PRIMAL-DUAL INTERIOR-POINT METHODS FOR SELF-SCALED CONES *

YU. E. NESTEROV† AND M. J. TODD‡

Abstract. In this paper we continue the development of a theoretical foundation for efficient primal-dual interior-point algorithms for convex programming problems expressed in conic form, when the cone and its associated barrier are *self-scaled* (see [NT97]). The class of problems under consideration includes linear programming, semidefinite programming and convex quadratically constrained quadratic programming problems. For such problems we introduce a new definition of affine-scaling and centering directions. We present efficiency estimates for several symmetric primal-dual methods that can loosely be classified as path-following methods. Because of the special properties of these cones and barriers, two of 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. Convex programming, conical form, interior-point algorithms, self-concordant barrier, self-scaled cone, self-scaled barrier, path-following algorithms, symmetric primal-dual algorithms.

AMS subject classifications. 90C05, 90C25, 65Y20

1. Introduction.

This paper continues the development begun in [NT97] of a theoretical foundation for efficient interior-point methods for problems that are extensions of linear programming. Here we concentrate on symmetric primal-dual algorithms that can loosely be classified as path-following methods. While 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. Note that any convex programming problem can be expressed in this conical form.

Nesterov and Nemirovskii [NN94] 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 [Ka84]) 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.

In [NT97] Nesterov and Todd introduced a special class of *self-scaled* convex cones and associated barriers. While they are required to satisfy certain apparently restrictive conditions, this class includes some important instances, for example the cone of (symmetric) positive semidefinite matrices and the second-order cone, as well as the nonnegative orthant in R^n. In fact, these cones (and their direct products) are perhaps the only self-scaled cones interesting for optimization. It turns out that self-scaled cones coincide with homogeneous self-dual cones as was pointed out by Güler [Gu96] (see also the discussion in [NT97]), and these have been characterized as being built up via direct products from the cones listed above, cones of positive semidefinite Hermitian complex and quaternion matrices, and one exceptional cone. However, for the reasons given in detail at the end of this introduction, we believe that it is simpler to carry out the analysis and develop algorithms in the abstract setting of self-scaled cones. We maintain the name self-scaled to emphasize our primary interest in the associated self-scaled barriers.

---

† CORE, Catholic University of Louvain, Louvain-la-Neuve, Belgium. E-mail: nesterov@core.ucl.ac.be. Part of this research was supported by the Russian Fund of Fundamental Studies, Grant 96-01-00293.

‡ School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853. E-mail: miketodd@cs.cornell.edu. Some of this work was done while the author was on a sabbatical leave from Cornell University visiting the Department of Mathematics at the University of Washington and the Operations Research Center at the Massachusetts Institute of Technology. Partial support from the University of Washington and from NSF Grants CCR-9103804, DMS-9303772, and DDM-9158118 is gratefully acknowledged.

For such cones, the Hessian of the barrier at any interior point maps the cone onto its dual cone, and *vice versa* for 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 a unique *scaling* 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).

These conditions have extensive consequences. 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 [NT97] developed primal long-step potential-reduction and path-following algorithms as well as a symmetric long-step primal-dual potential-reduction method.

In this paper we present some new properties of self-scaled cones and barriers that are necessary for deriving and analyzing primal-dual path-following interior-point methods.

The first part of the paper continues the study of the properties of self-scaled barriers started in [NT97]. In Section 2 we introduce the problem formulation and state the main definitions and our notation. In Section 3 we prove some symmetry results for primal and dual norms and study the properties of the third derivative of a self-scaled barrier and the properties of the scaling point. In Section 4 we introduce the definitions of the primal-dual central path and study different proximity measures. Section 5 is devoted to generalized affine-scaling and centering directions.

The second part of the paper applies these results to the derivation of short- and long-step symmetric primal-dual methods that follow the central path in some sense, using some proximity measure. Primal-dual path-following methods like those of Monteiro and Adler [MA89] and Kojima, Mizuno and Yoshise [KMY89] are studied in Section 6. These methods follow the path closely using a local proximity measure, and take short steps. At the end of this section we consider an extension of the predictor-corrector method of Mizuno, Todd and Ye [MTY90], which uses the affine-scaling and centering directions of Section 5, also relying on the same proximity measure. This algorithm employs an adaptive step size rule. In Section 7 we present another predictor-corrector algorithm, a variant of the functional-proximity path-following scheme proposed by Nesterov [Ne96] for general nonlinear problems. In the case of self-scaled cones the method is based on the affine-scaling and centering directions. This algorithm uses a global proximity measure and allows considerable deviation from the central path; thus much longer steps are possible than in the algorithms of the previous section. Nevertheless, the results here can also be applied to the predictor-corrector method considered there.

All these methods require $O(\sqrt{\nu}\ln(1/\epsilon))$ iterations to generate a feasible solution with objective function within ϵ of the optimal value, where ν 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, for the predictor-corrector methods of the last two sections we can prove that the predictor step size is at least a constant times the square root of the maximal possible step size along the affine-scaling direction. Alternatively, it can be bounded from below by a constant times the maximal possible step size divided by $\nu^{1/4}$. These results were previously unknown even in the context of linear programming.

In order to motivate our development, let us explain why we carry out our analysis in the abstract setting of self-scaled cones, rather than restricting ourselves to say semidefinite programming. After all, the latter context includes linear programming, and, as has been shown by several authors including Nesterov and Nemirovskii [NN94], convex quadratically constrained quadratic programming. However, in the latter case, the associated barrier parameter from the semidefinite programming formulation is far larger than it need be, and the complexity of the resulting algorithm also correspondingly too large. In contrast, using the second-order (or Lorentz, or ice-cream) cone permits one to employ the appropriate barrier and corresponding efficient algorithms. The abstract setting of self-scaled cones and barriers allows us to treat each problem with its natural formulation and barrier. Thus if a semidefinite programming problem has a positive semidefinite matrix as a variable, and also involves inequality constraints, we do not need to use the slack variables for the constraints to augment the positive semidefinite matrix and thus deal with block diagonal matrices; instead, our variable is just the combination of the original matrix variable and the slack vector, and the self-scaled cone in which it is constrained to lie is the product of a positive semidefinite cone and a nonnegative orthant of appropriate dimensions. In this way, it is in fact easier, although more abstract, to work in the more general setting. Similarly, nonnegative variables are dealt with directly (with the cone the nonnegative orthant), rather than embedded into the diagonal of a positive semidefinite matrix, so our algorithms apply naturally to linear programming as well as semidefinite programming, without having to consider what happens when the appropriate matrices are diagonal. We can also handle more general constraints, such as convex quadratic inequalities on symmetric positive semidefinite matrices, by using combinations of the positive semidefinite cone and the second-order cone. Finally, we can write linear transformations in the general form $x \to Ax$, rather than considering specific forms for semidefinite programming such as $X \to PXP^T$; similar comments apply to the bilinear and trilinear expressions that arise when dealing with second and third derivatives. Thus we obtain unified methods and unified analyses for methods that apply to all of these cases directly, rather than through reformulations; moreover, we see exactly what properties of the cones and the barriers lead to the desired results and algorithms. We will explain several of our constructions below for the case of linear or semidefinite programming to aid the reader, but the development will be carried out in the framework of abstract self-scaled cones for the reasons given above.

In what follows we often refer to different statements of [NN94] and [NT97]. The corresponding references are indicated by asterisks. An upper-case asterisk in 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 [NN94]. A lower-case asterisk, for example T_* 1.1 (or $(1.1)_*$), corresponds to the first theorem (or equation) in Section 1 of [NT97].

2. Problem Formulation and Notation. 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).

The problem we are concerned with is:

$$\begin{array}{llrcl} \text{(P)} & \min & \langle c, x\rangle & & \\ & \text{s.t.} & Ax & = & b, \\ & & x & \in & K, \end{array} \tag{2.1}$$

where we assume that

$$\begin{array}{l} A \text{ is a surjective linear operator from } E \text{ to} \\ \text{another finite-dimensional real vector space } Y^*. \end{array} \tag{2.2}$$

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).

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. Then the dual to problem (P) is (see [ET76]):

$$\begin{array}{llrcrcl} \text{(D)} & \max & \langle b, y\rangle & & & & \\ & \text{s.t.} & A^*y & + & s & = & c, \\ & & & & s & \in & K^*, \end{array} \tag{2.3}$$

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

If K is the nonnegative orthant in R^n, then (P) and (D) are the standard-form linear programming problem and its dual; if K is the cone of positive semidefinite matrices of order n, $\langle c, x\rangle := \text{Trace}\,(c^T x)$, and $Ax := (\langle a_k, x\rangle) \in R^m$ where the a_k's are symmetric matrices of order n, we obtain the standard-form semidefinite programming problem and its dual; here A^*y is $\sum y_k a_k$. In both these cases, K^* can be identified with K. The surjectivity assumption requires that the constraint matrix have full row rank in the first case, and that the a_k's be linearly independent in the second.

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

$$S^0(P) := \{x \in \operatorname{int} K : Ax = b\} \text{ is nonempty,} \tag{2.4}$$

and

$$S^0(D) := \{(y,s) \in Y \times \operatorname{int} K^* : A^*y + s = c\} \text{ is nonempty.} \tag{2.5}$$

These assumptions imply (see [NN94], 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 [Ne96]). 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.*

Our algorithms for (P) and (D) require that the cone K be *self-scaled,* i.e., admit a *self-scaled barrier.* We now introduce the terminology to define this concept.

Let F be a ν-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:

$$F(\tau x) \equiv F(x) - \nu \ln \tau. \tag{2.6}$$

Since K is a pointed cone, $\nu \geq 1$ in view of C^* 2.3.3. The reader might want to keep in mind for the properties below the trivial example where K is the nonnegative half-line and F the function $-\ln(\cdot)$, with $\nu = 1$. The two main interesting cases are where K is the nonnegative orthant R^n_+ and F the standard logarithmic barrier function ($F(x) := -\ln(x) := -\sum_{j=1}^{n} \ln x^{(j)}$), and where K is the cone S^n_+ of symmetric positive semidefinite $n \times n$ matrices and F the log-determinant barrier function ($F(x) := -\ln\det x$); in both cases, $\nu = n$. In the first case, $F'(x) = -([x^{(j)}]^{-1})$, the negative of the vector of reciprocals of the components of x, with $F''(x) = X^{-2}$, where X is the diagonal matrix containing the components of x; in the second, $F'(x) = -x^{-1}$, the negative of the inverse matrix, while $F''(x)$ is the linear transformation defined by $F''(x)v = x^{-1}vx^{-1}$.

We will often use the following straightforward consequences of (2.6): for all $x \in \operatorname{int} K$ and $\tau > 0$,

$$F'(\tau x) = \frac{1}{\tau} F'(x), \quad F''(\tau x) = \frac{1}{\tau^2} F''(x), \tag{2.7}$$

$$F''(x)x = -F'(x), \quad F'''(x)[x] = -2F''(x), \tag{2.8}$$

$$\langle F'(x), x\rangle = -\nu, \tag{2.9}$$

$$\langle F''(x)x, x\rangle = \nu, \quad \langle F'(x), [F''(x)]^{-1}F'(x)\rangle = \nu \tag{2.10}$$

(see P^* 2.3.4).

Let the function F_* on $\operatorname{int} K^*$ be conjugate to F, namely:

$$F_*(s) := \max\{-\langle s, x\rangle - F(x) : x \in K\}. \tag{2.11}$$

In accordance with T^* 2.4.4, F_* is a ν-self-concordant logarithmically homogeneous barrier for K^*. (If $F(x) = -\ln(x)$ for $x \in \operatorname{int} R^n_+$, $F_*(s) = -\ln(s) - n$, while if $F(x) = -\ln\det x$ for $x \in \operatorname{int} S^n_+$, $F_*(s) = -\ln\det s - n$.) We will often use the following properties of conjugate self-concordant barriers for dual cones: for any $x \in \operatorname{int} K$ and $s \in \operatorname{int} K^*$,

$$-F'(x) \in \operatorname{int} K^*, \quad -F'_*(s) \in \operatorname{int} K, \tag{2.12}$$

$$F_*(-F'(x)) = \langle F'(x), x\rangle - F(x) = -\nu - F(x) \tag{2.13}$$

(using (2.9)),

$$F(-F'_*(s)) = -\nu - F_*(s), \tag{2.14}$$

$$F'_*(-F'(x)) = -x, \quad F'(-F'_*(s)) = -s, \tag{2.15}$$

$$F''(-F'_*(s)) = [F''_*(s)]^{-1}, \quad F''_*(-F'(x)) = [F''(x)]^{-1}, \tag{2.16}$$

$$F(x) + F_*(s) \geq -\nu + \nu \ln \nu - \nu \ln\langle s, x\rangle, \tag{2.17}$$

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, as in [NT97], we consider cones and barriers of a rather special type. Let us give our main definition.

