle + \langle s, x\rangle$$

(of $x$) attains its minimum for any $s \in \operatorname{int} K^*$. In our case $s = F''(z)w$, which belongs to $\operatorname{int} K^*$ in view of Definition 2.1. Thus, the (unique) minimizer of this function can be found from the equation

$$F''(x)w = F''(z)w;$$

it is $z$ by part (ii) of Theorem 3.1. This establishes the inequality and hence the convexity of $g$ in this case.

The convexity of $g$ defined by a boundary point $w$ follows from continuity reasons. □

**Corollary 3.2** *(i) For any $w \in K$ and $x \in \operatorname{int} K$ the operator $F'''(x)[w]$ is negative semidefinite.*

*(ii) Let $x \in \operatorname{int} K$ and $x + p \in K$ for some $p \in E$. Then*

$$F'''(x)[p] \leq 2F''(x). \tag{3.4}$$

*(This inequality, and similar ones following, is with respect to the cone of positive semi-definite operators; thus it is equivalent to saying that the right hand side minus the left hand side is positive semidefinite.)*

*(iii) Let us fix a point $z \in \operatorname{int} K$. Then the function*

$$h(x) := \langle [F''(z)]^{-1}F'(x), F'(x)\rangle$$

*is strictly convex on $\operatorname{int} K$.*

*(iv) The mapping $\Phi := F' : \operatorname{int} K \to -\operatorname{int} K^*$ is concave on $\operatorname{int} K$ with respect to $K^*$.*

**Proof:**
Part (i) of the corollary is a straightforward consequence of Lemma 3.1. For part (ii), let $w = x + p \in K$. By part (i) and (2.3), for any $v \in E$ we have:

$$\langle F'''(x)[p]v, v\rangle = \langle F'''(x)[w]v, v\rangle - \langle F'''(x)[x]v, v\rangle \leq 2\langle F''(x)v, v\rangle.$$

In order to prove part (iii) we need only compute the Hessian of the function $h$:

$$h''(x) = 2F'''(x)[[F''(z)]^{-1}F'(x)] + 2F''(x)[F''(z)]^{-1}F''(x),$$

and this is positive definite using part (i) since the point $w := -[F''(z)]^{-1}F'(x)$ belongs to $\operatorname{int} K$.

Let us now consider part (iv). In view of the definition of a concave mapping (see, for example, [7], Section 5.1.2), we shall prove that for any $x, w \in \operatorname{int} K$

$$s := \Phi(x) + \Phi'(x)(w - x) - \Phi(w) \in K^*.$$

Indeed, for any $z \in K$ we have:

$$\begin{aligned}
\langle s, z\rangle &= -\int_0^1 \langle [\Phi'(x + \tau(w - x)) - \Phi'(x)](w - x), z\rangle d\tau \\
&= -\int_0^1 \int_0^\tau \langle F'''(x + \theta(w - x))[w - x, w - x], z\rangle d\theta d\tau \\
&= -\int_0^1 \int_0^\tau \langle F'''(x + \theta(w - x))[z](w - x), w - x\rangle d\theta d\tau \geq 0
\end{aligned}$$

in view of (i). This implies that $s \in K^*$. $\square$

# 4 Distance to the boundary

Here we introduce several functions which provide a measure of the distance to the boundary of a convex cone with respect to some interior point $x$. In the general case, the only such measure available is the local norm defined by the Hessian of the barrier function. The new measures here allow us to obtain approximations of $F$ and $F'$ which are valid in a much larger neighborhood of a point than one given by a bound on the local norm.

Let us fix $x \in \operatorname{int} K$ and let $p \in E$. Denote

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

so $0 \leq \sigma_x(p) < \infty$. Thus, $x - \alpha p \in \operatorname{int} K$ for all

$$\alpha \in \left[0, \frac{1}{\sigma_x(p)}\right).$$

(Here $1/\sigma_x(p)$ is interpreted as $+\infty$ if $\sigma_x(p) = 0$.)

It is clear that for any $w \in K$ we have:

$$\sigma_x(x - w) \leq 1, \quad \sigma_x(-w) = 0.$$

Indeed,

$$\sigma_x(-w) = 0 \text{ iff } w \in K.$$

In particular,

$$\sigma_x(x) = 1, \quad \sigma_x(-x) = 0.$$

It is also clear that for any $\lambda > 0$ we have:

$$\sigma_x(\lambda p) = \lambda \sigma_x(p) = \sigma_{x/\lambda}(p).$$

For $p \in E$ and $q \in E^*$ denote

$$\| p \|_x := \langle F''(x)p, p\rangle^{1/2}, \quad \| q \|_x^* := \langle q, [F''(x)]^{-1}q\rangle^{1/2}.$$

We also write, for $s \in \operatorname{int} K^*$,

$$\| p \|_s := \langle [F_*''(s)]^{-1}p, p\rangle^{1/2}, \quad \| q \|_s^* := \langle q, F_*''(s)q\rangle^{1/2}.$$

Using Theorem 3.2, we have the following intriguing symmetry result: for $x \in \operatorname{int} K$, $s \in \operatorname{int} K^*$,

$$\begin{array}{rclcl} \| s \|_x^* & = & \langle s, [F''(x)]^{-1}s\rangle^{1/2} & = & \langle s, [F''(z)]^{-1}[F_*''(s)]^{-1}[F''(z)]^{-1}s\rangle^{1/2} \\ & = & \langle [F_*''(s)]^{-1}x, x\rangle^{1/2} & = & \| x \|_s, \end{array} \tag{4.1}$$

where $z \in \operatorname{int} K$ is such that $F''(z)x = s$.

When $K$ is the nonnegative orthant in $R^n$, $\sigma_x(p)$ is the maximum component of the scaled vector $[\operatorname{diag}(x)]^{-1}p$, and $-\sigma_x(-p)$ is its minimum component, assuming the vector has both nonnegative and nonpositive components; $\| p \|_x$ is the Euclidean norm of this scaled vector. Similarly, if $K$ is the cone of positive semidefinite matrices, then $\sigma_X(P)$ is the largest eigenvalue of the (scaled) matrix $X^{-1/2}PX^{-1/2}$ if this is nonnegative, and $\| P \|_X$ is the Euclidean norm of the vector of its eigenvalues. Finally, in the case of the second-order cone, $\| \cdot \|_{(\tau,x)}$ and $\sigma_{(\tau,x)}(\cdot)$ are easy to compute.

For any $x \in \operatorname{int} K$, $u \in E^*$, and $v$, $w \in E$, we have the Cauchy-Schwartz inequalities

$$\begin{array}{rcl} \langle F''(x)v, w\rangle & \leq & \| v \|_x \| w \|_x, \\ \langle u, w\rangle & \leq & \| u \|_x^* \| w \|_x . \end{array}$$

Applying this with $v = x$ and $w = p$, we find (using (2.3) and (2.5))

$$\langle -F'(x), p\rangle \leq \| x \|_x \| p \|_x = \sqrt{\nu} \| p \|_x . \tag{4.2}$$

Another useful inequality comes from $\langle F''(x)x, x - \alpha p\rangle \geq 0$ as long as $\alpha \in [0, 1/\sigma_x(p)]$, which yields

$$\langle -F'(x), p\rangle \leq \nu \sigma_x(p). \tag{4.3}$$

In view of $T^*$ 2.1.1,

$$\{w : \| w - x \|_x \leq 1\} \subseteq K.$$

Therefore

$$\sigma_x(p) \leq \| p \|_x, \quad \sigma_x(-p) \leq \| p \|_x \,. \tag{4.4}$$

Let us define also the corresponding measure for the dual cone. Let $s \in \operatorname{int} K^*$ and $q \in E^*$. Denote

$$\sigma_s^*(q) = \frac{1}{\sup\{\beta : s - \beta q \in K^*\}}, \quad \sigma_x(q)^* = \sigma^*_{-F'(x)}(q).$$

Let $x \in \operatorname{int} K$, $s \in \operatorname{int} K^*$, $p \in E$, and $q \in E^*$. As a straightforward consequence of the definitions we get the following inequality:

$$\langle q, p\rangle \geq \sigma_s^*(q)\langle s, p\rangle + \sigma_x(p)\langle q, x\rangle - \sigma_x(p)\sigma_s^*(q)\langle s, x\rangle.$$

(For example, if $\sigma_x(p)$ and $\sigma_s^*(q)$ are positive, this follows from the fact that the scalar product of $x - p/\sigma_x(p)$ and $s - q/\sigma_s^*(q)$ is nonnegative.) Since $\sigma_x(-x) = 0$, we obtain the following particular case for $p = -x$:

$$\langle q, x\rangle \leq \sigma_s^*(q)\langle s, x\rangle.$$

In view of (2.4) this implies that

$$\langle q, x\rangle \leq \nu \sigma_x(q)^*$$

for any $q \in E^*$.

The following theorem underlines the main difference between self-scaled and self-concordant barriers. By $T^*$ 1.1.1, for general self-concordant barriers, we can bound the Hessian at a point $z$ in terms of the Hessian at $x$ only when $\|z - x\|_x$ less than one.

**Theorem 4.1** *For any $\alpha \in [0, 1/\sigma_x(p))$ the following inequality holds:*

$$\frac{1}{(1 + \alpha\sigma_x(-p))^2} F''(x) \leq F''(x - \alpha p) \leq \frac{1}{(1 - \alpha\sigma_x(p))^2} F''(x). \tag{4.5}$$

**Proof:**

We first prove the right hand inequality. Let $\bar{p} := -p/\sigma_x(p)$ if $\sigma_x(p) > 0$, $\bar{p} := -p/\sigma$ for arbitrary $\sigma > 0$ otherwise, so that $x + \bar{p} \in K$. Let us fix a direction $v \in E$. Consider the following function:

$$\phi(\beta) := \langle F''(x(\beta))v, v\rangle, \quad x(\beta) := x + \beta\bar{p}, \quad \beta \in [0, 1).$$

Note that

$$\phi'(\beta) = \langle F'''(x(\beta))[\bar{p}]v, v\rangle.$$

Since $x(\beta) + (1 - \beta)\bar{p} \in K$, we have from (3.4):

$$\langle F'''(x(\beta))[\bar{p}]v, v\rangle \leq \frac{2}{1 - \beta}\langle F''(x(\beta))v, v\rangle.$$

This corresponds to the following inequality:

$$\phi'(\beta) \le \frac{2}{1-\beta}\phi(\beta).$$

Introducing $\psi(\beta) := \ln \phi(\beta)$ we can rewrite this inequality as follows:

$$\psi'(\beta) \le \frac{2}{1-\beta}.$$

This implies that

$$\psi(\beta) \le \psi(0) - 2\ln(1-\beta),$$

or, which is the same,

$$\phi(\beta) \le \frac{\phi(0)}{(1-\beta)^2}.$$

If $\sigma_x(p) > 0$, using the definition of $\phi$ and setting $\beta = \alpha\sigma_x(p)$ immediately yields the right hand inequality of (4.5); otherwise, use the definition of $\phi$ and set $\beta = \alpha\sigma$, and then take limits as $\sigma \to 0$ to get the result.

The inequality we have proved shows that $F''(x+w) \le F''(x)$ for any $x \in \operatorname{int} K$, $w \in K$. Hence if $\sigma_x(-p) = 0$, the left hand inequality of (4.5) follows directly by substituting $x - \alpha p$ for $x$ and $\alpha p$ for $w$. Otherwise, we have

$$(x - \alpha p) + (\alpha + 1/\sigma_x(-p))p \in K,$$

which shows that

$$\sigma_{x-\alpha p}(-p) \le \sigma_x(-p)/(1 + \alpha\sigma_x(-p)).$$

Then applying the right hand inequality of (4.5) again with $x$ and $x - \alpha p$ interchanged and the sign of $p$ reversed yields the left hand inequality. □

**Corollary 4.1** *(i) For any $x \in \operatorname{int} K$, $w \in K$ and $\alpha \in [0,1)$ we have:*

$$F''(x+w) \le F''(x) \le F''\left(x - \frac{\alpha}{\sigma_x(w)}w\right).$$

*(ii) For any $x$, $z \in \operatorname{int} K$ the following inequality holds:*

$$F''(x) \le \sigma_x^2(z)F''(z).$$

*(iii) For any $x, w \in \operatorname{int} K$ we have:*

$$\frac{1}{\sigma_x(w)}F''(x) \le F''(z) \le \sigma_w(x)F''(x), \tag{4.6}$$

*where $z \in \operatorname{int} K$ is such that $F'(w) = -F''(z)x$.*

**Proof:**
The left hand inequality in (i) is a direct consequence of the right hand inequality of (4.5) since $\sigma_x(-w) = 0$. The right hand inequality in (i) holds for the same reason: we simply say that for

$$v := x - \frac{\alpha}{\sigma_x(w)} w$$

we have $x - v \in K$.