DEFINITION 2.1. *Let K be a pointed cone with nonempty interior and let F be a ν-self-concordant logarithmically homogeneous barrier for cone K. We call F a ν-self-scaled barrier for K if for any w and x from* $\operatorname{int} K$*,*

$$F''(w)x \in \operatorname{int} K^* \tag{2.18}$$

and

$$F_*(F''(w)x) = F(x) - 2F(w) - \nu. \tag{2.19}$$

If K admits such a barrier, we call it a self-scaled cone.
Note that the identity (2.19) has an equivalent form: if $x \in \operatorname{int} K$ and $s \in \operatorname{int} K^*$ then

$$F([F''(x)]^{-1}s) = F_*(s) + 2F(x) + \nu \tag{2.20}$$

(merely write x for w and $[F''(x)]^{-1}s$ for x in (2.19)). Tunçel [Tu96] has given a simpler equivalent form (using (2.15)): if $F''(w)x = -F'(v)$, then

$$F(w) = [F(x) + F(v)]/2.$$

In fact, self-scaled cones coincide with homogeneous self-dual cones (see Güler [Gu96] and the discussion in [NT97]), but we will maintain the name self-scaled to emphasize our interest in the associated self-scaled barriers.

One important property of a self-scaled barrier is the existence of *scaling* points (see T_* 3.2): for every $x \in \operatorname{int} K$ and $s \in \operatorname{int} K^*$, there is a unique point $w \in \operatorname{int} K$ satisfying

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

For example, when K is the nonnegative orthant in R^n the scaling point w is given by the following componentwise expression:

$$w^{(j)} = [x^{(j)}/s^{(j)}]^{1/2};$$

when K is the cone of positive semidefinite matrices of order n then

$$w = x^{1/2}[x^{1/2}sx^{1/2}]^{-1/2}x^{1/2} = s^{-1/2}(s^{1/2}xs^{1/2})^{1/2}s^{-1/2}.$$

In what follows we always assume that cone K is self-scaled and F is a corresponding ν-self-scaled barrier. Even though we know that we are in a self-dual setting, so that we could identify E and E^* and

K and K^*, we prefer to keep the distinction, which allows us to distinguish primal and dual norms merely by their arguments, and makes sure that our methods are scale-invariant; indeed, it is impossible to write methods that are not, such as updating x to $x - F'(x)$ or to $x - s$, since the points lie in different spaces. We note that the two examples given above, the logarithmic barrier function for the nonnegative orthant in R^n and the log-determinant barrier for the cone of symmetric positive semidefinite $n \times n$ matrices, are both self-scaled barriers with parameter n. For example, in the second case we find $F''(v)x = v^{-1}xv^{-1}$ is positive definite if x and v are, with

$$F_*(F''(v)x) = -\ln\det(v^{-1}xv^{-1}) - n = -\ln\det x + 2\ln\det v - n = F(x) - 2F(v) - n,$$

as desired. In addition, [NT97] shows that the standard barrier for the second-order cone is self-scaled, and also that the Cartesian product of self-scaled cones is also self-scaled, with associated barrier the sum of the individual self-scaled barriers on the components; its parameter is just the sum of the parameters of the individual barriers.

3. Some results for self-scaled barriers. In this section we present some properties of self-scaled barriers, which complement the results of Sections $3_* - 5_*$. The reader may prefer to skip the proofs at a first reading.

We will use the following convention in notation: the meaning of the terms $\| u \|_v$, $| u |_v$ and $\sigma_v(u)$ is completely defined by the spaces of arguments (which are always indicated explicitly in the text). Thus, if $v \in \operatorname{int} K$ then

- $\| u \|_v$ means $\langle F''(v)u, u\rangle^{1/2}$ if $u \in E$, and $\langle u, [F''(v)]^{-1}u\rangle^{1/2}$ if $u \in E^*$;
- $\sigma_v(u) := \frac{1}{\alpha}$, where $\alpha > 0$ is either the maximal possible step (possibly $+\infty$) in the cone K such that $v - \alpha u \in K$, if $u \in E$, or the maximal possible step in the cone K^* such that $-F'(v) - \alpha u \in K^*$, if $u \in E^*$; equivalently, $\sigma_v(u)$ is the minimum nonnegative β such that either $\beta v - u \in K$, if $u \in E$, or $-\beta F'(v) - u \in K^*$, if $u \in E^*$.

Similarly, if $v \in \operatorname{int} K^*$ then

- $\| u \|_v$ means $\langle u, F_*''(v)u\rangle^{1/2}$ if $u \in E^*$, and $\langle [F_*''(v)]^{-1}u, u\rangle^{1/2}$ if $u \in E$;
- $\sigma_v(u) := \frac{1}{\alpha}$, where $\alpha > 0$ is either the maximal possible step in the cone K^* such that $v - \alpha u \in K^*$, if $u \in E^*$, or the maximal possible step in the cone K such that $-F_*'(v) - \alpha u \in K$, if $u \in E$; equivalently, $\sigma_v(u)$ is the minimum nonnegative β such that either $\beta v - u \in K^*$, if $u \in E^*$, or $-\beta F_*'(v) - u \in K$, if $u \in E$.

In this notation

$$| u |_v := \max\{\sigma_v(u), \sigma_v(-u)\}.$$

For a self-adjoint linear operator B from E to E^* or vice versa we define its norm as follows: for $v \in \operatorname{int} K$ or $v \in \operatorname{int} K^*$, let

$$\| B \|_v := \max\{\| Bu \|_v : \| u \|_v \leq 1\},$$

where u belongs to the domain of B.

Thus if, for example, $x \in \operatorname{int} K$, $v, z \in E$, and $B : E \to E^*$, then

$$\langle Bv, z\rangle \leq \| Bv \|_x \cdot \| z \|_x \leq \| B \|_x \cdot \| v \|_x \cdot \| z \|_x . \tag{3.1}$$

Note also that

$$\| B \|_x \leq \alpha \quad \Leftrightarrow \quad -\alpha F''(x) \leq B \leq \alpha F''(x). \tag{3.2}$$

Let us illustrate these definitions when E and E^* are R^n, and K and K^* their nonnegative orthants. Let $u.{*}v$ ($u./v$) denote the vector whose components are the products (ratios) of the corresponding components of u and v. Then $\| u \|_v$ is the 2-norm of $u./v$ if u and v belong to the same space and of $u.{*}v$ if they belong to different spaces; $\sigma_v(u)$ is the maximum of 0 and the maximum component of $u./v$ or of $u.{*}v$; and $| u |_v$ is the infinity- or max-norm of the appropriate vector. Suppose next that E and E^* are spaces of symmetric matrices, and K and K^* their cones of positive semidefinite elements. Writing $\chi(x)$ for the vector of eigenvalues of the matrix x and $\chi_{\max}(x)$ and $\chi_{\min}(x)$ for its largest and smallest eigenvalues, we find that $\| u \|_v$ is the 2-norm of $\chi(v^{-1/2}uv^{-1/2})$ if both u and v lie in the same space, or of $\chi(v^{1/2}uv^{1/2})$ otherwise. Similarly, $\sigma_v(u)$ is the maximum of 0 and $\chi_{\max}(v^{-1/2}uv^{-1/2})$ or $\chi_{\max}(v^{1/2}uv^{1/2})$, and $| u |_v$ is the infinity- or max-norm of the appropriate vector of eigenvalues. Finally, if $B : E \to E^*$ is defined by $Bu := w^{-1}uw^{-1}$, with u and w in E, then for $v \in E$, $\| B \|_v = \max\{\| \chi(v^{1/2}w^{-1}uw^{-1}v^{1/2}) \|_2 \colon \| \chi(v^{-1/2}uv^{-1/2}) \|_2 \leq 1\} = \max\{\| \chi(v^{1/2}w^{-1}v^{1/2}xv^{1/2}w^{-1}v^{1/2}) \|_2 \colon \| \chi(x) \|_2 \leq 1\}$. It can be seen that our general notation is somewhat less cumbersome.

Returning to the general case, let us first demonstrate that $| \cdot |_v$ is indeed a norm (it is clear that $\| \cdot \|_v$ is):

PROPOSITION 3.1. *$| \cdot |_v$ defines a norm on E and on E^*, for v in $\operatorname{int} K$ or $\operatorname{int} K^*$.*

Proof. Without loss of generality, we assume $v \in \operatorname{int} K$ and that we are considering $| \cdot |_v$ defined on E. From the definition it is clear that each of $\sigma_v(z)$ and $\sigma_v(-z)$ is positively homogeneous, and hence so is $| \cdot |_v$. Moreover, clearly $| -z |_v = | z |_v$. It is also immediate that $\sigma_v(z) = 0$ if $-z \in K$, and (using the second definition of $\sigma_v(z)$) that the converse holds also. Thus, since K is pointed, $| z |_v = 0$ iff $z = 0$. It remains to show that $| \cdot |_v$ is subadditive.

Let $\beta_i := \sigma_v(z_i)$ for $i = 1, 2$. Then $\beta_i v - z_i \in K$ for each i, so that $(\beta_1 + \beta_2)v - (z_1 + z_2) \in K$, and thus $\sigma_v(z_1 + z_2) \leq \beta_1 + \beta_2$. We conclude that $\sigma_v(z_1 + z_2) \leq \sigma_v(z_1) + \sigma_v(z_2) \leq | z_1 |_v + | z_2 |_v$. This inequality also holds when we change the signs of the z_i's, and then taking the maximum shows that $| \cdot |_v$ is subadditive as required. ∎

Henceforth, we usually follow the convention that s, t, and u belong to E^*, with s and t in $\operatorname{int} K^*$, while p, v, w, x, and z belong to E, with x and w in $\operatorname{int} K$. We reserve w for the *scaling* point such that $F''(w)x = s$.

3.1. Some symmetry results. Let us prove several symmetry results for the norms defined by a self-scaled barrier.

LEMMA 3.2. *a) Let $x \in \operatorname{int} K$ and $s \in \operatorname{int} K^*$. Then for any values of τ and ξ we have:*

$$\| \tau x + \xi F_*'(s) \|_x = \| \tau s + \xi F'(x) \|_s, \tag{3.3}$$

$$\| \tau x + \xi F_*'(s) \|_s = \| \tau s + \xi F'(x) \|_x . \tag{3.4}$$

b) For any $x, v \in \operatorname{int} K$

$$\| v \|_x = \| F'(x) \|_v .$$

Proof. For part (a), let us choose $w \in K$ such that $s = F''(w)x$ (see T_* 3.2). Then

$$F'(x) = F''(w)F_*'(s), \quad F''(x) = F''(w)F_*''(s)F''(w).$$

Also, let $v \in E$ and $t := F''(w)v \in E^*$. Then we obtain:

$$\| v \|_x^2 = \langle F''(x)v, v\rangle = \langle F''(w)F_*''(s)F''(w)v, v\rangle = \langle t, F_*''(s)t\rangle = \| t \|_s^2 .$$

In particular we can take $v = \tau x + \xi F_*'(s)$ and then $t = \tau s + \xi F'(x)$. This proves the first equation; the second is proved similarly. Part (b) follows from the first equation by taking $s = -F'(v)$, $\tau = 0$, and $\xi = 1$. ∎

For semidefinite programming, (3.3) above states that the 2-norm of the vector of eigenvalues of $\tau i - \xi x^{-1/2}s^{-1}x^{-1/2} = x^{-1/2}(\tau x - \xi s^{-1})x^{-1/2}$ is the same as for $\tau i - \xi s^{-1/2}x^{-1}s^{-1/2} = s^{-1/2}(\tau s - \xi x^{-1})s^{-1/2}$; (3.7) below says the same for the infinity-norm (here i denotes the identity). This is clear, since the two matrices are similar and hence have the same eigenvalues. To see the steps of the proof in this setting, recall that, in this case, the point w is $s^{-1/2}(s^{1/2}xs^{1/2})^{1/2}s^{-1/2} = x^{1/2}(x^{1/2}sx^{1/2})^{-1/2}x^{1/2}$.