Further, from the definition of $\sigma_x(z)$ we have:

$$w := \sigma_x(z)x - z \in K.$$

Therefore

$$F''(\sigma_x(z)x) = F''(z + w) \leq F''(z).$$

It remains to use relation (2.2). Part (ii) is proved.

We now prove part (iii) of the corollary. (Note that Corollary 3.1 demonstrates the existence of the required $z$.) Again, we assume that the Hessians $F''(\cdot)$ are represented as symmetric positive definite matrices with respect to dual bases of $E$ and $E^*$. Let

$$Q_w = [F''(x)]^{-1/2} F''(w) [F''(x)]^{-1/2}, \quad Q_z = [F''(x)]^{-1/2} F''(z) [F''(x)]^{-1/2}.$$

Note that in view of Theorem 3.1(i) and (2.11) we have:

$$F''(w) = F''(z)[F''(x)]^{-1} F''(z)$$

and therefore $Q_w = Q_z^2$. Using (ii) we obtain

$$\frac{1}{\sigma_x^2(w)} I \leq Q_w \leq \sigma_w^2(x) I.$$

Hence

$$\frac{1}{\sigma_x(w)} I \leq Q_z \leq \sigma_w(x) I,$$

which yields the desired conclusion. □

Let us now prove two important inequalities. The first one provides bounds on the variation of a self-scaled barrier which we will use in the analysis of both our primal (potential-reduction and path-following) and our primal-dual algorithms. Note that our result allows good approximations even for long steps, i.e., those taking us a fixed fraction of the way to the boundary of the cone $K$.

**Theorem 4.2** *Let $x \in \operatorname{int} K$ and $p \in E$ be such that $\sigma_x(p) > 0$. Then for any $\alpha \in [0, 1/\sigma_x(p))$ we have:*

$$F(x - \alpha p) \leq F(x) - \alpha \langle F'(x), p \rangle + \frac{\| p \|_x^2}{\sigma_x^2(p)} \left( -\alpha \sigma_x(p) - \ln(1 - \alpha \sigma_x(p)) \right). \tag{4.7}$$

**Proof:**
Let $\bar{p} := -p/\sigma_x(p)$, and note that $x + \bar{p} \in K$. Consider the following function:

$$\theta(\beta) := F(x(\beta)), \quad x(\beta) := x + \beta\bar{p}, \quad \beta \in [0, 1).$$

Note that in view of Theorem 4.1

$$\theta''(\beta) \leq \frac{\theta''(0)}{(1-\beta)^2}, \quad \beta \in [0, 1).$$

Therefore

$$\theta(\beta) - \theta(0) = \int_0^\beta \theta'(\lambda) d\lambda = \theta'(0)\beta + \int_0^\beta \int_0^\lambda \theta''(\tau) d\tau d\lambda$$

$$\leq \theta'(0)\beta + \theta''(0) \int_0^\beta \int_0^\lambda \frac{d\tau d\lambda}{(1-\tau)^2} = \theta'(0)\beta + \theta''(0)(-\beta - \ln(1-\beta)).$$

It remains to use the concrete form of the function $\theta$. □

We can follow the argument used in the proof of Theorem 4.1 in the case that $\sigma_x(p) = 0$. Thus for $p \in K$ we get the following form of (4.7):

$$F(x + \alpha p) \leq F(x) + \alpha\langle F'(x)p\rangle + \frac{\alpha^2}{2} \parallel p \parallel_x^2 . \tag{4.8}$$

We will often use the inequality (4.7) combined with the following simple result.

**Proposition 4.1** *(i) For any $\tau \geq 0$ the function*

$$\tau - \ln(1 + \tau)$$

*is monotonically increasing.*

*(ii) For any $\tau > -1$ the function*

$$\frac{\tau - \ln(1 + \tau)}{\tau^2}$$

*is monotonically decreasing.*

*(iii) For any $\tau < 1$ the function*

$$\frac{-\tau - \ln(1 - \tau)}{\tau^2}$$

*is monotonically increasing.*

Our second inequality estimates the second-order term in the Taylor approximation of the gradient. It is very useful in the analysis of Newton's method and path-following strategies. Let

$$| p |_x = \max\{\sigma_x(p), \sigma_x(-p)\}. \tag{4.9}$$

Note that $| p |_x \leq \| p \|_x$. In the case of $K = R^n_+$ this measure is exactly the infinity norm of the vector $[\mathrm{diag}\,(x)]^{-1}p$.

**Theorem 4.3** *For any $x, w \in \mathrm{int}\, K$ the following inequality holds:*

$$\| F'(w) - F'(x) - F''(x)(w - x) \|^*_w \leq | w - x |_x \cdot \| w - x \|_x .$$

**Proof:**
Let us choose $z \in \mathrm{int}\, K$ such that $F'(w) = -F''(z)x$ (which is possible by Corollary 3.1). Then in view of Corollary 3.1,

$$F'(w) - F'(x) - F''(x)(w - x) = (F''(z) - F''(x))(w - x).$$

Let us estimate the matrix

$$Q := (F''(z) - F''(x))[F''(w)]^{-1}(F''(z) - F''(x))$$

from above using the matrix $F''(x)$. (We assume that the Hessians $F''(\cdot)$ are represented as symmetric positive definite matrices with respect to dual bases of $E$ and $E^*$.) Note that by Theorem 3.2 and (2.11)

$$F''(w) = F''(z)[F''(x)]^{-1}F''(z).$$

Therefore

$$\begin{aligned} Q &= F''(z)[F''(w)]^{-1}F''(z) - F''(x)[F''(w)]^{-1}F''(z) \\ &\quad -F''(z)[F''(w)]^{-1}F''(x) + F''(x)[F''(w)]^{-1}F''(x) \\ &= F''(x) - 2F''(x)[F''(z)]^{-1}F''(x) + F''(x)[F''(z)]^{-1}F''(x)[F''(z)]^{-1}F''(x) \\ &= [F''(x)]^{1/2}(Q_z - I)^2[F''(x)]^{1/2}, \end{aligned}$$

where

$$Q_z = [F''(x)]^{1/2}[F''(z)]^{-1}[F''(x)]^{1/2}.$$

Note that in view of Corollary 4.1 (iii)

$$\frac{1}{\sigma_w(x)} I \leq Q_z \leq \sigma_x(w) I.$$

Therefore

$$(Q_z - I)^2 \leq \left(\max\left\{\sigma_x(w) - 1, 1 - \frac{1}{\sigma_w(x)}\right\}\right)^2 I$$

(note that at least one of the arguments of the maximum in the right hand side of the inequality is nonnegative since $\sigma_x(w)\sigma_w(x) \geq 1$, see (4.6)).

Further, denote $p := w - x$. Then for any $\alpha > 0$ such that $x - \alpha p \in K$ we have:

$$x - \alpha p = (1 + \alpha)x - \alpha w \in K.$$

Therefore $\sigma_x(w) \leq 1 + \sigma_x(p)$. Similarly, for any $\alpha \geq 1$ such that $x + \alpha p \in K$ we have:

$$x + \alpha p = \alpha w - (\alpha - 1)x \in K,$$

and so $\sigma_w(x) \leq 1/(1 - \sigma_x(-p))$.

Thus, we have proved that

$$(Q_z - I)^2 \leq | p |_x^2 I.$$

By definition of $Q$ this implies that

$$(\| F'(w) - F'(x) - F''(x)(w - x) \|_w^*)^2 = \langle Q(w - x), w - x \rangle$$

$$= \langle (Q_z - I)^2 [F''(x)]^{1/2}(w - x), [F''(x)]^{1/2}(w - x) \rangle \leq | p |_x^2 \cdot \| p \|_x^2 .$$

□

# 5 Behavior of a self-scaled barrier along orthogonal feasible directions

Let us fix $v \in \partial K$, $v \neq 0$. Then there exists a point $s \in \partial K^*$ such that $\langle s, v \rangle = 0$. Let us fix also $z \in \operatorname{int} K$. By part (iii) of Theorem 3.1 $K^* = F''(z)K$. Therefore there exists a point $w \in \partial K$ such that $s = F''(z)w$. Thus, we can find a point $w \in \partial K$ such that

$$\langle F''(z)v, w \rangle = 0.$$

We will call direction $w$ *orthogonal* to $v$ with respect to $z$. The main result of this section is that the self-scaled barrier $F(\cdot)$ is in a sense separable on the two-dimensional plane defined by the directions $v$ and $w$ that passes through the point $z$. Besides its intrinsic interest, we use this to derive a key inequality to be used in the analysis of our primal-dual algorithm.

Let

$$D := \{x : x = z + \alpha v + \beta w, \ \alpha \geq 0, \ \beta \geq 0\}.$$

**Lemma 5.1** *For any $x \in D$ the following identities hold:*

$$\langle F''(x)v, w \rangle = 0, \tag{5.1}$$

$$\langle F'''(x)[w]v, v \rangle = 0, \quad \langle F'''(x)[v]w, w \rangle = 0. \tag{5.2}$$

**Proof:**
Let

$$\psi(\alpha) := -\langle F'(z+\alpha v), w\rangle, \quad \alpha \geq 0.$$

In view of Lemma 3.1 this function is convex. Therefore for any $\alpha \geq 0$ we have:

$$\psi'(\alpha) \geq \psi'(0) = -\langle F''(z)v, w\rangle = 0.$$

However

$$\psi'(\alpha) = -\langle F''(z+\alpha v)v, w\rangle \leq 0$$

from part (iii) of Theorem 3.1, and we conclude that

$$\langle F''(z+\alpha v)v, w\rangle = 0$$

for any $\alpha \geq 0$.

Now we can replace in this reasoning the point $z$ by $\bar{z} = z + \alpha v$ for some fixed $\alpha \geq 0$ and prove that $\langle F''(\bar{z}+\beta w)v, w\rangle = 0$ for any $\beta \geq 0$. Thus, identity (5.1) is proved.

The relations (5.2) are just the right derivatives of identity (5.1) with respect to $\alpha$ and $\beta$ respectively. □

**Theorem 5.1** *For any $x = z + \alpha v + \beta w$ with $\alpha$, $\beta \geq 0$ we have:*

$$F(x) = F(z+\alpha v) + F(z+\beta w) - F(z), \tag{5.3}$$

$$\langle F'(x), v\rangle = \langle F'(z+\alpha v), v\rangle, \quad \langle F'(x), w\rangle = \langle F'(z+\beta w), w\rangle, \tag{5.4}$$

$$\langle F''(x)v, v\rangle = \langle F''(z+\alpha v)v, v\rangle, \quad \langle F''(x)v, w\rangle = 0, \quad \langle F''(x)w, w\rangle = \langle F''(z+\beta w)w, w\rangle. \tag{5.5}$$

**Proof:**
Relations (5.4) and (5.5) can be obtained by differentiating (5.3) with respect to $\alpha$ and $\beta$. Therefore we need only prove (5.3).

In view of (5.1) we have:

$$F(x) - F(z+\alpha v) = \int_0^\beta \langle F'(z+\alpha v+\tau w), w\rangle d\tau$$

$$= \int_0^\beta \left( \langle F'(z+\tau w), w\rangle + \int_0^\alpha \langle F''(z+\rho v+\tau w)v, w\rangle d\rho \right) d\tau$$

$$= \int_0^\beta \langle F'(z+\tau w), w\rangle d\tau = F(z+\beta w) - F(z).$$

□

Let us now prove a very useful inequality.

**Theorem 5.2** *For any $x \in \operatorname{int} K$ and $s \in \operatorname{int} K^*$ we have:*

$$\langle F'(x), F'_*(s) \rangle \geq \frac{\nu(\nu-1)}{\langle s, x \rangle} + \frac{3}{4}\sigma_x^2(z), \tag{5.6}$$

*where the point $z \in \operatorname{int} K$ is such that $s = F''(z)x$.*

(This theorem is an extension of Lemma 2.5 of Kojima, Mizuno, and Yoshise [6], and is used similarly to prove constant decrease in a primal-dual potential function.)

**Proof:**

Note that from Theorem 3.2 $F'(x) = F''(z)F'_*(s)$, and therefore the inequality (5.6) can be rewritten as follows:

$$(\| F'(x) \|_z^*)^2 \geq \frac{\nu(\nu-1)}{\| x \|_z^2} + \frac{3}{4}\sigma_x^2(z). \tag{5.7}$$

Further, from the definition of the coefficient $\sigma_x(z)$ we have:

$$v := \sigma_x(z)x - z \in \partial K.$$

Therefore $x = (z+v)/\sigma_x(z)$ and (5.7) is equivalent to the following inequality:

$$(\| F'(z+v) \|_z^*)^2 \geq \frac{\nu(\nu-1)}{\| z+v \|_z^2} + \frac{3}{4} \tag{5.8}$$

(using (2.2)). We want to prove that (5.8) is true for any $v \in \partial K$. In fact, we prove the stronger inequality

$$(\| F'(z+v) \|_z^*)^2 \geq \frac{(\nu-1)^2}{\| z+v \|_z^2 - 1} + 1. \tag{5.9}$$

To show that (5.9) implies (5.8), let

$$\rho := \| z+v \|_z^2;$$

we then want to show that

$$\frac{(\nu-1)^2}{\rho - 1} + 1 \geq \frac{\nu(\nu-1)}{\rho} + \frac{3}{4}.$$

Note that $\rho \geq \| z+v \|_{z+v}^2 = \nu$ by part (i) of Corollary 4.1. The inequality above is equivalent to

$$\frac{1}{4}\rho(\rho-1) - (\nu-1)(\rho-\nu) \geq 0,$$

which holds for any $\rho \geq \nu \geq 1$ since the left hand side equals $\frac{1}{4}(\rho - 1 + [(\rho-\nu) - (\nu-1)]^2)$.

In order to establish (5.9) for any $v \in \partial K$, let $w$ be a direction orthogonal to $v$ with respect to $z$. Define

$$\mu = \frac{\rho - \nu}{\nu - 1}, \quad \tau = -\frac{\mu}{\langle F'(z+v), w \rangle} = -\frac{\mu}{\langle F'(z), w \rangle}$$

(see (5.4) with $\alpha = 1$, $\beta = 0$). Then in view of (2.4) and (5.4)

$$\langle F'(z+v), z+v+\tau w\rangle = -\nu + \tau\langle F'(z), w\rangle = -\nu - \mu.$$

Therefore

$$(\nu+\mu)^2 \le \left(\| F'(z+v) \|_z^*\right)^2 \| z+v+\tau w \|_z^2 .$$

Since $w$ and $v$ are orthogonal with respect to $z$, we have:

$$\| z+v+\tau w \|_z^2 = \rho + 2\tau\langle F''(z)z, w\rangle + \tau^2 \| w \|_z^2$$

$$= \rho + 2\mu + \mu^2 \frac{\langle F''(z)w, w\rangle}{\langle F'(z), w\rangle^2} \le \rho + 2\mu + \mu^2$$

(we have used $C^*$ 2.3.1). Note that

$$\nu + \mu = \frac{\rho - 2\nu + \nu^2}{\nu - 1},$$

$$\rho + 2\mu + \mu^2 = \rho - 1 + (\mu+1)^2 = (\rho - 1)\left(1 + \frac{\rho - 1}{(\nu-1)^2}\right)$$

$$= \frac{(\rho-1)(\rho - 2\nu + \nu^2)}{(\nu-1)^2}.$$

Combining these inequalities,

$$\left(\| F'(z+v) \|_z^*\right)^2 \ge \frac{(\nu+\mu)^2}{\rho + 2\mu + \mu^2} = 1 + \frac{(\nu-1)^2}{\rho - 1},$$

which is exactly (5.9). □

# 6 Problems, assumptions, and projections

In this section, we state the problems we are concerned with, list our assumptions, and then prove some results about (oblique) projections and relate them to Euclidean orthogonal projections. These projections will frequently arise as search directions in algorithms described in succeeding sections.

The problem we address is:

$$\begin{array}{llrcl} \text{(P)} & \min & \langle c, x\rangle & & \\ & \text{s.t.} & Ax & = & b, \\ & & x & \in & K, \end{array}$$

where we assume that

$$\begin{array}{l} A \text{ is a surjective linear operator from } E \text{ to} \\ \text{another finite-dimensional real vector space } Y^*. \end{array} \qquad (6.1)$$

Here, $b \in Y^*$ and $c \in E^*$. The assumption that $A$ is surjective is without loss of generality (else replace $Y^*$ with its range).

The dual to problem (P) is then (see [3]):

$$\begin{array}{llrcrcl} \text{(D)} & \max & \langle b, y \rangle & & & & \\ & \text{s.t.} & A^* y & + & s & = & c, \\ & & & & s & \in & K^*, \end{array}$$

where $A^*$ denotes the adjoint of $A$, mapping $Y$ to $E^*$, and $y \in Y$.

We make the following assumptions about (P) and (D):

$$S^0(P) := \{x \in \operatorname{int} K : Ax = b\} \text{ is nonempty,} \tag{6.2}$$

and

$$S^0(D) := \{(y, s) \in Y \times \operatorname{int} K^* : A^* y + s = c\} \text{ is nonempty.} \tag{6.3}$$

These assumptions imply (see [7], $T^*$ 4.2.1) that both (P) and (D) have optimal solutions and that their optimal values are equal, and that the sets of optimal solutions of both problems are bounded (see [8]). Also, it is easy to see that, if $x$ and $(y, s)$ are feasible in (P) and (D) respectively, then

$$\langle c, x \rangle - \langle b, y \rangle = \langle s, x \rangle.$$

This quantity is the (nonnegative) *duality gap.*

For primal-dual algorithms, we will assume we have available a point $x_0 \in S^0(P)$ and a point $(y_0, s_0) \in S^0(D)$ from which to start the iterations; primal algorithms only need to have $x_0 \in S^0(P)$ and a lower bound $\zeta_0$ on the optimal value $\zeta^*$ of (P) and (D).

Both kinds of algorithm obtain search directions by computing solutions to linear systems of the form

$$\begin{array}{rcl} Ap(u) & = & 0, \\ A^* y(u) + F''(z) p(u) & = & u, \end{array} \tag{6.4}$$

where $z$ is a fixed point in $\operatorname{int} K$ and $u$ a point of $E^*$. We seek $y(u)$ in $Y$ and $p(u)$ in $E$.

**Proposition 6.1** *Under the assumptions above, there is a unique solution to (6.4).*

We call the solution $p(u)$ the *projection* of $u$ into the kernel of $A$ with respect to the positive definite operator $F''(z)$.

**Proof:**

It is enough to show that the linear operator defined by the left hand side of (6.4) is injective, since it maps the finite-dimensional vector space $Y \times E$ into its dual, which has the same dimension. Thus we need to show that any solution $(y(0), p(0))$ with $u = 0$ is uniquely zero. But taking the scalar product of the second set of equations with $p(0)$ we obtain $\langle F''(z) p(0), p(0) \rangle = 0$, whence $p(0)$ must be zero since $F''(z)$ is positive definite. Then the second set of equations yields $A^* y(0) = 0$, which implies $y(0) = 0$ since $A^*$ is injective. Alternatively, note that if we represent the linear transformation defined by the left hand side of (6.4) with respect to dual bases of $E$ and $E^*$ and $Y$ and $Y^*$, we get a symmetric indefinite matrix of an invertible form that is standard in interior-point methods (the matrix representing the surjective mapping $A$ has full row rank). $\square$

Next we prove a simple proposition generalizing the fact that the square of the norm of the projection of a vector is equal to the scalar product of the projection with the vector.

**Proposition 6.2** *If $p(u)$ is a projection of $u$ with respect to the positive definite operator $F''(z)$, then*

$$\| p(u) \|_z^2 = \langle u, p(u) \rangle. \tag{6.5}$$

**Proof:**
From the definition, $F''(z)p(u) = u - A^* y(u)$. Taking the scalar product of both sides with $p(u)$ gives the result. $\square$

Applying the Cauchy-Schwartz inequality to the right hand side of (6.5), we have $\| p(u) \|_z \leq \| u \|_z^*$. More generally, we have

**Proposition 6.3** *For any $y \in Y$ the following inequality holds:*

$$\| p(u) \|_z \leq \| u - A^* y \|_z^* . \tag{6.6}$$

*This inequality becomes an equality only when $y$ is an exact minimizer of the right hand side.*

**Proof:**
Let us consider the following convex quadratic optimization problem:

$$\min\{-\langle u, p \rangle + \tfrac{1}{2}\langle F''(z)p, p \rangle : Ap = 0\}. \tag{6.7}$$

Let $\mu^*$ be the optimal value and $p^*$ be the optimal solution of this problem. Note that $p^*$ can be found from the following system of linear equations:

$$F''(z)p^* + A^* y = u, \quad Ap^* = 0,$$

where $y \in Y$. Therefore we conclude that $p^* = p(u)$ and

$$\mu^* = -\langle u, p^* \rangle + \tfrac{1}{2}\langle F''(z)p^*, p^* \rangle = -\tfrac{1}{2} \| p(u) \|_z^2$$

in view of Proposition 6.2.

Let us write the Lagrangean for the problem (6.7):

$$\mathcal{L}(p, y) = -\langle u, p \rangle + \tfrac{1}{2}\langle F''(z)p, p \rangle + \langle Ap, y \rangle.$$

Therefore the problem dual to (6.7) is:

$$\max\{-\tfrac{1}{2}\langle u - A^* y, [F''(z)]^{-1}(u - A^* y) \rangle : y \in Y\}.$$

Thus, we conclude that

$$-\tfrac{1}{2} \parallel p(u) \parallel_z^2 = \mu^* \geq -\tfrac{1}{2}(\parallel u - A^*y \parallel_z^*)^2$$

for any $y \in Y$. $\square$

We now relate the projections above to Euclidean orthogonal projections. Note that we can rewrite (6.4) as

$$\begin{array}{rcrcrcl} & & & & Ap(u) & = & 0, \\ -A^*y(u) & + & s(u) & & & = & 0, \\ & & s(u) & + & F''(z)p(u) & = & u. \end{array} \tag{6.8}$$

Thus $p(u)$ is in the kernel of $A$, $s(u)$ in the range of $A^*$, and the sum of $s(u)$ and a transformation of $p(u)$ equals the given vector $u$.

Let us choose dual bases for the spaces $E$ and $E^*$, and let $H$ be the symmetric positive definite matrix representing the transformation $F''(z)$ with respect to these bases, of order $n := \dim(E)$. Let $V := R^n$ with the standard coordinate basis, so that $V^*$ can be identified with $V$ and the basis is then self-dual. $V$ is then a Euclidean space with the usual inner product and norm.

Let $J$ be the mapping from $E$ to $V$ represented by the symmetric positive definite matrix $H^{1/2}$. Then $J^* : V^* = V \to E^*$ is also represented by this matrix, and thus $J^*J : E \to E^*$ is represented by the matrix $H$, and thus coincides with $F''(z)$. We have therefore factored the operator $F''(z) : E \to E^*$ into the product of two mappings from $E$ to $V$ and from $V$ to $E^*$. (This factorization corresponds to a Cholesky factorization of the matrix $H$.) Then (6.8) can be rewritten as

$$\begin{array}{rcrcrcl} & & & & (AJ^{-1})(Jp(u)) & = & 0, \\ -(AJ^{-1})^*y(u) & + & J^{-*}s(u) & & & = & 0, \\ & & J^{-*}s(u) & + & Jp(u) & = & J^{-*}u, \end{array} \tag{6.9}$$

which says that the vectors $Jp(u)$ and $J^{-*}s(u)$, both in the Euclidean space $V$, lie respectively in the kernel of $AJ^{-1}$ and the range of its adjoint, two complementary orthogonal subspaces of $V$, and have as their sum the vector $J^{-*}u$ in $V$. They are therefore the orthogonal projections of this vector into the appropriate subspaces.