For $\tau = 1$ and $\xi = 0$, part (a) of the lemma yields $(4.1)_*$. Note that, for w as above, we also have $\| \tau s + \xi F'(x) \|_w = \| \tau x + \xi F_*'(s) \|_w$.

LEMMA 3.3. *Let $x \in \operatorname{int} K$ and $s \in \operatorname{int} K^*$. Then for any values of τ and ξ we have:*

$$\sigma_x(\tau x + \xi F'_*(s)) = \sigma_s(\tau s + \xi F'(x)), \tag{3.5}$$

$$\sigma_s(\tau x + \xi F'_*(s)) = \sigma_x(\tau s + \xi F'(x)), \tag{3.6}$$

$$|\ \tau x + \xi F'_*(s)\ |_x = |\ \tau s + \xi F'(x)\ |_s, \tag{3.7}$$

$$|\ \tau x + \xi F'_*(s)\ |_s = |\ \tau s + \xi F'(x)\ |_x\ . \tag{3.8}$$

Proof. Again choose $w \in K$ such that $s = F''(w)x$ (see T_* 3.2). Then $F'(x) = F''(w)F'_*(s)$. Then in view of T_* 3.1 (iii), the inclusion

$$x - \alpha(\tau x + \xi F'_*(s)) \in K$$

is true for some $\alpha \geq 0$ if and only if

$$s - \alpha(\tau s + \xi F'(x)) \in K^*.$$

This proves (3.5); the proof of the (3.6) is similar. Then (3.7) and (3.8) follow from these. $\square$
Again, for w as above, we have also $|\ \tau s + \xi F'(x)\ |_w = |\ \tau x + \xi F'_*(s)\ |_w$.

Let us prove now some relations among the σ-measures for x, s, and the scaling point w. (The result below is referred to in the penultimate paragraph of Section 8 of [NT97].)

LEMMA 3.4. *Let $x \in \operatorname{int} K$ and $s \in \operatorname{int} K^*$, and define w as the unique point in $\operatorname{int} K$ such that $F''(w)x = s$. Then*

$$\sigma_s(-F'(x)) = \sigma_x(-F'_*(s)) = \sigma_x(w)^2 = \sigma_w(-F'(x))^2 = \sigma_s(-F'(w))^2 = \sigma_w(-F'_*(s))^2,$$

$$\sigma_s(x) = \sigma_x(s) = \sigma_w(x)^2 = \sigma_x(-F'(w))^2 = \sigma_w(s)^2 = \sigma_s(w)^2.$$

Proof. Let us prove the first sequence of identities. The equality of $\sigma_s(-F'(x))$ and $\sigma_x(-F'_*(s))$ follows from Lemma 3.3, as does that of $\sigma_x(w)$ and $\sigma_w(-F'(x))$ and that of $\sigma_s(-F'(w))$ and $\sigma_w(-F'_*(s))$. Moreover, the proof of $\sigma_x(w) = \sigma_s(-F'(w))$ is exactly analogous. It remains to show that $\sigma_x(-F'_*(s)) = \sigma_x(w)^2$. Let us write $z := -F'_*(s)$ and abbreviate $\sigma_x(w)$ to σ. Since x and w lie in $\operatorname{int} K$, σ is positive and finite, and

$$\sigma x - w \in \partial K.$$

We want to show that

$$\sigma^2 x - z = \sigma(\sigma x - w) + (\sigma w - z) \in \partial K$$

also.

We saw above that $\sigma = \sigma_w(z)$, so $\sigma w - z \in \partial K$ and thus $\sigma^2 x - z \in K$. We need to show that it lies in the boundary of K. Because $\sigma x - w \in \partial K$, there is some $u \in \partial K^*$ with $\langle u, \sigma x - w\rangle = 0$. Let $v := [F''(w)]^{-1}u \in \partial K$, so that

$$\langle F''(w)v, \sigma x - w\rangle = 0,$$

i.e., v and $\sigma x - w$ are orthogonal with respect to w (see Section 5 of [NT97]). We aim to show that v is also orthogonal to $\sigma w - z$ with respect to w, so that it is orthogonal to $\sigma^2 x - z$ and thus the latter lies in ∂K.

By L_* 5.1, v and $\sigma x - w$ are also orthogonal with respect to $w(\alpha) := w + \alpha(\sigma x - w)$ for any $\alpha \geq 0$. Hence we find $\langle F''(w(\alpha))(\sigma x - w), v\rangle = 0$ for all $\alpha \geq 0$, which implies

$$\langle F'(\sigma x) - F'(w), v\rangle = \langle \int_0^1 F''(w(\alpha))(\sigma x - w)d\alpha, v\rangle = 0.$$

This gives

$$\begin{aligned}\langle F''(w)v, \sigma w - z\rangle &= \langle F''(w)v, \sigma w + F'_*(s)\rangle \\ &= \langle F''(w)v, [F''(w)]^{-1}(F'(x) - \sigma F'(w))\rangle \\ &= \sigma\langle F'(\sigma x) - F'(w), v\rangle = 0,\end{aligned}$$

where the last equation used (2.7). Hence $\sigma^2 x - z = \sigma(\sigma x - w) + (\sigma w - z)$ is orthogonal to v with respect to w, and thus lies in ∂K as desired.

The proof of the second line of identities is similar. □

In the semidefinite programming case, the lemma says that the maximum eigenvalue of $s^{-1/2}x^{-1}s^{-1/2}$ is equal to that of $x^{-1/2}s^{-1}x^{-1/2}$, and that this is the square of the maximum eigenvalue of $x^{-1/2}wx^{-1/2} = (x^{1/2}sx^{1/2})^{-1/2}$.

3.2. Inequalities for norms. We have the following relationship between the two norms defined at a point.

PROPOSITION 3.5. *For $a \in E$ or E^* and for $b \in \operatorname{int} K$ or $\operatorname{int} K^*$ we have*

$$| a |_b \leq \| a \|_b \leq \nu^{1/2} | a |_b .$$

Proof. Since all the cases are similar, we assume $a \in E$ and $b \in \operatorname{int} K$. The first inequality follows from $(4.5)_*$. For the second, suppose that $b + \alpha a$ and $b - \alpha a \in K$ for some $\alpha > 0$. Then

$$\nu - \alpha^2 \| a \|_b^2 = \langle F''(b)b, b\rangle - \alpha^2\langle F''(b)a, a\rangle = \langle F''(b)(b - \alpha a), b + \alpha a\rangle \geq 0,$$

so that $\| a \|_b \leq \nu^{1/2}/\alpha$. Hence this holds for the supremum of all such α's, which gives the right-hand inequality. □

We now give a result that relates the norms at two neighboring points. (Compare with Theorem 3.8 below.) Note that the outer bounds in (3.9) and (3.10) are valid for general self-concordant functions under the much stronger assumption that $\| \tilde{x} - x \|_x \leq \delta < 1$ (see T^* 2.1.1).

THEOREM 3.6. *Suppose x and $\tilde{x}$ lie in $\operatorname{int} K$ with $| \tilde{x} - x |_x \leq \delta < 1$. Also, let $p := x - \tilde{x}$, so that $p_+ := \sigma_x(p) \leq \delta$ and $p_- := \sigma_x(-p) \leq \delta$. Then, for any $v \in E$ and $u \in E^*$, we have*

$$(1+\delta)^{-1} \| v \|_x \leq (1+p_-)^{-1} \| v \|_x \leq \| v \|_{\tilde{x}} \leq (1-p_+)^{-1} \| v \|_x \leq (1-\delta)^{-1} \| v \|_x, \tag{3.9}$$

$$(1-\delta) \| u \|_x \leq (1-p_+) \| u \|_x \leq \| u \|_{\tilde{x}} \leq (1+p_-) \| u \|_x \leq (1+\delta) \| u \|_x, \tag{3.10}$$

$$(1+\delta)^{-1} | v |_x \leq (1+p_-)^{-1} | v |_x \leq | v |_{\tilde{x}} \leq (1-p_+)^{-1} | v |_x \leq (1-\delta)^{-1} | v |_x, \tag{3.11}$$

$$(1-\delta) | u |_x \leq (1-p_-) | u |_x \leq | u |_{\tilde{x}} \leq (1+p_+) | u |_x \leq (1+\delta) | u |_x . \tag{3.12}$$

Proof. The inequalities in (3.9) and (3.10) follow directly from T_* 4.1. For (3.11), suppose first that $0 < \alpha < 1/\sigma_x(v)$. Then $x - \alpha v$ lies in K, and hence

$$(1-p_+)x - \alpha(1-p_+)v \in K.$$

Also,

$$-(1-p_+)x + \tilde{x} = p_+ x - (x - \tilde{x}) \in K,$$

so that $\tilde{x} - \alpha(1-p_+)v \in K$ and $1/\sigma_{\tilde{x}}(v) \geq \alpha(1-p_+)$. Since $0 < \alpha < 1/\sigma_x(v)$ was arbitrary, we have $\sigma_{\tilde{x}}(v) \leq (1-p_+)^{-1}\sigma_x(v)$, and because this holds also for $-v$, we conclude that

$$| v |_{\tilde{x}} \leq (1-p_+)^{-1} | v |_x .$$

Similarly, if $0 < \alpha < 1/\sigma_{\tilde{x}}(v)$, then $\tilde{x} - \alpha v$, and hence

$$(1 + p_-)^{-1}\tilde{x} - \alpha(1 + p_-)^{-1}v \in K.$$

Also,

$$(1 + p_-)x - \tilde{x} = p_- x + (x - \tilde{x}) \in K,$$

so $x - (1 + p_-)^{-1}\tilde{x} \in K$ and thus $x - \alpha(1 + p_-)^{-1}v \in K$. We conclude that $1/\sigma_x(v) \geq \alpha(1 + p_-)^{-1}$ and hence $\sigma_{\tilde{x}}(v) \geq (1 + p_-)^{-1}\sigma_x(v)$. Since this holds also for $-v$, we obtain

$$\mid v \mid_{\tilde{x}} \geq (1 + p_-)^{-1} \mid v \mid_x,$$

finishing the proof of (3.11).

Now (3.5) in Lemma 3.3 with $s := -F'(\tilde{x})$ shows that

$$\sigma_{\tilde{x}}(F'(x) - F'(\tilde{x})) = p_+, \quad \sigma_{\tilde{x}}(-F'(x) + F'(\tilde{x})) = p_-,$$

so (3.12) follows from (3.11) by replacing v, x, and $\tilde{x}$ with u, $-F'(\tilde{x})$, and $-F'(x)$ respectively. □

3.3. Properties of the third derivative. Let us establish now some properties of the third derivative of the self-scaled barrier. Let $x \in \operatorname{int} K$ and p_1, p_2, $p_3 \in E$. Then

$$F'''(x)[p_1, p_2, p_3] = \left(\langle F''(x + \alpha p_3)p_1, p_2\rangle\right)'_{\alpha=0}.$$

Note that the third derivative is a symmetric trilinear form. We denote by $F'''(x)[p_1, p_2]$ a vector from E^* such that

$$F'''(x)[p_1, p_2, p_3] = \langle F'''(x)[p_1, p_2], p_3\rangle$$

for any $p_3 \in E$. Similarly, $F'''(x)[p_1]$ is a linear operator from E to E^* such that

$$F'''(x)[p_1, p_2, p_3] = \langle F'''(x)[p_1]p_2, p_3\rangle$$

for any $p_2, p_3 \in E$.