(Note that, when $K$ is the nonnegative orthant in $E := R^n$, $x$ and $s$ in $\operatorname{int} K$, and $z \in \operatorname{int} K$ so that $F''(z)x = s$, then $F''(z)$ is represented with respect to the standard basis as $\operatorname{diag}(s)[\operatorname{diag}(x)]^{-1}$ and thus $J$ by the matrix $[\operatorname{diag}(s)]^{1/2}[\operatorname{diag}(x)]^{-1/2}$, which is the usual *scaling matrix* used in primal-dual methods.)

Finally, let us discuss the arithmetical complexity of computing solutions to (6.8), either to obtain just $p(u)$ (for primal methods) or to get $(p(u), y(u), s(u))$ (for primal-dual methods). We have seen that we can alternatively solve (6.9), which is a Euclidean projection or least-squares problem, for which many methods are available. One simple scheme for solving (6.8) is as follows. Usually the Hessians of self-scaled barriers have a rather simple form and it is not too difficult to invert them. Thus we first compute $v := [F''(z)]^{-1}u$, thence $y(u) = (A[F''(z)]^{-1}A^*)^{-1}Av$, and so obtain $s(u) = A^*y(u)$ and $p(u) = [F''(z)]^{-1}(u - s(u))$. If all linear mappings are represented by matrices and $Y$ is $m$-dimensional, this requires the solution of a linear system with an $m \times m$ symmetric positive definite matrix and some matrix multiplications.

# 7 Primal potential-reduction methods

In this section, we discuss extensions of two primal potential reduction methods for problem (P), where assumptions (6.1)–(6.3) hold. (Such an extension of Karmarkar's method [5] is already given in [7], Section 4.3. But here we allow "long steps," i.e., steps that typically go a large fraction of the way to the boundary of $K$, whereas the theory in [7] only applies to "short steps," i.e., those which go only a fraction of the way to the boundary of a ball defined by the local norm $\|\cdot\|_x$.) We assume that we have available an initial point $x_0 \in S^0(P)$ and a lower bound $\zeta_0$ on the optimal value of (P). We also assume that *the objective function is not constant on the feasible region of* (P), so that it is strictly greater than the optimal value at any point in $S^0(P)$.

For both algorithms, convergence is based on forcing a decrease in the primal *potential function*, parametrized by $\zeta \leq \zeta^*$, defined by

$$\Phi(x;\zeta) := \mu \ln(\langle c, x\rangle - \zeta) + F(x). \tag{7.1}$$

for various values of $\mu$. Our last assumption above implies that this is well-defined at all points in $S^0(P)$. (In fact, this assumption is made only for convenience. If it fails, we may apply the algorithms exactly as stated, except that if we ever generate a feasible solution $x$ and a lower bound $\zeta$ with $\langle c, x\rangle = \zeta$, we terminate with an indication that $x$ is exactly optimal. In any case, the same complexity bounds hold.) The algorithms of this section are guaranteed to generate a feasible solution $x$ whose objective function value is within $\epsilon$ of the optimal value in $O(\nu \ln(1/\epsilon))$ iterations.

## 7.1 Karmarkar's method

Here we assume further that the constraints $Ax = b$ can be written as $Bx = 0, \langle d, x\rangle = 1$, with $d \in K^*$. (This can always be achieved by replacing $E$ by $E \times R$, $x$ by $(x, \tau)$, $K$ by $K \times R_+$, and the constraints by $Ax - b\tau = 0, \quad \tau = 1$.) Let $\tilde{Y}^*$ be the range of $B$, so that $Y^* = \tilde{Y}^* \times R$, and let us write elements of $Y$ as $(y, \zeta) \in \tilde{Y} \times R$. In this case, our problem takes the form

$$\begin{array}{rrcl} (\mathrm{P}) & \min & \langle c, x\rangle & \\ & \text{s.t.} & Bx & = & 0, \\ & & \langle d, x\rangle & = & 1, \\ & & x & \in & K, \end{array}$$

and its dual can be written as:

$$\begin{array}{llcccccl} \text{(D)} & \max & & & \zeta & & & \\ & \text{s.t.} & B^* y & + & d\zeta & + & s & = & c, \\ & & & & & & s & \in & K^*, \end{array}$$

where $B^*$ denotes the adjoint of $B$ and $y \in \tilde{Y}$. Under the assumptions we have made, this has an optimal solution $(y^*, \zeta^*, s^*)$, and $\zeta^*$ equals the optimal value of (P).

We use the potential function $\Phi$ of (7.1) with $\mu := \nu$, but extend it to $\{x \in \operatorname{int} K : Bx = 0\}$, by defining

$$\Phi(x; \zeta) := \nu \ln\langle c - \zeta d, x\rangle + F(x).$$

Note that this function is homogeneous (of degree 0) in $x$, since $F$ is $\nu$-logarithmically homogeneous. Our first result shows that decreasing this potential function sufficiently yields an approximately optimal solution to (P). From our assumptions,

$$\bar{K} := \{x \in K : Bx = 0, \langle c, x\rangle \le \gamma_0 := \max(\langle c, x_0\rangle, 0) + 1, \langle d, x\rangle \le 1\}$$

is bounded. The barrier $F$ is bounded below on this set: let $\underline{F}$ denote its minimum there.

**Theorem 7.1** *Let $x \in S^0(P)$ and $\zeta \in [\zeta_0, \zeta^*]$ be such that $\Phi(x;\zeta) \le \Phi(x_0;\zeta_0) - \Delta$. Then, as long as*

$$\Delta \ge \nu \ln(\langle c, x_0\rangle - \zeta_0) + F(x_0) - \underline{F},$$

*we have*

$$\langle c, x\rangle - \zeta \le (\langle c, x_0\rangle - \zeta_0) \cdot \exp\left(\frac{F(x_0) - \underline{F}}{\nu}\right) \cdot \exp\left(-\frac{\Delta}{\nu}\right). \tag{7.2}$$

**Proof:**
Let $\lambda := \max(1, \langle c, x\rangle/\gamma_0)$ and $\bar{x} := x/\lambda$. Then if $x \in \bar{K}$, $\lambda = 1$ and $\bar{x} = x$, while if not, $\bar{x} \in \bar{K}$ with $\langle c, \bar{x}\rangle = \gamma_0$. (We will show that the first case holds for $\Delta$ sufficiently large.) In either case, since $\Phi$ is homogeneous, our hypothesis implies that

$$\langle c - \zeta d, \bar{x}\rangle \le (\langle c, x_0\rangle - \zeta_0) \cdot \exp\left(\frac{F(x_0) - F(\bar{x})}{\nu}\right) \cdot \exp\left(-\frac{\Delta}{\nu}\right),$$

and since $\bar{x} \in \bar{K}$ and thus $F(\bar{x}) \ge \underline{F}$, we deduce

$$\langle c - \zeta d, \bar{x}\rangle \le (\langle c, x_0\rangle - \zeta_0) \cdot \exp\left(\frac{F(x_0) - \underline{F}}{\nu}\right) \cdot \exp\left(-\frac{\Delta}{\nu}\right).$$

If $x = \bar{x}$, this immediately yields the conclusion of the theorem. Suppose not. Then $\lambda > 1$ and

$$\langle c - \zeta d, \bar{x}\rangle = \gamma_0 - \zeta/\lambda \ge 1 + \langle c, x_0\rangle/\lambda - \zeta/\lambda \ge 1,$$

and we deduce that

$$1 \le (\langle c, x_0\rangle - \zeta_0) \cdot \exp\left(\frac{F(x_0) - \underline{F}}{\nu}\right) \cdot \exp\left(\frac{-\Delta}{\nu}\right).$$

But this contradicts our hypothesis on the size of $\Delta$. $\square$

From the theorem, it suffices to show that we can reduce $\Phi$ by a constant at each iteration (possibly adjusting the lower bound $\zeta$), and then we will have an $O(\nu \ln(1/\epsilon))$-iteration algorithm to obtain a feasible solution with objective value within $\epsilon$ of the optimal value of (P). (Here the $O(\cdot)$ notation hides additive constants that depend on the initial solution $x_0$ and the initial lower bound $\zeta_0$.) We now show how this can be achieved.

Suppose at any iteration we have $\hat{x} \in S^0(P)$ and $\hat{\zeta} \leq \zeta^*$. We then solve the two linear systems

$$\begin{aligned} Bp(u) &= 0, \\ \langle F'(\hat{x}), p(u) \rangle &= 0, \\ B^* y(u) + F'(\hat{x})\mu(u) + F''(\hat{x})p(u) &= u, \end{aligned} \tag{7.3}$$

for the two right hand sides $u = c$ and $u = d$, to get $p(c)$ and $p(d)$, etc. By Proposition 6.1 these have unique solutions. Note that we have

$$\langle c, \hat{x} \rangle = \langle B^* y(c) + F'(\hat{x})\mu(c) + F''(\hat{x})p(c), \hat{x} \rangle = \mu(c)\langle F'(\hat{x}), \hat{x} \rangle - \langle F'(\hat{x}), p(c) \rangle = -\nu\mu(c)$$

so that $\mu(c) = -\langle c, \hat{x} \rangle / \nu$, and similarly $\mu(d) = -\langle d, \hat{x} \rangle / \nu = -1/\nu$. Our search direction will be of the form $p := p(c) - \zeta p(d)$ for some $\zeta$. We may be able to choose $\zeta = \hat{\zeta}$, but sometimes it is necessary to first update our lower bound.

Lower bounds will be derived from feasible solutions to (D). In order to try to find these, let us define, for $\zeta \leq \zeta^*$,

$$\tilde{x}(\zeta) := \left( \frac{\langle c, \hat{x} \rangle - \zeta}{\nu} \right) \hat{x} + (p(c) - \zeta p(d)) = -(\mu(c) - \zeta\mu(d))\hat{x} + (p(c) - \zeta p(d)) \tag{7.4}$$

and

$$\tilde{s}(\zeta) := F''(\hat{x})\tilde{x}(\zeta) = (\mu(c) - \zeta\mu(d))F'(\hat{x}) + F''(\hat{x})(p(c) - \zeta p(d)).$$

We then have

$$B^*(y_c - \zeta y_d) + \tilde{s}(\zeta) = c - \zeta d$$

so that, as long as $\tilde{s}(\zeta) \in K^*$, $\zeta$ is a lower bound on the optimal value of (P) since it is the value of a feasible solution to (D).

There are now two cases. If $\tilde{x}(\hat{\zeta}) \notin \operatorname{int} K$, we set $\zeta^+ := \hat{\zeta}$, choose the search direction as

$$p := p(c) - \zeta^+ p(d), \tag{7.5}$$

and note that by (7.4) $\nu/(\langle c, \hat{x} \rangle - \zeta^+) \geq 1/\sigma_{\hat{x}}(-p)$, so that $\langle c, \hat{x} \rangle - \zeta^+ \leq \nu\sigma_{\hat{x}}(-p)$. On the other hand, if $\tilde{x}(\hat{\zeta}) \in \operatorname{int} K$, then note that $\tilde{x}(\zeta) = \tilde{x}(\hat{\zeta}) - (\zeta - \hat{\zeta})\tilde{p}$, where $\tilde{p} := p(d) + \hat{x}/\nu$. In this case, we update $\hat{\zeta}$ to

$$\zeta^+ := \hat{\zeta} + 1/\sigma_{\tilde{x}(\hat{\zeta})}(\tilde{p}). \tag{7.6}$$

Then set the search direction again to $p$ given by (7.5), and note that $\tilde{x}(\zeta^+) \in \partial K$, so that $\nu/(\langle c, \hat{x} \rangle - \zeta^+) = 1/\sigma_{\hat{x}}(-p)$ and $\langle c, \hat{x} \rangle - \zeta^+ = \nu\sigma_{\hat{x}}(-p)$. Also, $\tilde{s}(\zeta^+) = F''(\hat{x})\tilde{x}(\zeta^+) \in K^*$ by part (iii) of Theorem 3.1, which implies that $\zeta^+$ is a valid lower bound. We therefore have

**Lemma 7.1** *After possibly updating the lower bound as shown above, we have $\zeta^+ \leq \zeta^*$ and, with $p$ given by (7.5),*

$$\langle c, \hat{x}\rangle - \zeta^+ \leq \nu \sigma_{\hat{x}}(-p). \tag{7.7}$$

□

This lemma generalizes a result of Anstreicher [1] for the case of standard linear programming. Note that it implies (if $\hat{x}$ is not optimal) that $\sigma_{\hat{x}}(-p) > 0$.

We need one more observation before we prove the main theorem of this subsection. Applying Proposition 6.2 with $p$ given by (7.5), we have

$$\| p \|_{\hat{x}}^2 = \langle c - \zeta^+ d, p\rangle. \tag{7.8}$$

We now see how these results and Theorem 4.2 enable us to establish a fixed decrease in potential. For simplicity, let $p_+ := \sigma_{\hat{x}}(p)$ and $p_- := \sigma_{\hat{x}}(-p)$.

**Theorem 7.2** *Let the (possibly updated) lower bound $\zeta^+$ and the search direction $p$ be as above. Then, for a suitable value of $\alpha$, $x^+ := \hat{x} - \alpha p$ lies in $\{x \in \operatorname{int} K : Bx = 0\}$ and satisfies*

$$\Phi(x^+; \zeta^+) \leq \Phi(\hat{x}; \hat{\zeta}) - \frac{\| p \|_{\hat{x}}^2}{| p |_x^2}(1 - \ln 2) \leq \Phi(\hat{x}; \hat{\zeta}) - (1 - \ln 2). \tag{7.9}$$

(Note that $x^+$ may not be feasible, because it may not satisfy the constraint $\langle d, x\rangle = 1$; but because $\Phi$ is homogeneous, we may take as our next iterate $\hat{x}^+ := x^+/\langle d, x^+\rangle$ and achieve the same reduction in the potential function.)

**Proof:**

Let $\Delta\Phi(\alpha) := \Phi(\hat{x} - \alpha p; \zeta^+) - \Phi(\hat{x}; \zeta^+)$. Since $\zeta^+ \geq \hat{\zeta}$ and thus $\Phi(\hat{x}; \zeta^+) \leq \Phi(\hat{x}; \hat{\zeta})$, it suffices to prove that $\Delta\Phi(\alpha)$ is suitably small for some positive $\alpha \leq 1/p_+$.

Let $\bar{c} := c - \zeta^+ d$. Then, from Theorem 4.2,

$$\begin{aligned} \Delta\Phi(\alpha) &= \nu \ln\left(1 - \alpha\tfrac{\langle \bar{c}, p\rangle}{\langle \bar{c}, \hat{x}\rangle}\right) + F(\hat{x} - \alpha p) - F(\hat{x}) \\ &= \nu \ln\left(1 - \alpha\tfrac{\|p\|_{\hat{x}}^2}{\langle \bar{c}, \hat{x}\rangle}\right) + F(\hat{x} - \alpha p) - F(\hat{x}) \\ &\leq \nu \ln\left(1 - \alpha\tfrac{\|p\|_{\hat{x}}^2}{\nu p_-}\right) + F(\hat{x} - \alpha p) - F(\hat{x}) \\ &\leq -\alpha\tfrac{\|p\|_{\hat{x}}^2}{p_-} + \tfrac{\|p\|_{\hat{x}}^2}{p_+^2}[-\alpha p_+ - \ln(1 - \alpha p_+)] \\ &\leq -\alpha \| p \|_{\hat{x}}^2 \left(\tfrac{1}{p_-} + \tfrac{1}{p_+}\right) - \tfrac{\|p\|_{\hat{x}}^2}{p_+^2} \ln(1 - \alpha p_+). \end{aligned} \tag{7.10}$$

The right hand side is minimized by choosing

$$\alpha = \frac{1}{p_+ + p_-} < \frac{1}{p_+},$$

and then $x^+$ lies in $\operatorname{int} K$ (and hence in $\{x \in \operatorname{int} K : Bx = 0\}$) and we have

$$\Delta\Phi(\alpha) \leq -\frac{\| p \|_{\hat{x}}^2}{p_+ p_-} + \frac{\| p \|_{\hat{x}}^2}{p_+^2} \ln\left(1 + \frac{p_+}{p_-}\right)$$

$$= -\frac{\| p \|_{\hat{x}}^2}{\max\{p_+, p_-\}^2}\left[\frac{(\max\{p_+, p_-\})^2}{p_+p_-} - \frac{(\max\{p_+, p_-\})^2}{p_+^2}\ln\left(1 + \frac{p_+}{p_-}\right)\right].$$

Let us write the expression in brackets as

$$\frac{\max\{1, \tau^2\}}{\tau} - \frac{\max\{1, \tau^2\}}{\tau^2}\ln(1+\tau) =: g(\tau),$$

where $\tau := p_+/p_- > 0$. Then we see that for $\tau \geq 1$, in view of Propsition 4.1(i), the function $g(\tau)$ is monotonically increasing, so that $g(\tau) \geq g(1) = 1 - \ln 2 > 0$, while for $\tau \leq 1$, in view of Propsition 4.1(ii), the function $g(\tau)$ is monotonically decreasing, so that $g(\tau) \geq g(1) = 1 - \ln 2 > 0$. Hence

$$\Delta\Phi(\alpha) \leq -\frac{\| p \|_{\hat{x}}^2}{(\max\{p_+, p_-\})^2}(1 - \ln 2) = -\frac{\| p \|_{\hat{x}}^2}{| p |_{\hat{x}}^2}(1 - \ln 2) \leq -(1 - \ln 2).$$

This completes the proof. For future reference, we note that if $p_-$ is replaced in (7.10) by any $\lambda \in (0, \| p \|_{\hat{x}}]$, the proof above can be applied to give

$$\Delta\Phi(\alpha) \leq -\frac{\| p \|_{\hat{x}}^2}{(\max\{p_+, \lambda\})^2}(1 - \ln 2) \leq -(1 - \ln 2).$$

□

The algorithm is now clear. Given $\hat{x} \in S^0(P)$ and $\hat{\zeta} \leq \zeta^*$, first try to update the lower bound as described above Lemma 7.1 to get $\zeta^+$, and hence define $p$ from (7.5). Now choose the step size $\alpha \in [0, 1/\sigma_{\hat{x}}(p))$ so that (7.9) holds for $x^+ := \hat{x} - \alpha p$ and set $\hat{x}^+ := x^+/\langle d, x^+\rangle$. Now $\hat{x}^+$ and $\zeta^+$ replace $\hat{x}$ and $\hat{\zeta}$ and the iteration is finished. Starting with $x_0$ and $\zeta_0$, we continue these iterations until $\langle c, \hat{x}\rangle - \hat{\zeta} \leq \epsilon$, which, by Theorems 7.1 and 7.2, occurs within $O(\nu \ln(1/\epsilon))$ iterations.

## 7.2 An affine potential-reduction method

Now we discuss a method that solves (P) without assuming that the constraints have a special form. This algorithm extends that of Gonzaga in [4]. We use the potential function $\Phi$ of (7.1), but here with $\mu > \nu$. We now let

$$\bar{K} := \{x \in K : Ax = b\tau, 0 \leq \tau \leq 1, \langle c, x\rangle \leq \gamma_0 := \max(\langle c, x_0\rangle, 0) + 1\}$$

which is again bounded by our assumptions. The barrier $F$ is bounded below on this set, and we let $\underline{F}$ denote its minimum there. As before, we first show that decreasing the potential function sufficiently is enough to give an approximately optimal solution to (P).

**Theorem 7.3** *Let $x \in S^0(P)$ and $\zeta \in [\zeta_0, \zeta^*]$ be such that $\Phi(x; \zeta) \leq \Phi(x_0; \zeta_0) - \Delta$. Then, as long as*

$$\Delta \geq \mu \ln(\langle c, x_0\rangle - \zeta_0) + F(x_0) - \underline{F},$$

*we have*

$$\langle c, x\rangle - \zeta \leq (\langle c, x_0\rangle - \zeta_0) \cdot \exp\left(\frac{F(x_0) - \underline{F}}{\mu}\right) \cdot \exp\left(-\frac{\Delta}{\mu}\right). \tag{7.11}$$

**Proof:**
We follow the argument of Theorem 7.1, letting $\lambda := \max(1, \langle c, x \rangle / \gamma_0)$ and $\bar{x} := x/\lambda$. Since $\mu > \nu$, $\Phi$ is not homogeneous, but we have instead $\Phi(\bar{x}; \zeta/\lambda) = \Phi(x;\zeta) - (\mu - \nu) \ln \lambda \leq \Phi(x; \zeta)$. The rest of the proof is exactly as before. □

Hence we wish to decrease $\Phi$ by a constant at each iteration (possibly adjusting the lower bound $\zeta$) in order to have an $O(\mu \ln(1/\epsilon))$-iteration algorithm to obtain a feasible solution with objective value within $\epsilon$ of the optimal value of (P). As in the previous subsection, we show how this can be achieved from any $\hat{x} \in S^0(P)$ and $\hat{\zeta} \leq \zeta^*$.

First we solve the two linear systems

$$\begin{aligned} Ap(u) &= 0, \\ A^* y(u) + F''(\hat{x}) p(u) &= u, \end{aligned} \tag{7.12}$$

for $u = c$ and $u = d := F'(\hat{x})$. Noting that

$$\Phi'(\hat{x}; \zeta) = \frac{\mu}{\langle c, \hat{x} \rangle - \zeta} c + F'(\hat{x}), \tag{7.13}$$

we see that

$$p := p(c) + \lambda p(d), \tag{7.14}$$

for

$$\lambda = \lambda(\zeta) := \frac{\langle c, \hat{x} \rangle - \zeta}{\mu}, \tag{7.15}$$

is $\lambda$ times the projection of the gradient of $\Phi(\cdot; \zeta)$ into the kernel of $A$. Our search direction will have the form of $p$ above where $\zeta$ is a possibly updated lower bound on the optimal value $\zeta^*$ of (P). (Note that the search direction used in the previous subsection can also be viewed as a scaled projection of this gradient, there into the kernel of $B$ intersected with the kernel of $F'(\hat{x})$.)

We first consider the lower bound update. Note that, for any $\lambda \in R$,

$$A^*(y(c) + \lambda y(d)) + F''(\hat{x})(p(c) + \lambda p(d)) - \lambda F'(\hat{x}) = c.$$

We therefore define

$$\tilde{x}(\lambda) := \lambda \hat{x} + p(c) + \lambda p(d) \tag{7.16}$$

and

$$\tilde{s}(\lambda) := F''(\hat{x}) \tilde{x}(\lambda) = F''(\hat{x})(p(c) + \lambda p(d)) - \lambda F'(\hat{x}).$$

Thus, as long as $\tilde{s}(\lambda) \in K^*$, $\langle c, \hat{x} \rangle - \langle \tilde{s}(\lambda), \hat{x} \rangle$ is a lower bound on the optimal value of (P) since it is the value of a feasible solution to (D). The gap

$$\begin{aligned} \langle \tilde{s}(\lambda), \hat{x} \rangle &= \langle -F'(\hat{x}), p(c) \rangle + \lambda((\| F'(\hat{x}) \|_{\hat{x}}^*)^2 - \langle F'(\hat{x}), p(d) \rangle) \\ &= \langle -F'(\hat{x}), p(c) \rangle + \lambda((\| F'(\hat{x}) \|_{\hat{x}}^*)^2 - \| p(d) \|_{\hat{x}}^2) \end{aligned}$$

is an increasing function of $\lambda$, using Propositions 6.2 and 6.3. Thus we choose $\hat{\lambda}$ as the minimum $\lambda$ such that $\tilde{s}(\lambda) \in K^*$ (or $+\infty$ if there is no such $\lambda$), and then set

$$\zeta^+ := \max(\hat{\zeta}, \langle c, \hat{x} \rangle - \langle \tilde{s}(\hat{\lambda}), \hat{x} \rangle), \tag{7.17}$$

where the second argument of the "max" is taken to be $-\infty$ if $\hat{\lambda} = \infty$. (Note that we could use a similar update in our extension of Karmarkar's algorithm, updating $\hat{\zeta}$ whenever an improvement is possible using some $\tilde{s}(\zeta) \in K^*$.)