In the semidefinite programming case, $F'''(x)[p_1, p_2]$ is the matrix $-x^{-1}p_1x^{-1}p_2x^{-1} - x^{-1}p_2x^{-1}p_1x^{-1}$.

The results below are based on the following important property.

Lemma 3.7. *For any $x \in \operatorname{int} K$ and $p \in E$*

$$\| F'''(x)[p] \|_x \leq 2 \mid p \mid_x . \tag{3.13}$$

Proof. Without loss of generality, we can assume that $\mid p \mid_x = 1$. Then $x + p \in K$ and $x - p \in K$.

Note that in view of C_* 3.2 (i), the operator $F'''(x)[v]$ is negative semidefinite for any $v \in K$. Therefore,

$$F'''(x)[p] = F'''(x)[(x + p) - x] = 2F''(x) + F'''(x)[x + p] \leq 2F''(x).$$

Similarly,

$$F'''(x)[p] = F'''(x)[x - (x - p)] = -2F''(x) - F'''(x)[x - p] \geq -2F''(x)$$

and (3.13) follows. □

Using the lemma above we can obtain some useful inequalities.

Theorem 3.8. *For any $x \in \operatorname{int} K$ and $p_1, p_2 \in E$ the following inequalities hold:*

$$\| F'''(x)[p_1, p_2] \|_x \leq 2 \mid p_1 \mid_x \cdot \| p_2 \|_x . \tag{3.14}$$

Moreover, if $v = x - p_1 \in \operatorname{int} K$ *then*

$$\| (F''(v) - F''(x))p_2 \|_x \leq \frac{2-\sigma_x(p_1)}{(1-\sigma_x(p_1))^2} \| p_1 \|_x \cdot | p_2 |_x . \tag{3.15}$$

Proof. The proof of (3.14) immediately follows from (3.13) and (3.1). Let us prove (3.15).
In view of (3.14), Theorem 3.6 and T_* 4.1 we have:

$$\| (F''(v) - F''(x))\, p_2 \|_x = \| \int_0^1 F'''(x-\tau p_1)[p_1,p_2] d\tau \|_x \leq \int_0^1 \| F'''(x-\tau p_1)[p_1,p_2] \|_x \, d\tau$$

$$\leq \int_0^1 \frac{\| F'''(x-\tau p_1)[p_1,p_2] \|_{x-\tau p_1}}{1-\tau\sigma_x(p_1)} d\tau \leq \int_0^1 \frac{2}{1-\tau\sigma_x(p_1)} \| p_1 \|_{x-\tau p_1} \cdot | p_2 |_{x-\tau p_1} \, d\tau$$

$$\leq \| p_1 \|_x \cdot | p_2 |_x \int_0^1 \frac{2 d\tau}{(1-\tau\sigma_x(p_1))^3} = \frac{2-\sigma_x(p_1)}{(1-\sigma_x(p_1))^2} \| p_1 \|_x \cdot | p_2 |_x . \qquad \square$$

Note that from (3.14) for any $p_3 \in E$ we have:

$$| F'''(x)[p_1,p_2,p_3] | \leq 2 \, | p_1 |_x \cdot \| p_2 \|_x \cdot \| p_3 \|_x . \tag{3.16}$$

This inequality highlights the difference between self-scaled barriers and arbitrary self-concordant functions; see D^* 2.1.1 and its seemingly more general consequence in T^* 2.1.1, which is (3.16) but with $| p_1 |_x$ replaced by $\| p_1 \|_x$.

We will need also a kind of "parallelogram rule" for the self-scaled barrier. Let us prove first two simple lemmas.

LEMMA 3.9. *Let* $x, v \in \operatorname{int} K$. *Then*

$$[F''(v)]^{-1} F'(x) = \tfrac{1}{2} [F''(x)]^{-1} F'''(x)[v,v]. \tag{3.17}$$

Proof. Let $s = -F'(v)$. Then, in view of Lemma 3.2(a) we have

$$\langle F''_*(s) F'(x), F'(x) \rangle \equiv \langle F''(x) F'_*(s), F'_*(s) \rangle.$$

By differentiating this identity with respect to x we get

$$2F''(x) F''_*(s) F'(x) \equiv F'''(x)[F'_*(s), F'_*(s)].$$

And that is exactly (3.17) since

$$F'_*(s) = -v, \quad F''_*(s) = [F''(v)]^{-1}$$

using (2.15) and (2.16). $\square$

LEMMA 3.10. *Let* $x \in \operatorname{int} K$ *and* $s \in \operatorname{int} K^*$. *Then*

$$F_*(s + \tau F'(x)) + F(x) \equiv F_*(s) + F(x + \tau F'_*(s)) \tag{3.18}$$

for any τ *such that* $s + \tau F'(x) \in \operatorname{int} K^*$.

Proof. We simply note that

$$s + \tau F'(x) = F''(w)(x + \tau F'_*(s)),$$

where $w \in \operatorname{int} K$ is such that $s = F''(w)x$. Therefore (3.18) immediately follows from (2.19). ∎
(We note that this lemma was proved for the universal barrier of K by Rothaus in Theorem 3.8 of [Ro60].)

Now we can prove the following important theorem.

THEOREM 3.11. *Let $x \in \operatorname{int} K$, $p \in E$, and the positive constants α and β be such that $x - \alpha p \in \operatorname{int} K$ and $x + \beta p \in \operatorname{int} K$. Then*

$$F(x-\alpha p) + F(x+\beta p) = F(x) + F\left(x + (\beta-\alpha)p + \tfrac{1}{2}\alpha\beta[F''(x)]^{-1}F'''(x)[p,p]\right).$$

Proof. Denote

$$v = x + (\beta-\alpha)p + \tfrac{1}{2}\alpha\beta[F''(x)]^{-1}F'''(x)[p,p], \quad x_1 = x - \alpha p, \quad x_2 = x + \beta p.$$

Then in view of (2.8) and Lemma 3.9

$$\begin{aligned} v &= x + (\beta-\alpha)p + \tfrac{\alpha}{2}[F''(x)]^{-1}F'''(x)[p, x_2 - x] \\ &= x_2 + \tfrac{\alpha}{2}[F''(x)]^{-1}F'''(x)[p, x_2] \\ &= x_2 + \tfrac{\alpha}{2\beta}[F''(x)]^{-1}F'''(x)[(x_2 - x, x_2] \\ &= \left(1 + \tfrac{\alpha}{\beta}\right) x_2 + \tfrac{\alpha}{2\beta}[F''(x)]^{-1}F'''(x)[x_2, x_2] \\ &= \left(1 + \tfrac{\alpha}{\beta}\right) x_2 + \tfrac{\alpha}{\beta}[F''(x_2)]^{-1}F'(x) \\ &= [F''(x_2)]^{-1}\left[-\left(1 + \tfrac{\alpha}{\beta}\right)F'(x_2) + \tfrac{\alpha}{\beta}F'(x)\right] \end{aligned}$$

Now, let $s = -F'(x_2)$. Then, in view of (2.20), (2.6) and Lemma 3.10 we have:

$$\begin{aligned} F(v) &= F\left([F''(x_2)]^{-1}\left[\left(1 + \tfrac{\alpha}{\beta}\right)s + \tfrac{\alpha}{\beta}F'(x)\right]\right) \\ &= F_*\left(\left(1 + \tfrac{\alpha}{\beta}\right)s + \tfrac{\alpha}{\beta}F'(x)\right) + 2F(x_2) + \nu \\ &= F_*\left(s + \tfrac{\alpha}{\alpha+\beta}F'(x)\right) + 2F(x_2) + \nu + \nu\ln\tfrac{\beta}{\alpha+\beta} \\ &= -F(x) + F_*(s) + F\left(x - \tfrac{\alpha}{\alpha+\beta}x_2\right) + 2F(x_2) + \nu + \nu\ln\tfrac{\beta}{\alpha+\beta} \\ &= -F(x) + F\left(\tfrac{\beta x - \alpha\beta p}{\alpha+\beta}\right) + F(x_2) + \nu\ln\tfrac{\beta}{\alpha+\beta} \\ &= -F(x) + F(x_1) + F(x_2). \end{aligned}$$

Of course, here we have assumed that the appropriate points lie in the domains of F and F_*. But $x_1 = x - \alpha p \in \operatorname{int} K$ implies

$$x - \frac{\alpha}{\alpha+\beta}x_2 = \frac{\beta x - \alpha\beta p}{\alpha+\beta} \in \operatorname{int} K.$$

Hence, using w with $F''(w)x = s$, we find

$$s + \frac{\alpha}{\alpha+\beta}F'(x) \text{ and hence } (1 + \frac{\alpha}{\beta})s + \frac{\alpha}{\beta}F'(x) \in \operatorname{int} K^*.$$

This shows $v \in \operatorname{int} K$, and hence all the terms above are well defined. ∎

3.4. Some properties of the scaling point. In this section we prove some relations for $x \in \operatorname{int} K$, $s \in \operatorname{int} K^*$, and the scaling point $w \in \operatorname{int} K$ such that $F''(w)x = s$.

Denote by $\mu(x,s)$ the *normalized duality gap*:

$$\mu(x,s) := \frac{1}{\nu}\langle s, x\rangle.$$

LEMMA 3.12. *Let $\bar{w} := \sqrt{\mu(x,s)}w$, $\bar{x} := x/\sqrt{\mu(x,s)}$. Then*

$$2(F(x) - F(\bar{w})) \equiv F(x) + F_*(s) + \nu\ln\mu(x,s) + \nu \geq 0, \tag{3.19}$$

$$2[F(x) - F(\bar{w})] \geq \| \bar{x} - w \|_w^2 = \| \bar{w} - x \|_{\bar{w}}^2, \tag{3.20}$$

$$2[F(x) - F(\bar{w})] \leq [\mu(x,s)\langle F'(x), F'_*(s)\rangle - \nu] - \| \bar{w} - x \|_x^2 . \tag{3.21}$$

Proof. Denote $\mu = \mu(x,s)$. Then,

$$F(\bar{w}) = F(w) - \tfrac{1}{2}\nu \ln \mu = \tfrac{1}{2}[F(x) - F_*(s) - \nu - \nu \ln \mu].$$

(We have used (2.6) and (2.19).) Therefore, in view of (2.17) we obtain:

$$2(F(x) - F(\bar{w})) = F(x) + F_*(s) + \nu \ln \mu + \nu \geq 0.$$

Further, from (2.9) and the convexity of F, we have:

$$\| \bar{x} - w \|_w^2 = \frac{1}{\mu}\langle s, x\rangle + \frac{2}{\sqrt{\mu}}\langle F'(w), x\rangle + \nu$$

$$= 2[\nu + \langle F'(\bar{w}), x\rangle] = 2\langle F'(\bar{w}), x - \bar{w}\rangle \leq 2(F(x) - F(\bar{w})),$$

which gives (3.20); the equation is trivial by (2.7).