With this update, we set

$$\lambda^+ := \lambda(\zeta^+), \quad p := p(c) + \lambda^+ p(d).$$

**Lemma 7.2** *If $\mu \geq \nu + \sqrt{\nu}$, then $\lambda^+ \leq \| p \|_{\hat{x}}$, while if $\mu \geq 2\nu$, then $\lambda^+ \leq | p |_{\hat{x}}$.*

Recall that $| p |_{\hat{x}}$ is defined as $\max\{\sigma_{\hat{x}}(p), \sigma_{\hat{x}}(-p)\}$ (see (4.9)).

**Proof:**

Suppose $\lambda^+ > \| p \|_{\hat{x}}$ or $\lambda^+ > | p |_{\hat{x}}$, so that in either case $\lambda^+ > \sigma_{\hat{x}}(-p)$. Then $\lambda^+ \hat{x} + p \in K$, so $\tilde{s}(\lambda^+) \in K^*$. Now $\zeta^+$ is defined as the best lower bound that can be deduced from an $s$ of this form, so we have

$$\begin{aligned} \langle c, \hat{x} \rangle - \zeta^+ &\leq \langle \tilde{s}(\lambda^+), \hat{x} \rangle = \langle F''(\hat{x})(\lambda^+ \hat{x} + p), \hat{x} \rangle \\ &= \lambda^+ \langle F''(\hat{x})\hat{x}, \hat{x} \rangle + \langle F''(\hat{x})p, \hat{x} \rangle = \lambda^+ \nu + \langle -F'(\hat{x}), p \rangle. \end{aligned}$$

If $\mu \geq \nu + \sqrt{\nu}$, we use (4.2), so that $\lambda^+ > \| p \|_{\hat{x}}$ implies

$$\langle c, \hat{x} \rangle - \zeta^+ \leq \lambda^+ \nu + \sqrt{\nu} \, \| p \|_{\hat{x}} < (\nu + \sqrt{\nu})\lambda^+ \leq \mu \lambda^+,$$

contradicting $\lambda^+ = (\langle c, \hat{x} \rangle - \zeta^+)/\mu$.

If $\mu \geq 2\nu$, we use (4.3), so that $\lambda^+ > | p |_{\hat{x}} \geq \sigma_{\hat{x}}(p)$ implies

$$\langle c, \hat{x} \rangle - \zeta^+ \leq \lambda^+ \nu + \nu \sigma_{\hat{x}}(p) < 2\nu \lambda^+ \leq \mu \lambda^+,$$

again contradicting the definition of $\lambda^+$. □

We can now show that a constant decrease in potential is possible. As above, let us write $p_+ := \sigma_{\hat{x}}(p)$ and $p_- := \sigma_{\hat{x}}(-p)$.

**Theorem 7.4** *Let the (possibly updated) lower bound $\zeta^+$ and the search direction $p$ be as above.*

*(a) If $\mu \geq \nu + \sqrt{\nu}$, a suitable value of $\alpha$ can be chosen so that $x^+ := \hat{x} - \alpha p$ lies in $S^0(P)$ and satisfies*

$$\Phi(x^+; \zeta^+) \leq \Phi(\hat{x}; \hat{\zeta}) - \frac{\| p \|_{\hat{x}}^2}{(\max\{p_+, \| p \|_{\hat{x}}\})^2}(1 - \ln 2) = \Phi(\hat{x}; \hat{\zeta}) - (1 - \ln 2). \tag{7.18}$$

*(b) If $\mu \geq 2\nu$, a suitable value of $\alpha$ can be chosen so that $x^+ := \hat{x} - \alpha p$ lies in $S^0(P)$ and satisfies*

$$\begin{aligned} \Phi(x^+;\zeta^+) &\leq \Phi(\hat{x};\hat{\zeta}) - \tfrac{\|p\|_{\hat{x}}^2}{(\max\{p_+,|p|_{\hat{x}}\})^2}(1-\ln 2) = \Phi(\hat{x};\hat{\zeta}) - \tfrac{\|p\|_{\hat{x}}^2}{|p|_{\hat{x}}^2}(1-\ln 2) \\ &\leq \Phi(\hat{x};\hat{\zeta}) - (1-\ln 2). \end{aligned} \tag{7.19}$$

**Proof:**
Note first that, since $Ap = 0$, $x - \alpha p$ lies in $S^0(P)$ as long as $\alpha \in (0, 1/p_+)$. Now, as in the analysis of Karmarkar's method, let $\Delta\Phi(\alpha) := \Phi(\hat{x} - \alpha p; \zeta^+) - \Phi(\hat{x};\zeta^+)$. Since $\zeta^+ \geq \hat{\zeta}$ and thus $\Phi(\hat{x};\zeta^+) \leq \Phi(\hat{x};\hat{\zeta})$, it suffices to prove that $\Delta\Phi(\alpha)$ is suitably less than 0.

We find

$$\begin{aligned} \Delta\Phi(\alpha) &= \mu \ln\left(1 - \alpha\tfrac{\langle c,p\rangle}{\langle c,\hat{x}\rangle - \zeta^+}\right) + F(\hat{x} - \alpha p) - F(\hat{x}) \\ &\leq -\tfrac{\alpha}{\lambda^+}\langle c,p\rangle - \alpha\langle F'(x), p\rangle + \tfrac{\|p\|_{\hat{x}}^2}{p_+^2}[-\alpha p_+ - \ln(1-\alpha p_+)] \\ &= -\tfrac{\alpha}{\lambda^+}\langle c + \lambda^+ F'(x), p\rangle + \tfrac{\|p\|_{\hat{x}}^2}{p_+^2}[-\alpha p_+ - \ln(1-\alpha p_+)] \\ &= -\tfrac{\alpha}{\lambda^+} \| p \|_{\hat{x}}^2 + \tfrac{\|p\|_{\hat{x}}^2}{p_+^2}[-\alpha p_+ - \ln(1-\alpha p_+)] \\ &= -\alpha \| p \|_{\hat{x}}^2 \left(\tfrac{1}{\lambda^+} + \tfrac{1}{p_+}\right) - \tfrac{\|p\|_{\hat{x}}^2}{p_+^2}\ln(1-\alpha p_+). \end{aligned} \tag{7.20}$$

If $\mu \geq \nu + \sqrt{\nu}$, Lemma 7.2 yields $\lambda^+ \leq \| p \|_{\hat{x}}$, and we can substitute this in (7.20) to get an upper bound on $\Delta\Phi(\alpha)$. As in the proof of Theorem 7.2, this bound is minimized by taking

$$\alpha = \frac{1}{p_+ + \| p \|_{\hat{x}}} < \frac{1}{p_+},$$

and then we have

$$\Delta\Phi(\alpha) \leq -\frac{\| p \|_{\hat{x}}^2}{(\max\{p_+, \| p \|_{\hat{x}}\})^2}(1 - \ln 2)$$

using the proof of Theorem 7.2, which gives (7.18).

On the other hand, if $\mu \geq 2\nu$, Lemma 7.2 yields $\lambda^+ \leq | p |_{\hat{x}}$, and substituting this in (7.20) yields a different upper bound on $\Delta\Phi(\alpha)$. Following the proof of Theorem 7.2, we minimize this bound by taking

$$\alpha = \frac{1}{p_+ + | p |_{\hat{x}}} < \frac{1}{p_+},$$

and then we have

$$\Delta\Phi(\alpha) \leq -\frac{\| p \|_{\hat{x}}^2}{(\max\{p_+, | p |_{\hat{x}}\})^2}(1 - \ln 2),$$

which gives (7.19) as desired. □

Of course, this argument also naturally yields an $O(\mu \ln(1/\epsilon))$-iteration algorithm to obtain a feasible solution with objective value within $\epsilon$ of the optimal value of (P). This is the same complexity bound as before as long as we choose $\mu = O(\nu)$.

## 7.3 Remarks

In the previous subsection, we made a distinction between the cases with $\mu \geq \nu + \sqrt{\nu}$ and $\mu \geq 2\nu$ which might have seemed irrelevant. However, in the first case, we were only able to show a decrease in $\Phi$ for $\alpha = 1/(p_+ + \parallel p \parallel_{\hat{x}}) \leq 1/ \parallel p \parallel_{\hat{x}}$, i.e., only for a step within a unit ball in the local norm, and the (bound on the) decrease was only $1 - \ln 2$ (see (7.18)). In the second case, we could choose $\alpha = 1/(p_+ + \mid p \mid_{\hat{x}})$, which is typically a reasonable fraction of the way to the boundary of $K$ (for example, half way if $p_- = p_+$), and the (bound on the) decrease was then $[\parallel p \parallel_{\hat{x}}^2 / \mid p \mid_{\hat{x}}^2](1 - \ln 2)$ (see (7.19)), typically much larger. For Karmarkar's algorithm, even with $\mu = \nu$, a large step and a typically large decrease in $\Phi$ are possible (see (7.9)).

Another interesting observation concerns the derivation of lower bounds. Given a current iterate $\hat{x}$ and search direction $p$, the new iterate is of the form $\hat{x} - \alpha p$ for some positive $\alpha$. However, if the lower bound is updated, it corresponds to a dual feasible solution with $s$ equal to a positive multiple of $F''(\hat{x})(\hat{x} + \alpha p)$ for some positive $\alpha$, i.e., to a search in the opposite direction (see (7.4),(7.16)).

# 8 Joint scaling primal-dual interior-point method

In this section we describe an algorithm that works symmetrically in the primal and dual spaces. At each iteration it strives to decrease the symmetric primal-dual potential function given by

$$\phi(s, x) = (\nu + \rho) \ln\langle s, x\rangle + F(x) + F_*(s),$$

where $\rho$ is a constant at least equal to $\sqrt{\nu}$, defined for $x \in S^0(P)$ and $(y, s) \in S^0(D)$. Recall that $\langle s, x\rangle$ is the duality gap corresponding to feasible $x$ and $(y, s)$. We again assume (6.1)–(6.3), and now we need to have available a point $x_0 \in S^0(P)$ and a point $(y_0, s_0) \in S^0(D)$.

As in the previous section, a suitable reduction in the potential function guarantees that we have near-optimal solutions, here to both (P) and (D). Indeed, by [7], $P^*$ 4.5.1, we have

**Theorem 8.1** *Let $x \in S^0(P)$, $(y, s) \in S^0(D)$ be such that $\phi(x, s) \leq \phi(x_0, s_0) - \Delta$. Then*

$$\langle s, x\rangle \leq \langle s_0, x_0\rangle \cdot R(x_0, s_0) \cdot \exp\left(-\frac{\Delta}{\rho}\right),$$

*where*

$$R(x_0, s_0) := \exp\{(\rho \ln\langle s_0, x_0\rangle + F(x_0) + F_*(s_0) - \nu \ln(\nu) + \nu)/\rho\}.$$

□

The theorem implies that, if we can assure a constant decrease in the potential function at each iteration, we can obtain primal and dual feasible solutions that are within $\epsilon$ of the common optimal value in $O(\sqrt{\nu} \ln(1/\epsilon))$ iterations by choosing $\rho = \gamma\sqrt{\nu}$ for constant $\gamma \geq 1$, an improvement of a factor of $\sqrt{\nu}$ over the methods of the previous section. This symmetric primal-dual potential function was first used to establish such a complexity bound in Todd and Ye [10] for a short-step symmetric primal-dual algorithm. Ye [11] obtained a long-step algorithm which was not symmetric, and Kojima *et al.* [6] developed a symmetric primal-dual method. Nesterov and Nemirovskii [7], Section 4.5, extended Ye's algorithm to the setting of general cones. Here we extend the method of Kojima *et al.* when the cones are self-scaled:

1. **Initialization.** Choose $x_0 \in S^0(P)$ and $(y_0, s_0) \in S^0(D)$.

2. **$k$-th iteration ($k \geq 0$).**

   a) Compute the scaling point $z_k \in \operatorname{int} K$ such that

   $$s_k = F''(z_k)x_k,$$

   and the coefficient

   $$\sigma_k = \sigma_{x_k}(z_k).$$

   b) Compute the displacement $(\Delta x_k, \Delta s_k, \Delta y_k)$ as a solution of the following system of linear equations:

   $$F''(z_k)\Delta x_k + \Delta s_k = u_k := \frac{\nu+\rho}{\langle s_k, x_k\rangle} s_k + F'(x_k), \tag{8.1}$$

   $$A\Delta x_k = 0, \quad A^*\Delta y_k + \Delta s_k = 0.$$

   c) Find $\sigma_{x_k}(\Delta x_k)$ and $\sigma^*_{s_k}(\Delta s_k)$ and set $\bar{\sigma}_k := \max\{\sigma_{x_k}(\Delta x_k), \sigma^*_{s_k}(\Delta s_k)\}$. Choose the initial step size

   $$\bar{\alpha}_k = \frac{1}{\sigma_k^2 + \bar{\sigma}_k}$$

   and find the step size $\alpha_k$ from the condition

   $$\phi(x_k - \alpha_k\Delta x_k, s_k - \alpha_k\Delta s_k) \leq \phi(x_k - \bar{\alpha}_k\Delta x_k, s_k - \bar{\alpha}_k\Delta s_k).$$

   d) Set

   $$x_{k+1} = x_k - \alpha_k\Delta x_k, \quad s_{k+1} = s_k - \alpha_k\Delta s_k, \quad y_{k+1} = y_k - \alpha_k\Delta y_k.$$

   **End of iteration.**

Note that $u_k = \phi'_x(x_k, s_k)$, the partial derivative of $\phi$ with respect to the $x$-variable. We also observe that we can write (8.1) as

$$\Delta x_k + F''_*(t_k)\Delta s_k = w_k := \frac{\nu+\rho}{\langle s_k, x_k\rangle} x_k + F'_*(s_k),$$

where $t_k := -F'(z_k)$ so that $F''_*(t_k)s_k = x_k$, and note that $w_k = \phi'_s(x_k, s_k)$. Thus the direction choice is symmetric between $K$ and $K^*$. In addition, it is easy to see that $\sigma_k$ can equally be defined as $\sigma^*_{s_k}(t_k)$, so the choice of step size is also symmetric. Finally, note the similarity between (8.1) and the projection equations (6.8), demonstrating that $\Delta x_k$ is the projection of $\phi'_x(x_k, s_k)$ into the kernel of $A$ with respect to $F''(z)$, and similarly $\Delta s_k$ is the projection of $\phi'_s(x_k, s_k)$ into the range of $A^*$ with respect to $F''_*(t)$. Alternatively, in the Euclidean space $V$, the primal and dual steps (in that space) are the Euclidean projections of the corresponding gradients (which happen to be equal).

**Theorem 8.2** *For any $k \geq 0$ the following inequality holds:*

$$\phi(x_{k+1}, s_{k+1}) \leq \phi(x_k, s_k) - \frac{\sqrt{3}}{2} + \ln\left(1 + \frac{\sqrt{3}}{2}\right).$$

**Proof:**
Let us analyze one iteration of the method. In order to simplify the notation let us omit all indices denoting the number of the iteration. Denote $H = F''(z)$.

Note that in view of Theorem 3.2

$$\begin{array}{rcl} \mu^2 & := & (\| u \|^*_z)^2 = \frac{(\nu+\rho)^2}{\langle s,x\rangle^2}\langle s, H^{-1}s\rangle + 2\frac{\nu+\rho}{\langle s,x\rangle}\langle F'(x), H^{-1}s\rangle + \langle F'(x), H^{-1}F'(x)\rangle \\ & = & \frac{(\nu+\rho)^2 - 2\nu(\nu+\rho)}{\langle s,x\rangle} + \langle F'(x), F'_*(s)\rangle = \langle F'(x), F'_*(s)\rangle - \frac{\nu^2-\rho^2}{\langle s,x\rangle} \\ & \geq & \langle F'(x), F'_*(s)\rangle - \frac{\nu(\nu-1)}{\langle s,x\rangle} \end{array} \tag{8.2}$$

(recall that $\rho \geq \sqrt{\nu}$). On the other hand

$$\mu^2 = \langle H\Delta x + \Delta s, H^{-1}(H\Delta x + \Delta s)\rangle = \| \Delta x \|^2_z + (\| \Delta s \|^*_z)^2. \tag{8.3}$$

Let us introduce the function

$$\Delta\phi(\alpha) = \phi(x - \alpha\Delta x, s - \alpha\Delta s) - \phi(x, s).$$

Since

$$\begin{array}{rcl} \langle s - \alpha\Delta s, x - \alpha\Delta x\rangle - \langle s, x\rangle & = & -\alpha(\langle \Delta s, x\rangle + \langle s, \Delta x\rangle) \\ & = & -\alpha(\langle u - H\Delta x, x\rangle + \langle s, \Delta x\rangle) = -\alpha\langle u, x\rangle = -\alpha\rho \end{array}$$

(see (2.4)), in view of Theorem 4.2 we have:

$$\Delta\phi(\alpha) \leq (\nu+\rho)\ln\left(1 - \frac{\alpha\rho}{\langle s, x\rangle}\right) - \alpha\langle F'(x), \Delta x\rangle + \frac{\| \Delta x \|^2_x}{\sigma^2_x(\Delta x)}(-\alpha\sigma_x(\Delta x) - \ln(1 - \alpha\sigma_x(\Delta x)))$$

$$-\alpha\langle \Delta s, F'_*(s)\rangle + \frac{(\| \Delta s \|^*_s)^2}{(\sigma^*_s(\Delta s))^2}(-\alpha\sigma^*_s(\Delta s) - \ln(1 - \alpha\sigma^*_s(\Delta s))) \leq -\alpha\Delta_0 + \Delta_x(\alpha) + \Delta_s(\alpha),$$

where

$$\Delta_0 = \frac{\rho(\nu+\rho)}{\langle s, x\rangle} + \langle F'(x), \Delta x\rangle + \langle \Delta s, F'_*(s)\rangle,$$

$$\Delta_x(\alpha) = \frac{\| \Delta x \|_x^2}{\sigma_x^2(\Delta x)}(-\alpha\sigma_x(\Delta x) - \ln(1-\alpha\sigma_x(\Delta x))),$$

$$\Delta_s(\alpha) = \frac{(\| \Delta s \|_s^*)^2}{(\sigma_s^*(\Delta s))^2}(-\alpha\sigma_s^*(\Delta s) - \ln(1-\alpha\sigma_s^*(\Delta s))).$$

Let us estimate each $\Delta_{(\cdot)}$ separately.

In view of Theorem 3.2 and (2.4) we have:

$$\langle \Delta s, F_*'(s)\rangle = \langle \frac{\nu+\rho}{\langle s,x\rangle}s + F'(x) - H\Delta x, F_*'(s)\rangle$$

$$= -\frac{\nu(\nu+\rho)}{\langle s,x\rangle} + \langle F'(x), F_*'(s)\rangle - \langle F'(x), \Delta x\rangle.$$

Therefore

$$\Delta_0 = \langle F'(x), F_*'(s)\rangle - \frac{\nu^2-\rho^2}{\langle s,x\rangle} = \mu^2.$$

Further, by Corollary 4.1 (ii)

$$\| \Delta x \|_x^2 \leq \sigma^2 \| \Delta x \|_z^2, \tag{8.4}$$

and therefore from (4.4) and (8.3),

$$\sigma_x(\Delta x) \leq \| \Delta x \|_x \leq \sigma \| \Delta x \|_z \leq \sigma\mu \tag{8.5}$$

and similarly $\sigma_s^*(\Delta s) \leq \sigma\mu$. Recalling $\bar\sigma := \max\{\sigma_x(\Delta x), \sigma_s^*(\Delta s)\}$, we have

$$\max\{\sigma_x(\Delta x), \sigma_s^*(\Delta s)\} = \bar\sigma \leq \sigma\mu.$$

In view of Propsition 4.1(iii) and (8.4) we can estimate $\Delta_x(\alpha)$ as follows:

$$\Delta_x(\alpha) \leq \frac{\sigma^2 \| \Delta x \|_z^2}{\bar\sigma^2}(-\alpha\bar\sigma - \ln(1-\alpha\bar\sigma)).$$

The symmetric reasoning for $\Delta_s(\alpha)$ leads to the following estimate:

$$\Delta_s(\alpha) \leq \frac{\sigma^2(\| \Delta s \|_z^*)^2}{\bar\sigma^2}(-\alpha\bar\sigma - \ln(1-\alpha\bar\sigma)).$$

Therefore in view of (8.3) we conclude that

$$\Delta_x(\alpha) + \Delta_s(\alpha) \leq \frac{\sigma^2\mu^2}{\bar\sigma^2}(-\alpha\bar\sigma - \ln(1-\alpha\bar\sigma))$$

and we come to the following inequality:

$$\Delta\phi(\alpha) \leq -\alpha\sigma^2\mu^2\left(\frac{1}{\sigma^2} + \frac{1}{\bar\sigma}\right) - \frac{\sigma^2\mu^2}{\bar\sigma^2}\ln(1-\alpha\bar\sigma).$$

This is exactly the same form as the bound of (7.10), and is minimized by $\alpha = \bar{\alpha} := 1/[\sigma^2 + \bar{\sigma}]$, which yields

$$\Delta\phi(\bar{\alpha}) \leq -\frac{\mu^2}{\bar{\sigma}} + \frac{\sigma^2\mu^2}{\bar{\sigma}^2}\ln\left(1 + \frac{\bar{\sigma}}{\sigma^2}\right) = -\frac{\mu^2}{\sigma^2}\left[\frac{\tau - \ln(1+\tau)}{\tau^2}\right],$$

where $\tau := \bar{\sigma}/\sigma^2 \leq \mu/\sigma$. As before, in view of Proposition 4.1 (ii) the function in brackets is decreasing in $\tau$. Therefore we may replace $\tau$ by its upper bound:

$$\Delta\phi(\bar{\alpha}) \leq -\left(\frac{\mu}{\sigma} - \ln\left(1 + \frac{\mu}{\sigma}\right)\right).$$

It remains to note that by (8.2) and Theorem 5.2 we have $\mu/\sigma \geq \sqrt{3}/2$. $\square$

(The reader may be confused by the presence of $\sigma^2$ above, especially in conjunction with $\bar{\sigma}$ in the expression for $\bar{\alpha}$. Why is it squared? It turns out that $\sigma^2 = \sigma_x(-F'_*(s)) = \sigma^*_s(-F'(x))$, but the proof would take us rather far afield. Thus the step size is expressed in terms of the maximum steps to the boundary for $x$ or $s$ in the directions $\Delta x$ and $-F'_*(s)$ or the directions $\Delta s$ and $-F'(x)$, and is at least half the minimum of these steps.)

In view of Theorems 8.1 and 8.2, if we choose $\rho = \gamma\sqrt{\nu}$ for a constant $\gamma \geq 1$, we have the desired $O(\sqrt{\nu}\ln(1/\epsilon))$-iteration algorithm. In this algorithm, the fact that we have a self-scaled barrier is used to assure the existence of the scaling point $z$ (and hence get a symmetric method) as well as to allow a long step.