Finally, note that in view of Lemma 3.2(b), we have:

$$\begin{array}{rcl} \| \bar{w} - x \|_x^2 & = & \mu\langle F''(x)w, w\rangle + 2\langle F'(x), \bar{w}\rangle + \nu \\ & = & \mu\langle F'(x), [F''(w)]^{-1}F'(x)\rangle + 2\langle F'(x), \bar{w}\rangle + \nu \\ & = & [\mu\langle F'(x), F'_*(s)\rangle - \nu] + 2(\langle F'(x), \bar{w}\rangle + \nu). \end{array}$$

Therefore, using the convexity of $F(x)$, we get

$$2(F(x) - F(\bar{w})) \leq 2\langle F'(x), x - \bar{w}\rangle = -2(\langle F'(x), \bar{w}\rangle + \nu),$$

and (3.21) follows. ∎

Let $t := -F'(w)$, so that $F''_*(t) = [F''(w)]^{-1}$ takes s into x. We next relate $F''(w)$ to $F''(x)$ and $F''_*(t)$ to $F''_*(s)$:

Lemma 3.13. *Suppose we have $\delta := | s/\mu + F'(x) |_x < 1$ for some $\mu > 0$. Then*

$$\begin{array}{rcccl} (1-\delta)F''(x) & \leq & F''(w)/\mu & \leq & (1+\delta)F''(x), \\ (1-\delta)F''_*(s) & \leq & F''_*(t)/\mu & \leq & (1+\delta)F''_*(s). \end{array} \tag{3.22}$$

Proof. Let $v := -F'_*(s)$ and $u := -F'(x)$. We will use C_* 4.1 to prove our result, so we need bounds on $\sigma_v(x)$, $\sigma_x(v)$, $\sigma_u(s)$, and $\sigma_s(u)$. Since x, $v \in \operatorname{int} K$ and s, $u \in \operatorname{int} K^*$, these are all positive, and we have using Lemma 3.3

$$\begin{aligned} [\sigma_v(x)]^{-1} &= [\sigma_u(s)]^{-1} \\ &= \sup\{\alpha : -F'(x) - \alpha s \in K^*\} \\ &= \sup\{\alpha : (1-\alpha\mu)(-F'(x)) - \alpha\mu(s/\mu + F'(x)) \in K^*\}. \end{aligned}$$

Now for $\alpha < [\mu(1+\delta)]^{-1}$, $\alpha\mu < 1$ and $[\alpha\mu/(1-\alpha\mu)]\delta < 1$, so using $\delta := | s/\mu + F'(x) |_x$ we have $-F'(x) - [\alpha\mu/(1-\alpha\mu)](s/\mu + F'(x)) \in K^*$. Thus $[\sigma_v(x)]^{-1} \geq [\mu(1+\delta)]^{-1}$, or

$$\sigma_v(x) = \sigma_u(s) \leq \mu(1+\delta). \tag{3.23}$$

Similarly,

$$
\begin{aligned}
[\sigma_x(v)]^{-1} &= [\sigma_s(u)]^{-1} \\
&= \sup\{\alpha : s + \alpha F'(x) \in K\} \\
&= \sup\{\alpha : (\mu - \alpha)(-F'(x)) + \mu(s/\mu + F'(x)) \in K^*\}.
\end{aligned}
$$

Now for $\alpha < \mu(1-\delta)$, $\alpha < \mu$ and $[\mu/(\mu-\alpha)]\delta < 1$, so using $\delta := \mid s/\mu + F'(x) \mid_x$ again we have $-F'(x) + [\mu/(\mu-\alpha)](s/\mu + F'(x)) \in K^*$. Thus $[\sigma_x(v)]^{-1} \geq \mu(1-\delta)$, or

$$
\sigma_x(v) = \sigma_s(u) \leq [\mu(1-\delta)]^{-1}. \tag{3.24}
$$

Then (3.23) and (3.24) imply (3.22) by C_* 4.1(iii). □

Corollary 3.14. *Under the conditions of the lemma, for all $v \in E$, $u \in E^*$, we have*

$$
\| [F''(w)/\mu - F''(x)]v \|_x \leq \delta \| v \|_x, \tag{3.25}
$$

$$
\| [F''_*(t)/\mu - F''_*(s)]u \|_s \leq \delta \| u \|_s . \tag{3.26}
$$

Proof. This follows from the lemma and (3.2). □

4. Primal-Dual Central Path and Proximity Measures. In this section we consider *general* primal-dual proximity measures. These measures estimate the distance from a primal-dual pair $(x, s) \in \operatorname{int} K \times \operatorname{int} K^*$ to the nonlinear surface of "analytic centers"

$$
\mathcal{A} := \{(x, s) \in \operatorname{int} K \times \operatorname{int} K^* : \ s = -\mu F'(x), \ \mu > 0\}.
$$

(The intersection of the Cartesian product of this surface with Y and the strictly feasible set of the primal-dual problem (2.1)-(2.3), $S^0(P) \times S^0(D)$, is called the *primal-dual central path.*) Note that these general proximity measures do not involve at all the specific information about our problem (2.1), namely the objective vector c, the operator A and the right-hand-side b.

Let us introduce first the notion of the central path and give the main properties of this trajectory. In what follows we assume that Assumptions (2.4) and (2.5) are satisfied.

Define the *primal central path* $\{x(\mu) : \mu > 0\}$ for the problem (2.1) as follows:

$$
x(\mu) = \arg\min_x \left\{ \frac{1}{\mu}\langle c, x\rangle + F(x) : Ax = b, \ x \in K \right\}. \tag{4.1}
$$

The *dual central path* $\{(y(\mu), s(\mu)) : \mu > 0\}$ for the dual problem (2.3) is defined in a symmetric way:

$$
(y(\mu), s(\mu)) = \arg\min_{y,s} \left\{ -\frac{1}{\mu}\langle b, y\rangle + F_*(s) : A^*y + s = c, \ y \in Y, \ s \in K^* \right\}. \tag{4.2}
$$

The *primal-dual central* path for the primal-dual problem

$$
\begin{array}{ll}
\min & \langle c, x\rangle - \langle b, y\rangle \\
\text{s.t.} & Ax = b, \\
& A^*y + s = c, \\
& x \in K, \ y \in Y, \ s \in K^*,
\end{array} \tag{4.3}
$$

is simply the combination of these paths:

$$
\{v(\mu) = (x(\mu), y(\mu), s(\mu)) : \mu > 0\}.
$$

Let us present now the duality theorem (see [Ne96]).

THEOREM 4.1.
a) For any $\xi > 0$ the set

$$Q(\xi) = \{v = (x, y, s) : Ax = b,\ A^*y + s = c,\ \langle c, x\rangle - \langle b, y\rangle = \xi,\ x \in K,\ y \in Y,\ s \in K^*\}$$

is nonempty and bounded. The optimal set of the problem (4.3) is also nonempty and bounded.
b) The optimal value of the problem (4.3) is zero.
c) The points of the primal-dual central path are well defined and for any $\mu > 0$ the following relations hold:

$$\langle c, x(\mu)\rangle - \langle b, y(\mu)\rangle = \langle s(\mu), x(\mu)\rangle = \nu\mu, \tag{4.4}$$

$$F(x(\mu)) + F_*(s(\mu)) = -\nu - \nu \ln \mu, \tag{4.5}$$

$$s(\mu) = -\mu F'(x(\mu)), \tag{4.6}$$

$$x(\mu) = -\mu F'_*(s(\mu)). \qquad \square \tag{4.7}$$

For linear programming, where K is the nonnegative orthant in $E := R^n$, (4.6) and (4.7) state that the componentwise product of x and s is μ times the vector of ones in R^n. For semidefinite programming, where K is the cone of positive semidefinite matrices in the space E of symmetric matrices of order n, they state that the matrix product of x and s is μ times the identity matrix of order n.

The relations of part (c) of the theorem explain the construction of all general primal-dual proximity measures: most of them measure in a specific way the residual

$$\frac{1}{\mu}s + F'(x)$$

with $\mu = \mu(x, s)$, where

$$\mu(x, s) := \frac{1}{\nu}\langle s, x\rangle.$$

In our paper we will consider two groups of primal-dual proximity measures. The first group consists of the following *global* proximity measures:

- the *functional* proximity measure

$$\gamma_F(x, s) := F(x) + F_*(s) + \nu \ln \mu(x, s) + \nu, \tag{4.8}$$

- the *gradient* proximity measure

$$\gamma_G(x, s) := \mu(x, s)\langle F'(x), F'_*(s)\rangle - \nu, \tag{4.9}$$

- the *uniform* proximity measure

$$\gamma_\infty(x, s) := \mu(x, s)\sigma_s(-F'(x)) - 1 = \mu(x, s)\sigma_x(-F'_*(s)) - 1 \tag{4.10}$$

(see Lemma 3.3).

The second group is comprised of the *local* proximity measures:

$$\lambda_\infty(x, s) := \left| \frac{1}{\mu(x, s)}s + F'(x) \right|_x = \left| \frac{1}{\mu(x, s)}x + F'_*(s) \right|_s \tag{4.11}$$

(see Lemma 3.3),

$$\lambda_\infty^+(x,s) := \frac{\sigma_s(x)}{\mu(x,s)} - 1 = \frac{\sigma_x(s)}{\mu(x,s)} - 1, \tag{4.12}$$

(see Lemma 3.3 and compare with (4.10)),

$$\lambda_2(x,s) := \parallel \frac{1}{\mu(x,s)} s + F'(x) \parallel_x = \parallel \frac{1}{\mu(x,s)} x + F_*'(s) \parallel_s \tag{4.13}$$

(see Lemma 3.2(a)), and

$$\lambda(x,s) := \left[\nu - \frac{\nu^2\mu^2(x,s)}{\parallel s \parallel_x^2}\right]^{1/2} = \left[\nu - \frac{\nu^2\mu^2(x,s)}{\parallel x \parallel_s^2}\right]^{1/2} \tag{4.14}$$

(see Lemma 3.2(a)). The proof of Theorem 4.2 below shows that $\lambda(x,s)$ is well-defined. We will see later that the value of any global proximity measure can be used as an estimate for the distance to the primal-dual central path at any strictly feasible primal-dual point. The value of a local proximity measure has a sense only in a small neighborhood of the central path.

For comparison with the standard proximity measures developed in linear and semidefinite programming, let us present the specific form of the above proximity measures for the cases $K = R^n_+$ and $K = S^n_+$, where i denotes the identity matrix:

$$\begin{array}{rcll}
\gamma_F(x,s) & = & -\sum\limits_{j=1}^n \ln\left(x^{(j)}s^{(j)}\right) + n\ln\left(\frac{x^Ts}{n}\right), & -\ln\det x - \ln\det s + n\ln\frac{\mathrm{Trace}\,(x^{1/2}sx^{1/2})}{n}, \\
\gamma_G(x,s) & = & \frac{x^Ts}{n}\sum\limits_{j=1}^n \frac{1}{x^{(j)}s^{(j)}} - n, & \frac{\mathrm{Trace}\,(x^{1/2}sx^{1/2})\mathrm{Trace}\,(x^{-1/2}s^{-1}x^{-1/2})}{n} - n, \\
\gamma_\infty(x,s) & = & \frac{x^Ts}{n\min\limits_{1\le j\le n} x^{(j)}s^{(j)}} - 1, & \frac{\mathrm{Trace}\,(x^{1/2}sx^{1/2})}{n\chi_{\min}(x^{1/2}sx^{1/2})} - 1, \\
\lambda_\infty(x,s) & = & \parallel \frac{n}{x^Ts}Xs - e \parallel_\infty, & \parallel \frac{x^{1/2}sx^{1/2}}{\mathrm{Trace}\,(x^{1/2}sx^{1/2})/n} - i \parallel_2, \\
\lambda_\infty^+(x,s) & = & \frac{n}{x^Ts}\max\limits_{1\le j\le n} x^{(j)}s^{(j)} - 1, & \chi_{\max}\left(\frac{x^{1/2}sx^{1/2}}{\mathrm{Trace}\,(x^{1/2}sx^{1/2})/n} - i\right), \\
\lambda_2(x,s) & = & \parallel \frac{n}{x^Ts}Xs - e \parallel_2, & \parallel \frac{x^{1/2}sx^{1/2}}{\mathrm{Trace}\,(x^{1/2}sx^{1/2})/n} - i \parallel_F, \\
\lambda(x,s) & = & \left[n - \frac{\left(\sum\limits_{j=1}^n x^{(j)}s^{(j)}\right)^2}{\sum\limits_{j=1}^n (x^{(j)}s^{(j)})^2}\right]^{1/2}, & \left[n - \frac{[\mathrm{Trace}\,(x^{1/2}sx^{1/2})]^2}{\|x^{1/2}sx^{1/2}\|_F^2}\right]^{1/2}.
\end{array}$$

Thus level sets of λ_2, λ_∞, and γ_∞ correspond respectively to the 2-norm, infinity-norm, and "minus-infinity-norm" neighborhoods of the central path used in interior-point methods for linear programming. Also, γ_F is the Tanabe-Todd-Ye primal-dual potential function with parameter n. (We have written these measures for semidefinite programming in a symmetric way that stresses the similarity to linear programming. However, simpler formulae should be used for computation; for example, $\mathrm{Trace}\,(x^{1/2}sx^{1/2}) = \mathrm{Trace}\,(xs)$, and the eigenvalues of $x^{1/2}sx^{1/2}$ are the same as those of l^Tsl, where ll^T is the Cholesky factorization of x.)

Now we return to the general case.

REMARK 4.1. *1. Note that the gradient proximity measure can be written as a squared norm of the vector* $\frac{1}{\mu(x,s)}s + F'(x)$. *Indeed, let us choose* $w \in \mathrm{int}\,K$ *such that* $s = F''(w)x$ *(this is possible in view of* T_* *3.2). Then* $F'(x) = F''(w)F_*'(s)$. *Therefore (writing* μ *for* $\mu(x,s)$ *for simplicity)*

$$\langle \frac{1}{\mu}s + F'(x), [F''(w)]^{-1}\left(\frac{1}{\mu}s + F'(x)\right)\rangle$$

$$= \frac{1}{\mu^2}\langle s, [F''(w)]^{-1}s\rangle + \frac{2}{\mu}\langle F'(x), [F''(w)]^{-1}s\rangle + \langle F'(x), [F''(w)]^{-1}F'(x)\rangle$$

$$= \frac{\nu}{\mu} - \frac{2\nu}{\mu} + \langle F'(x), F'_*(s) \rangle = \frac{\gamma_G(x,s)}{\mu}$$

Thus,

$$\gamma_G(x,s) = \mu(x,s) \parallel \frac{1}{\mu(x,s)} s + F'(x) \parallel_w^2 = \mu(x,s) \parallel \frac{1}{\mu(x,s)} x + F'_*(s) \parallel_w^2 . \tag{4.15}$$

2. There is an interesting relation between λ_∞*,* λ_∞^+ *and* γ_∞*:*

$$\lambda_\infty(x,s) = \max\left\{ \lambda_\infty^+(x,s), \frac{\gamma_\infty(x,s)}{1+\gamma_\infty(x,s)} \right\}.$$

Note that all the proximity measures we have introduced are homogeneous of degree zero (indeed, separately in x and s). Let us prove some relations between these measures.

THEOREM 4.2. *Let* $x \in \operatorname{int} K$ *and* $s \in \operatorname{int} K^*$*. Then*

$$\lambda_\infty^+(x,s) \le \lambda_\infty(x,s) \le \lambda_2(x,s) \le \nu^{1/2}\lambda_\infty(x,s), \tag{4.16}$$

$$\gamma_F(x,s) \le \nu \ln\left(1 + \frac{1}{\nu}\gamma_G(x,s)\right) \le \gamma_G(x,s), \tag{4.17}$$

$$\frac{\gamma_\infty^2(x,s)}{1+\gamma_\infty(x,s)} \le \gamma_G(x,s) \le \nu\gamma_\infty(x,s), \tag{4.18}$$

$$\lambda_2(x,s) - \ln(1+\lambda_2(x,s)) \le \gamma_F(x,s), \tag{4.19}$$

$$\lambda(x,s) \le \lambda_2(x,s), \tag{4.20}$$

$$\gamma_G(x,s) \le \lambda_2^2(x,s)(1+\gamma_\infty(x,s)). \tag{4.21}$$

Moreover, if $\lambda_\infty(x,s) < 1$ *then*

$$\gamma_\infty(x,s) \le \frac{\lambda_\infty(x,s)}{1-\lambda_\infty(x,s)} \tag{4.22}$$

and therefore

$$\gamma_F(x,s) \le \gamma_G(x,s) \le \frac{\lambda_2^2(x,s)}{1-\lambda_\infty(x,s)}. \tag{4.23}$$

Proof. The fact that $\lambda_\infty(x,s) \le \lambda_2(x,s) \le \nu^{1/2}\lambda_\infty(x,s)$ follows from Proposition 3.5. In order to prove the inequality $\lambda_\infty^+(x,s) \le \lambda_\infty(x,s)$ we simply note that

$$\sigma_x(s/\mu(x,s)) = 1 + \sigma_x(s/\mu(x,s) + F'(x)).$$

Let us prove inequality (4.17). We have:

$$\begin{array}{rcll}
\gamma_F(s,x) & = & F(x) + F_*(s) + \nu\ln\mu(x,s) + \nu & \text{(in view of (4.8))} \\
 & = & -F_*(-F'(x)) - F(-F'_*(s)) + \nu\ln\mu(x,s) - \nu & \text{(using (2.13) and (2.14))} \\
 & \le & \nu\ln\langle F'(x), F'_*(s)\rangle + \nu\ln\mu(x,s) - \nu\ln\nu & \text{(see (2.17))} \\
 & = & \nu\ln\left(1 + \frac{1}{\nu}\gamma_G(x,s)\right) & \text{(see (4.9))}
\end{array}$$

and we get (4.17).

Next observe that, since all our proximity measures are homogeneous, we can assume for the rest of the proof that

$$\mu(x,s) = 1. \tag{4.24}$$

Thus, using (4.24) and $(4.4)_*$, we get

$$\gamma_G(x,s) = \langle F'(x), F'_*(s)\rangle - \nu \le \nu(\sigma_x(-F'_*(s)) - 1) = \nu\gamma_\infty(x,s),$$

which is exactly the right-hand inequality of (4.18). To prove the left-hand inequality, note that

$$\begin{array}{rcll} \gamma_G(x,s) & = & \parallel x + F'_*(s) \parallel_w^2 & \text{(by (4.15))} \\ & \ge & \parallel x + F'_*(s) \parallel_x^2 / \sigma_x^2(w) & \text{(by } C_* \text{ 4.1)} \\ & \ge & \mid x + F'_*(s) \mid_x^2 / \sigma_x(-F'_*(s)) & \text{(by Lemma 3.4 and Proposition 3.5)} \\ & \ge & \mid x + F'_*(s) \mid_x^2 / (1+\gamma_\infty(x,s)) & \text{(by the definition of } \gamma_\infty(x,s)) \\ & \ge & \sigma_x^2(-x - F'_*(s))/(1+\gamma_\infty(x,s)) & \\ & = & \gamma_\infty^2(x,s)/(1+\gamma_\infty(x,s)). & \end{array}$$

Note that the functional proximity measure is nonnegative in view of (2.17). Let us fix s and consider the function

$$\phi(v) := F(v) + F_*(s) + \langle s, v - x\rangle + \nu.$$

Note that $\phi(v) \ge \gamma_F(v,s)$ for any $v \in \operatorname{int} K$ and $\phi(x) = \gamma_F(x,s)$. Note also that $\phi(v)$ is a strictly self-concordant function. Therefore, from P^* 2.2.2, we have:

$$\phi(\bar v) \le \phi(x) - [\lambda_\phi(x) - \ln(1 + \lambda_\phi(x))],$$

where

$$\bar v := x - \frac{1}{\lambda_\phi(x)}[\phi''(x)]^{-1}\phi'(x),$$

and $\lambda_\phi(x) := \langle \phi'(x), [\phi''(x)]^{-1}\phi'(x)\rangle^{1/2}$. Since $\phi(\bar v) \ge \gamma_F(\bar v, s) \ge 0$ and $\lambda_\phi(x) = \lambda_2(x,s)$, we get (4.19).

Further,

$$\lambda_2^2(x,s) = \langle s + F'(x), [F''(x)]^{-1}(s + F'(x))\rangle = \langle s, [F''(x)]^{-1}s\rangle - \nu$$

(see (2.8) and (4.24)). Therefore $\lambda(x,s)$ is well-defined and we have

$$\lambda_2^2(x,s) = \frac{\nu\lambda^2(x,s)}{\nu - \lambda^2(x,s)} \ge \lambda^2(x,s) \tag{4.25}$$

which gives (4.20).

Let us now prove inequality (4.21). Let w be such that $s = F''(w)x$. In view of (4.15) and (4.24)

$$\gamma_G(x,s) = \langle s + F'(x), [F''(w)]^{-1}(s + F'(x))\rangle. \tag{4.26}$$

Using inequality $(4.7)_*$ in C_* 4.1, we have

$$F''(w) \ge \frac{1}{\sigma_x(-F'_*(s))} F''(x). \tag{4.27}$$

Note that $\sigma_x(-F'_*(s)) = 1 + \gamma_\infty(x,s)$. Therefore, combining (4.26), (4.27) and (4.13) we come to (4.21).

Let $\lambda = \lambda_\infty(x,s) < 1$. This implies that for $p = s + F'(x)$ we have $-\lambda F'(x) + p \in K^*$. Hence

$$s + (1-\lambda)F'(x) \in K^*.$$

This means that

$$\frac{1}{\sigma_s(-F'(x))} \geq 1 - \lambda.$$

Therefore

$$\gamma_\infty(x,s) = \sigma_s(-F'(x)) - 1 \leq \frac{\lambda}{1-\lambda}$$

and inequality (4.22) is proved.

The remaining inequality (4.23) is a direct consequence of (4.17), (4.21) and (4.22). □

From (2.17) we know that $\gamma_F(x,s)$ is nonnegative, and $\lambda(x,s)$ is nonnegative by definition. Also, if $\sigma := \sigma_x(s)$, then $-\sigma F'(x) - s \in K^*$, so $\nu\sigma - \langle s, x\rangle = \langle -\sigma F'(x) - s, x\rangle \geq 0$, which implies that $\lambda_\infty^+ = \sigma/\mu(x,s) - 1 \geq 0$. Hence the theorem shows that all the measures are nonnegative. Now suppose that $s = -\mu(x,s)F'(x)$ (or equivalently that $x = -\mu(x,s)F_*'(s)$). Then $\gamma_\infty(x,s)$ and $\lambda_\infty(x,s)$ are zero, and hence all the measures are zero. Indeed, if any one of the measures is zero, we have $s = -\mu(x,s)F'(x)$ and hence all of the measures are zero. This follows from the conditions for equality in (2.17) for $\gamma_F(x,s)$, and hence for all the global measures. It holds trivially for all the local measures except for $\lambda_\infty^+(x,s)$ and $\lambda(x,s)$. If $\lambda_\infty^+(x,s) = 0$, then as above $\sigma := \sigma_x(s) = \mu(x,s)$ and $\langle -\sigma F'(x) - s, x\rangle = 0$. But since $x \in \operatorname{int} K$ and $-\sigma F'(x) - s \in K^*$, we find $s + \mu(x,s)F'(x) = s + \sigma F'(x) = 0$. If $\lambda(x,s)$ is zero, so is $\lambda_2(x,s)$ by (4.25).

Observe that (4.18) provides a very simple proof of T_* 5.2. Indeed, in our present notation, we need to show that (using Lemma 3.4)

$$\gamma_G(x,s) \geq \frac{3}{4}\mu(x,s)\sigma_x^2(w) - 1 = \frac{3}{4}\mu(x,s)\sigma_x(-F_*'(s)) - 1 = \frac{3}{4}\gamma_\infty(x,s) - \frac{1}{4}. \tag{4.28}$$

But this follows directly from (4.18) since $\gamma^2/(1+\gamma) \geq (3\gamma - 1)/4$ for any $\gamma \geq 0$.

Note that all global proximity measures are unbounded: they tend to $+\infty$ when x or s approaches a nonzero point in the boundary of the corresponding cone. This is immediate for γ_F, and hence follows from the theorem for the other measures. All local proximity measures are bounded. Indeed, $v := [F''(x)]^{-1}s/\parallel s \parallel_x$ has $\parallel v \parallel_x = 1$, so $x - v \in K$ and $\langle s, x - v\rangle \geq 0$. This proves that $\langle s, x\rangle \geq \parallel s \parallel_x$ and hence $0 \leq \lambda(x,s)^2 \leq \nu - 1$. This bounds $\lambda_2(x,s)$ using (4.25) and hence all the local proximity measures. It can be also shown that no contour of any local proximity measure defined by its maximum value corresponds to the boundary of the primal-dual cone; thus, no such measure can be transformed to a global proximity measure.