# 9 Newton's method and the path-following approach

Here we discuss the possibilities of applying to problem (P) the path-following approach, i.e., following the trajectory $\{x(\tau) : \tau > 0\}$, where $x(\tau)$ is the minimizer of the penalty function

$$\psi(\tau, x) := \tau\langle c, x\rangle + F(x)$$

over $S^0(P)$. Under assumptions (6.1)–(6.3), this trajectory is well-defined (see [7], [8]). Each point $x(\tau)$ is the unique solution of the following system:

$$\tau c + F'(x(\tau)) - A^*y = 0, \quad Ax(\tau) = b, \quad x \in \operatorname{int} K.$$

All complexity estimates for path-following methods are based on the behavior of Newton's method as applied to minimizing the function $\psi(\tau, x)$ over $S^0(P)$. The complete description of this behavior in the case of general cones is presented in [7], Section 2.2. In this section we improve several estimates of [7] using the specific properties of self-scaled cones.

Let us fix $\tau > 0$. Consider Newton's method in the following form:

1. **Initialization** Choose $x_0 \in S^0(P)$.

2. $k$**th iteration**

   a) Find the search direction $p_k := p(\tau, x_k)$, where $p(\tau, x)$ is the solution of the following system:
   $$\tau c + F'(x) - F''(x)p(\tau, x) - A^* y(\tau, x) = 0, \quad Ap(\tau, x) = 0,$$
   and $y(\tau, x) \in Y$.

   b) Choose a step size $\alpha_k$ such that $x_k - \alpha_k p_k \in \operatorname{int} K$.

   c) Set $x_{k+1} = x_k - \alpha_k p_k$.

   **End of iteration.**

We will consider two strategies for choosing $\alpha_k$. The first one can be applied at the initial stages of the minimization process.

**Theorem 9.1** *Let us choose in Newton's method*
$$\alpha_k = \frac{1}{1 + \sigma_{x_k}(p_k)}.$$
*Then for any $k \geq 0$ we have:*
$$\begin{aligned} \psi(\tau, x_{k+1}) &\leq \psi(\tau, x_k) - \left(\tfrac{\|p_k\|_{x_k}}{\sigma_{x_k}(p_k)}\right)^2 (\sigma_{x_k}(p_k) - \ln(1 + \sigma_{x_k}(p_k))) \\ &\leq \psi(\tau, x_k) - (\| p_k \|_{x_k} - \ln(1 + \| p_k \|_{x_k}) \end{aligned} \tag{9.1}$$

**Proof:**
By Proposition 6.2
$$\langle \tau c + F'(x_k), p_k \rangle = \| p_k \|^2_{x_k} .$$
Therefore the first inequality of (9.1) is a direct consequence of the inequality (4.7). The second follows from Proposition 4.1(ii). □

Let us introduce the following proximity measure:
$$\pi(\tau, x) := \| p(\tau, x) \|_x .$$
It is clear that $\pi(\tau, x(\tau)) = 0$. We will prove that the inequality
$$\pi(\tau, x) < 1$$
defines a region of quadratic convergence of Newton's method (with coefficient 1), so that in this region we should choose $\alpha_k = 1$. This result extends Theorem 2.1 of Roos and Vial [9]. Note that in view of Proposition 6.3 we have
$$\pi(\tau, x) \leq \| F'(x) + \tau c - A^* y \|^*_x . \tag{9.2}$$
for any $\tau > 0$, $x \in \operatorname{int} K$, and $y \in Y$.

**Theorem 9.2** *Let $\pi(\tau, x_0) < 1$ and let us choose in Newton's method $\alpha_k = 1$. Then*

$$\pi(\tau, x_{k+1}) \leq | p_k |_{x_k} \pi(\tau, x_k) \leq \pi^2(\tau, x_k).$$

**Proof:**
Indeed, in view of (9.2) and the definition of $p_k$ we have:

$$\begin{aligned} \pi(\tau, x_{k+1}) &\leq \| F'(x_{k+1}) + \tau c - A^* y(\tau, x_k)) \|^*_{x_{k+1}} \\ &= \| F'(x_{k+1}) - F'(x_k) - F''(x_k)(x_{k+1} - x_k) \|^*_{x_{k+1}} . \end{aligned}$$

It remains to use Theorem 4.3. □

Thus, we have demonstrated that using Newton's method we can find a good approximation to a point on the central path $x(\tau)$. Let us prove that we can follow the central path when $\tau \to \infty$ by updating points in the region of quadratic convergence of Newton's method.

Consider the following scheme:

1. **Initialization** Choose a value $\tau_0 > 0$ and a point $x_0 \in S^0(P)$ such that

$$\pi(\tau_0, x_0) \leq \frac{1}{4}.$$

2. **$k$th iteration**

   a) Set $v_k = v(x_k)$, where $v(x)$ is the solution of the following linear system:

$$c - F''(x_k)v(x) - A^* \hat{y}(x) = 0, \quad Av(x) = 0,$$

   and $\hat{y}(x) \in Y$.

   b) Choose the tangent step size $\Delta\tau_k$ as follows:

$$\Delta\tau_k = \frac{3}{8\sqrt{| v_k |_{x_k} \| v_k \|_{x_k}}}.$$

   Set (predictor step)

$$z_k = x_k - \Delta\tau_k v_k, \quad \tau_{k+1} = \tau_k + \Delta\tau_k.$$

   c) Set (corrector step)

$$x_{k+1} = z_k - p(\tau_{k+1}, z_k).$$

   **End of iteration.**

**Theorem 9.3** *The above path-following scheme preserves the condition*

$$\pi(\tau_k, x_k) \leq \frac{1}{4}, \quad k = 0, 1, \ldots. \tag{9.3}$$

*Further, for any $k \geq 0$, the following inequality holds:*

$$\tau_{k+1} \geq \left(1 + \frac{3}{8\sqrt{\nu} + 2}\sqrt{\frac{\| v_k \|_{x_k}}{| v_k |_{x_k}}}\right) \tau_k \geq \left(1 + \frac{3}{8\sqrt{\nu} + 2}\right) \tau_k. \tag{9.4}$$

**Proof:**
In view of Theorem 9.2, in order to prove (9.3) we shall prove that

$$\pi(\tau_{k+1}, z_k) \leq \tfrac{1}{2}.$$

Therefore, by Proposition 6.3, it is enough to prove that

$$\| F'(z_k) + (\tau_k + \Delta\tau_k)c - A^*\bar{y} \|_{z_k}^* \leq \tfrac{1}{2}$$

for some $\bar{y} \in Y$.

Indeed, let us choose

$$\bar{y} = y(\tau_k, x_k) + \Delta\tau_k \hat{y}(x_k).$$

Then

$$\begin{aligned} F'(z_k) + (\tau_k + \Delta\tau_k)c - A^*\bar{y} &= F'(z_k) - F'(x_k) + \Delta\tau_k F''(x_k)v_k + F''(x_k)p(\tau_k, x_k) \\ &= F'(z_k) - F'(x_k) - F''(x_k)(z_k - x_k) + F''(x_k)p(\tau_k, x_k). \end{aligned}$$

Therefore in view of Theorem 4.3 we have:

$$\| F'(z_k) + (\tau_k + \Delta\tau_k)c - A^*\bar{y} \|_{z_k}^* \leq \Delta\tau_k^2 \mid v_k \mid_{x_k} \cdot \| v_k \|_{x_k} + \| F''(x_k)p(\tau_k, x_k) \|_{z_k}^* .$$

Note that from Theorem 4.1

$$F''(z_k) \geq \frac{1}{(1 + \Delta\tau_k \mid v_k \mid_{x_k})^2} F''(x_k).$$

This implies that

$$\| F''(x_k)p(\tau_k, x_k) \|_{z_k}^* \leq (1 + \Delta\tau_k \mid v_k \mid_{x_k})\pi(\tau_k, x_k)$$

$$\leq \frac{1}{4}(1 + \Delta\tau_k \mid v_k \mid_{x_k}) = \frac{1}{4}\left(1 + \frac{3}{8}\sqrt{\frac{\mid v_k \mid_{x_k}}{\| v_k \|_{x_k}}}\right) \leq \frac{11}{32}.$$

Thus, we conclude that

$$\| F'(z_k) + (\tau_k + \Delta\tau_k)c - A^*\bar{y} \|_{z_k}^* \leq \frac{9}{64} + \frac{11}{32} < \frac{1}{2}.$$

Let us now prove inequality (9.4). Note that

$$\tau_{k+1} = \tau_k + \Delta\tau_k = \tau_k + \frac{3}{8 \| v_k \|_{x_k}} \sqrt{\frac{\| v_k \|_{x_k}}{\mid v_k \mid_{x_k}}} \geq \tau_k + \frac{3}{8 \| v_k \|_{x_k}}.$$

Thus, we only have to prove that

$$\| \tau_k v_k \|_{x_k} \leq \sqrt{\nu} + \frac{1}{4}. \tag{9.5}$$

Let us choose

$$\bar{y} = \frac{1}{\tau_k} y(\tau_k, x_k).$$

Then in view of Proposition 6.3 and the definitions of $v_k$, $p(\tau_k, x_k)$ and $y(\tau_k, x_k)$ we have:

$$\begin{aligned}
\| \tau_k v_k \|_{x_k} &\leq \| \tau_k c - \tau_k A^* \bar{y} \|^*_{x_k} \\
&= \| F'(x_k) - F''(x_k) p(\tau_k, x_k) - A^*(y(\tau_k, x_k) - \tau_k \bar{y}) \|^*_{x_k} \\
&= \| F'(x_k) - F''(x_k) p(\tau_k, x_k) \|^*_{x_k} \\
&\leq \| F'(x_k) \|^*_{x_k} + \| F''(x_k) p(\tau_k, x_k) \|^*_{x_k} \leq \sqrt{\nu} + \frac{1}{4}.
\end{aligned}$$

Thus, (9.5) and therefore (9.4) is proved. $\square$

Note that $\Delta\tau_k$ in part (b) of the algorithm is chosen in accordance with a kind of "large-step" rule. Indeed, let

$$\bar{v}_k = \frac{v_k}{\| v_k \|_{x_k}}.$$

Then the maximal step $\alpha^*$ such that $x_k \pm \alpha^* \bar{v}_k \in K$ is given by the formula:

$$\alpha^* = \frac{\| v_k \|_{x_k}}{| v_k |_{x_k}}.$$

Note that $\alpha^* \geq 1$. And it is easy to see that our rule is

$$z_k = x_k - \frac{3\sqrt{\alpha^*}}{8} \bar{v}_k.$$

Of course, this rule depends on worst-case behavior. We can alternatively increase $\tau_k$ using an adaptive rule, which could allow even faster increase.

# References


[1] K. M. Anstreicher, A strengthened acceptance criterion for approximate projections in Karmarkar's algorithm. *Operations Research Letters*, 5:211–214, 1986.

[2] A. Ben-Tal and A. Nemirovskii, Potential reduction polynomial time method for truss topology design. Faculty of Industrial Engineering and Management, Technion, Haifa, 1993.

[3] I. Ekeland and R. Temam. Convex Analysis and Variational Problems. North-Holland, Amsterdam, 1976.

[4] C. C. Gonzaga. Polynomial affine algorithms for linear programming. *Mathematical Programming*, 49:7–21, 1990.

[5] N. K. Karmarkar. A new polynomial–time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[6] M. Kojima, S. Mizuno, and A. Yoshise. An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems. *Mathematical Programming*, 50:331–342, 1991.

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

[8] Yu. Nesterov, Long-step strategies in interior point potential-reduction methods. Department SES COMIN, Geneva University, 1993.

[9] C. Roos and J. P. Vial. A polynomial method of approximate centers for linear programming. *Mathematical Programming*, 54:295–305, 1992.

[10] M. J. Todd and Y. Ye. A centered projective algorithm for linear programming. *Mathematics of Operations Research*, 15:508–529, 1990.

[11] Y. Ye. An $O(n^3L)$ potential reduction algorithm for linear programming. In J. C. Lagarias and M. J. Todd, editors, *Mathematical Developments Arising from Linear Programming : Proceedings of a Joint Summer Research Conference held at Bowdoin College, Brunswick, Maine, USA, June/July 1988*, volume 114 of *Contemporary Mathematics*, pages 77–88. American Mathematical Society, Providence, Rhode Island, USA, 1990.