As we have seen, one of the key points in forming a primal-dual proximity measure is to decide what is the reference point on the central path. This decision results in the choice of the parameter μ involved in that measure. It is clear that any proximity measure leads to a specific rule for defining the optimal reference point. Indeed, we can consider a primal-dual proximity measure as a function of x, s and μ. Therefore, the optimal choice for μ is given by minimizing the value of this measure with respect to μ when x and s are fixed. Let us demonstrate how our choice $\mu = \mu(x,s)$ arises in this way for the functional proximity measure.

Consider the penalty function

$$\psi_\mu(v) := \psi_\mu(x,y,s) = \frac{1}{\mu}[\langle c, x\rangle - \langle b, y\rangle] + F(x) + F_*(s).$$

By definition, for any $\mu > 0$ the point $v(\mu)$ is a minimizer of this function over the set

$$Q = \{v = (x,y,s) : Ax = b,\ A^*y + s = c,\ x \in K,\ y \in Y,\ s \in K^*\}.$$

Note that from Theorem 4.1 we know exactly what the minimal value of $\psi_\mu(v)$ over Q is:

$$\psi_\mu^* = \psi_\mu(v(\mu)) = \frac{1}{\mu}[\langle c, x(\mu)\rangle - \langle b, y(\mu)\rangle] + F(x(\mu)) + F_*(s(\mu))$$

$$= \nu - \nu - \nu \ln \mu = -\nu \ln \mu.$$

Thus, we can use the function

$$\gamma_\mu(v) = \psi_\mu(v) + \nu \ln \mu$$

as a kind of functional proximity measure.

Note that the function $\gamma_\mu(v)$ can estimate the distance from a given point v to the point $v(\mu)$ for any value of μ. But what is the point of the primal-dual central path that is closest to v? To answer this question we have to minimize $\gamma_\mu(v)$ as a function of μ. Since

$$\gamma_\mu(v) = \frac{1}{\mu}[\langle c, x\rangle - \langle b, y\rangle] + F(x) + F_*(s) + \nu \ln \mu,$$

we easily get that the optimal value of μ is

$$\mu^* = \frac{1}{\nu}[\langle c, x\rangle - \langle b, y\rangle] = \frac{1}{\nu}\langle s, x\rangle = \mu(x, s).$$

It is easy to see that $\gamma_{\mu(x,s)}(v) = \gamma_F(x, s)$.

The main advantage of the choice $\mu = \mu(x, s)$ is that the restriction of $\mu(x, s)$ to the feasible set is a linear function. But this choice is not optimal for other proximity measures. For example, for the local proximity measure λ_2, let

$$\bar{\lambda}_2(x, s, \mu) :=\| \frac{1}{\mu}s + F'(x) \|_x, \tag{4.29}$$

so that $\lambda_2(x, s) = \bar{\lambda}_2(x, s, \mu(x, s))$. Then the optimal choice of μ for $\bar{\lambda}_2$ is not $\mu(x, s)$ but

$$\bar{\mu} = \frac{\| s \|_x^2}{\langle s, x\rangle}.$$

This leads to the following relation:

$$\bar{\lambda}_2(x, s, \bar{\mu}) = \left[\nu - \frac{\langle s, x\rangle^2}{\| s \|_x^2}\right]^{1/2} \equiv \lambda(x, s). \tag{4.30}$$

The following lemma will be useful below:

LEMMA 4.3. *Let $x \in \operatorname{int} K$, $s \in \operatorname{int} K^*$, and $w \in \operatorname{int} K$ be such that $F''(w)x = s$. Then for any $v \in E$, $u \in E^*$, we have*

$$\| v \|_x^2 \leq \frac{1}{\mu(x, s)}[1 + \gamma_\infty(x, s)] \| v \|_w^2,$$
$$\| u \|_s^2 \leq \frac{1}{\mu(x, s)}[1 + \gamma_\infty(x, s)] \| u \|_w^2 .$$

Proof. By $(4.7)_*$ in C_* 4.1,

$$\frac{1}{\sigma_x(-F_*'(s))} F''(x) \leq F''(w),$$

and similarly

$$\frac{1}{\sigma_s(-F'(x))} F_*''(s) \leq [F''(w)]^{-1}.$$

The result follows since $\sigma_x(-F_*'(s)) = \sigma_s(-F'(x)) = [1 + \gamma_\infty(x, s)]/\mu(x, s)$ by definition of $\gamma_\infty(x, s)$. ∎

5. Affine scaling and centering directions. In the next sections we will consider several primal-dual interior point methods, which are based on two directions: the *affine-scaling* direction and the *centering* direction. Let us present the main properties of these directions.

5.1. The affine-scaling direction. We start from the definition. Let us fix points $x \in S^0(P)$ and $(y, s) \in S^0(D)$, and let $w \in \operatorname{int} K$ be such that $s = F''(w)x$.

DEFINITION 5.1. *The affine-scaling direction for the primal-dual point (x, y, s) is the solution (p_x, p_y, p_s) of the following linear system:*

$$\begin{array}{rcrcrcl} & & F''(w)p_x & + & p_s & = & s, \\ & & Ap_x & & & = & 0, \\ A^* p_y & & & + & p_s & = & 0. \end{array} \tag{5.1}$$

Note that the first equation of (5.1) can be rewritten as follows:

$$p_x + [F''(w)]^{-1} p_s = x. \tag{5.2}$$

If $t := -F'(w) \in \operatorname{int} K^*$, then $F''_*(t)s = x$ and $F''_*(t) = [F''(w)]^{-1}$, so the affine-scaling direction is defined symmetrically. It is also clear from the other two equations that

$$\langle p_s, p_x \rangle = 0. \tag{5.3}$$

For K the nonnegative orthant in R^n this definition gives the usual primal-dual affine-scaling direction. For K the cone of positive semidefinite matrices of order n, efficient methods of computing w and the affine-scaling direction (and also the centering direction of the next subsection) are given in [TTT96].

The affine-scaling direction has several important properties.

LEMMA 5.2. *Let (p_x, p_y, p_s) be the affine-scaling direction for the strictly feasible point (x, y, s). Then*

$$\langle s, p_x \rangle + \langle p_s, x \rangle = \langle s, x \rangle, \tag{5.4}$$

$$\langle F'(x), p_x \rangle + \langle p_s, F'_*(s) \rangle = -\nu, \tag{5.5}$$

$$\| p_x \|_w^2 + \| p_s \|_w^2 = \langle s, x \rangle, \tag{5.6}$$

$$\bar{\sigma}(x, s) := \max\{| p_x |_x, | p_s |_s\} \geq 1. \tag{5.7}$$

Proof. Indeed, multiplying the first equation of (5.1) by x and using the definition of w we get:

$$\langle s, x \rangle = \langle F''(w)p_x, x \rangle + \langle p_s, x \rangle = \langle s, p_x \rangle + \langle p_s, x \rangle,$$

which is (5.4). From this relation it is clear that the point $(x - p_x, y - p_y, s - p_s)$ is not strictly feasible (since $\langle s - p_s, x - p_x \rangle = 0$) and therefore we get (5.7).

Further, multiplying the first equation of (5.1) by $F'_*(s)$ and using the relation $F'(x) = F''(w)F'_*(s)$, we get (5.5) in view of (2.9).

Finally, multiplying the first equation of (5.1) by p_x we get

$$\| p_x \|_w^2 = \langle s, p_x \rangle$$

from (5.3). Similarly, multiplying (5.2) by p_s we get

$$\| p_s \|_w^2 = \langle p_s, x \rangle.$$

Adding these equations we get (5.6) in view of (5.4). $\square$

It is interesting that the affine-scaling direction is related in a very nice way with the functional proximity measure.

THEOREM 5.3. *Let the point (x, y, s) be strictly feasible, and let $w \in \operatorname{int} K$ be such that $F''(w)x = s$. Then for small enough $\alpha \geq 0$ the point*

$$x(\alpha) := x - \alpha p_x, \quad y(\alpha) := y - \alpha p_y, \quad s(\alpha) := s - \alpha p_s$$

is also strictly feasible and for any such α

$$\begin{aligned} \gamma_F(x(\alpha), s(\alpha)) - \gamma_F(x, s) &= F(x - \alpha p_x) + F\left(x + \tfrac{\alpha}{1-\alpha} p_x\right) - 2F(x) \\ &= F\left(x - \tfrac{\alpha^2}{2(1-\alpha)}[F''(x)]^{-1}F'''(x)[p_x, [F''(w)]^{-1}p_s]\right) - F(x). \end{aligned} \tag{5.8}$$

Proof. Since the point (x, y, s) is strictly feasible, for small enough $\alpha \geq 0$ the point $(x(\alpha), y(\alpha), s(\alpha))$ is strictly feasible by the second and third equations of (5.1).

In view of (5.4) and (5.3) we get the following relation for the duality gap:

$$\langle s(\alpha), x(\alpha)\rangle = \langle s, x\rangle - \alpha[\langle s, p_x\rangle + \langle p_s, x\rangle] = (1-\alpha)\langle s, x\rangle. \tag{5.9}$$

Therefore

$$\nu \ln\langle s(\alpha), x(\alpha)\rangle - \nu \ln\langle s, x\rangle = \nu \ln(1-\alpha). \tag{5.10}$$

Further, from (5.1) we have:

$$\begin{aligned} s(\alpha) &= s - \alpha p_s \\ &= s - \alpha[s - F''(w)p_x] \\ &= (1-\alpha)s + \alpha F''(w)p_x \\ &= F''(w)[(1-\alpha)x + \alpha p_x]. \end{aligned}$$

Therefore, using (2.19) we obtain:

$$\begin{aligned} F_*(s(\alpha)) &= F((1-\alpha)x + \alpha p_x) - 2F(w) - \nu \\ &= F((1-\alpha)x + \alpha p_x) + F_*(s) - F(x). \end{aligned} \tag{5.11}$$

Thus, in view of (5.10) and (5.11) we get

$$\begin{aligned} \gamma_F(x(\alpha), s(\alpha)) - \gamma_F(x, s) &= \nu \ln(1-\alpha) + F(x(\alpha)) + F_*(s(\alpha)) - F(x) - F_*(s) \\ &= \nu \ln(1-\alpha) + F(x(\alpha)) + F((1-\alpha)x + \alpha p_x) - 2F(x), \end{aligned}$$

which is the first part of (5.8) by (2.6).

Further, let

$$v = x - \frac{\alpha^2}{2(1-\alpha)}[F''(x)]^{-1}F'''(x)[p_x, [F''(w)]^{-1}p_s], \quad \beta = \frac{\alpha}{1-\alpha}.$$

Then

$$\beta - \alpha = \beta\alpha = \frac{\alpha^2}{1-\alpha}$$

and we have:

$$\begin{aligned} v &= x - \tfrac{1}{2}\beta\alpha[F''(x)]^{-1}F'''(x)[p_x, (p_x + [F''(w)]^{-1}p_s) - p_x] \\ &= x - \tfrac{1}{2}\beta\alpha[F''(x)]^{-1}F'''(x)[p_x, x - p_x] \\ &= x + (\beta - \alpha)p_x + \tfrac{1}{2}\beta\alpha[F''(x)]^{-1}F'''(x)[p_x, p_x]. \end{aligned}$$

Therefore the second part of the equation (5.8) follows from Theorem 3.11. ☐

Denote

$$p(x,s) = -\tfrac{1}{2}[F''(x)]^{-1}F'''(x)[p_x, [F''(w)]^{-1}p_s].$$

In view of the above theorem we need more information about this direction.

LEMMA 5.4.

$$\langle F'(x), p(x,s)\rangle \le 2(1+\lambda^+_\infty(x,s))\sqrt{\gamma_G(x,s)} \mid p_x \mid_x \cdot \parallel p_s \parallel_s, \tag{5.12}$$

$$\langle F'(x), p(x,s)\rangle \le \nu(1+\gamma_\infty(x,s))\sqrt{\gamma_F(x,s)}, \tag{5.13}$$

$$\parallel p(x,s) \parallel_x \le \mid p_x \mid_x \parallel p_s \parallel_s \le \tfrac{1}{2}\nu(1+\gamma_\infty(x,s)). \tag{5.14}$$

Proof. Indeed,

$$\begin{array}{rcl} \langle F'(x), p(x,s)\rangle & = & -\frac{1}{2}\langle F'''(x)[p_x, [F''(w)]^{-1}p_s], [F''(x)]^{-1}F'(x)\rangle \\ & = & \frac{1}{2}F'''(x)[x, p_x, [F''(w)]^{-1}p_s] \\ & = & -\langle F''(x)p_x, [F''(w)]^{-1}p_s\rangle \\ & = & \langle \left[\frac{1}{\mu(x,s)}F''(w) - F''(x)\right] p_x, [F''(w)]^{-1}p_s\rangle. \end{array}$$

Note also that

$$\parallel [F''(w)]^{-1}p_s \parallel_x = \parallel p_s \parallel_s .$$

Therefore, using Theorem 3.8 we obtain:

$$\langle F'(x), p(x,s)\rangle \le \frac{2}{(1-\sigma_x(p))^2} \parallel p \parallel_x \cdot \mid p_x \mid_x \cdot \parallel p_s \parallel_s, \tag{5.15}$$

where $p = x - \sqrt{\mu(x,s)}w$. In view of (3.19) and (3.21) we have:

$$\parallel p \parallel_x \le \sqrt{\gamma_G(x,s)}. \tag{5.16}$$

Further, by definition $\sigma_x(p) = \frac{1}{\alpha}$, where α is the maximal step such that

$$x - \alpha(x - \sqrt{\mu(x,s)}w) \in K.$$

Clearly, such α is greater than one. Therefore

$$\frac{\alpha - 1}{\alpha\sqrt{\mu(x,s)}} = \frac{1}{\sigma_w(x)}.$$

Thus,

$$\frac{1}{(1-\sigma_x(p))^2} = \frac{\alpha^2}{(\alpha-1)^2} = \frac{\sigma_w(x)^2}{\mu(x,s)} = \frac{\sigma_s(x)}{\mu(x,s)} = \lambda^+_\infty(x,s) + 1$$

(see Lemma 3.4). Combining this equality with (5.15) and (5.16) we get (5.12).

In order to prove (5.13), note that in view of Theorem 3.8

$$\begin{array}{l} \mid \langle \left[\frac{1}{\mu(x,s)}F''(w) - F''(x)\right] p_x, [F''(w)]^{-1}p_s\rangle \mid \\ = \frac{1}{\mu(x,s)} \mid \langle \left[F''(x/\sqrt{\mu(x,s)}) - F''(w)\right] p_x, [F''(w)]^{-1}p_s\rangle \mid \\ \le \frac{2}{\mu(x,s)(1-\sigma_w(p))^2} \parallel p \parallel_w \cdot \mid p_x \mid_w \cdot \parallel p_s \parallel_w, \end{array}$$

where now $p = w - x/\sqrt{\mu(x,s)}$. Note that $\| p \|_w^2 \leq \gamma_F(x,s)$ by Lemma 3.12,

$$2 \mid p_x \mid_w \cdot \| p_s \|_w \leq \| p_x \|_w^2 + \| p_s \|_w^2 = \nu\mu(x,s)$$

in view of (5.6) and

$$\sigma_w(p) = 1 - \sqrt{\frac{1}{1+\gamma_\infty(x,s)}}$$

(the proof is quite straightforward). Combining these inequalities, we get (5.13).

Finally, the first inequality in (5.14) follows from Theorem 3.8. In order to prove the second one, note that from Lemma 4.3 and (5.6) we have

$$\begin{array}{rcl} 2 \mid p_x \mid_x \cdot \| p_s \|_s & \leq & \| p_x \|_x^2 + \| p_s \|_s^2 \\ & \leq & \frac{1}{\mu(x,s)}(1+\gamma_\infty(x,s))(\| p_x \|_w^2 + \| p_s \|_w^2) \\ & = & \nu(1+\gamma_\infty(x,s)). \quad \square \end{array}$$

5.2. Centering direction. Let us again fix points $x \in S^0(P)$ and $(y,s) \in S^0(D)$, and let $w \in \operatorname{int} K$ be such that $s = F''(w)x$.

Definition 5.5. *The centering direction for the primal-dual point (x,y,s) is the solution (d_x,d_y,d_s) of the following linear system:*

$$\begin{array}{ccccccc} & & F''(w)d_x & + & d_s & = & s + \mu(x,s)F'(x), \\ & & Ad_x & & & = & 0, \\ A^*d_y & & & + & d_s & = & 0. \end{array} \tag{5.17}$$

Note that the first equation of (5.17) can be rewritten as follows:

$$d_x + [F''(w)]^{-1}d_s = x + \mu(x,s)F'_*(s), \tag{5.18}$$

so the centering direction is also symmetric between the primal and dual. It is also clear from the other two equations that

$$\langle d_s, d_x \rangle = 0. \tag{5.19}$$

Let us present the main properties of the centering direction.

Lemma 5.6. *Let (d_x,d_y,d_s) be the centering direction for the strictly feasible point (x,y,s). Then*

$$\langle s, d_x \rangle + \langle d_s, x \rangle = 0, \tag{5.20}$$

$$\langle F'(x), d_x \rangle + \langle d_s, F'_*(s) \rangle = \gamma_G(x,s), \tag{5.21}$$

$$\| d_x \|_w^2 + \| d_s \|_w^2 = \mu(x,s)\gamma_G(x,s). \tag{5.22}$$

$$\| d_x \|_x^2 + \| d_s \|_s^2 \leq \gamma_G(x,s)(1+\gamma_\infty(x,s)). \tag{5.23}$$

Moreover, the affine-scaling and the centering directions are orthogonal in the metric defined by w:

$$\langle F''(w)p_x, d_x \rangle + \langle d_s, [F''(w)]^{-1}p_s \rangle = 0. \tag{5.24}$$

Proof. Indeed, multiplying the first equation of (5.17) by x, we get (5.20) in view of (2.9). Multiplying the first equation of (5.17) by $F'_*(s)$ and using again (2.9), we obtain:

$$\langle F'(x), d_x \rangle + \langle d_s, F'_*(s) \rangle = -\nu + \mu(x,s)\langle F'(x), F'_*(s) \rangle = \gamma_G(x,s).$$

Further, multiplying the first equation of (5.17) by d_x and using (5.19) we get

$$\| d_x \|_w^2 = \langle s, d_x \rangle + \mu(x,s)\langle F'(x), d_x \rangle.$$

Similarly, from (5.18) we get:

$$\| d_s \|_w^2 = \langle d_s, x \rangle + \mu(x,s)\langle d_s, F_*'(s) \rangle.$$

Adding these equations, we get (5.22) from (5.20) and (5.21).

Next, (5.23) follows from (5.22) and Lemma 4.3.

In order to prove (5.24), let us note that $\langle p_s, d_x \rangle = \langle d_s, p_x \rangle = 0$. Therefore

$$\begin{aligned} \langle F''(w)p_x, d_x \rangle + \langle d_s, [F''(w)]^{-1} p_s \rangle &= \langle s - p_s, d_x \rangle + \langle d_s, x - p_x \rangle \\ &= \langle s, d_x \rangle + \langle d_s, x \rangle = 0 \end{aligned}$$

in view of (5.20). $\square$

Note that (5.20) shows that the centering direction keeps the duality gap constant.

Let us describe the behavior of the Newton process based on the centering direction. Let the point (x, y, s) be strictly feasible. Consider the following process.

$$\begin{array}{l} x_0 = x, \; y_0 = y, \; s_0 = s; \\ \left.\begin{array}{rcl} x_{k+1} &=& x_k - \alpha_k d_{x_k}, \\ y_{k+1} &=& y_k - \alpha_k d_{y_k}, \\ s_{k+1} &=& s_k - \alpha_k d_{s_k}, \end{array}\right\} \; k = 0, \ldots, \end{array} \tag{5.25}$$

where

$$\alpha_k = \frac{1}{1 + \gamma_\infty(x_k, s_k) + \bar{\sigma}_k}, \qquad \bar{\sigma}_k = \max\{\sigma_{x_k}(d_{x_k}), \sigma_{s_k}(d_{s_k})\}. \tag{5.26}$$

THEOREM 5.7. *For any $k \geq 0$ we have*

$$\gamma_F(x_{k+1}, s_{k+1}) \leq \gamma_F(x_k, s_k) - [\tau_k - \ln(1 + \tau_k)], \tag{5.27}$$

where

$$\tau_k = \left(\frac{\gamma_G(x_k, s_k)}{1 + \gamma_\infty(x_k, s_k)} \right)^{1/2} \geq \frac{\gamma_\infty(x_k, s_k)}{1 + \gamma_\infty(x_k, s_k)}.$$

Proof. In order to simplify the notation let us omit subscripts for the current iterate and use a subscript $+$ for the next iterate.

Since from Lemma 5.6 the centering direction does not change the duality gap, the decrease of the functional proximity measure can be estimated as follows, using T_* 4.2 and P_* 4.1:

$$\begin{aligned} \gamma_F(x_+, s_+) - \gamma_F(x, s) &= F(x_+) - F(x) + F_*(s_+) - F_*(s) \\ &\leq -\alpha\langle F'(x), d_x \rangle + \left(\tfrac{\|d_x\|_x}{\sigma_x(d_x)} \right)^2 (-\alpha\sigma_x(d_x) - \ln(1 - \alpha\sigma_x(d_x))) \\ &\quad -\alpha\langle F_*'(s), d_s \rangle + \left(\tfrac{\|d_s\|_s}{\sigma_s(d_s)} \right)^2 (-\alpha\sigma_s(d_s) - \ln(1 - \alpha\sigma_s(d_s))) \\ &\leq -\alpha\gamma_G(x, s) + \tfrac{1}{\bar{\sigma}^2}(\| d_x \|_x^2 + \| d_s \|_s^2)(-\alpha\bar{\sigma} - \ln(1 - \alpha\bar{\sigma})) \\ &\leq -\gamma_G(x, s) \left[\alpha + \tfrac{1 + \gamma_\infty(x,s)}{\bar{\sigma}^2}(\alpha\bar{\sigma} + \ln(1 - \alpha\bar{\sigma})) \right]. \end{aligned}$$

Maximizing the expression in brackets with respect to α, we get the following step size yielding the optimal bound:

$$\alpha = \frac{1}{1 + \gamma_\infty(x, s) + \bar{\sigma}}.$$

Substituting this value in the above chain of inequalities, we obtain the following:

$$\gamma_F(x_+,s_+) - \gamma_F(x,s) \le -\frac{1}{\bar\sigma^2}\gamma_G(x,s)(1+\gamma_\infty(x,s))\left[\frac{\bar\sigma}{1+\gamma_\infty(x,s)} - \ln\left(1+\frac{\bar\sigma}{1+\gamma_\infty(x,s)}\right)\right].$$

Note that the expression in brackets above is monotonically decreasing in $\bar\sigma$ and that

$$\bar\sigma^2 \le \| d_x \|_x^2 + \| d_s \|_s^2 \le (1+\gamma_\infty(x,s))\gamma_G(x,s).$$

Substituting this bound in the previous expression, we get the desired inequality for the change in γ_F. The bound on τ_k follows from (4.18). ∎

6. Some Closely Path-Following Methods. In this section we describe some extensions of standard primal-dual path-following methods for linear programming to the setting of self-scaled cones. All our algorithms generate iterates in the (narrow) neighborhood

$$\mathcal{N}(\beta) := \{(x,y,s) \in S^0(P)\times S^0(D) : \lambda_2(x,s) = \| s/\mu(x,s) + F'(x) \|_x \le \beta\} \tag{6.1}$$

of the primal-dual central path, where $\beta \in (0,1)$. Note that this definition is symmetric between x and s (see (4.13)).

Suppose we are given, at the start of some iteration, a point $(x,y,s) \in \mathcal{N}(\beta)$ with

$$\delta := | s/\mu(x,s) + F'(x) |_x = \lambda_\infty(x,s) \le \epsilon := \| s/\mu(x,s) + F'(x) \|_x = \lambda_2(x,s) \le \beta. \tag{6.2}$$

Actually, we could work with the slightly wider neighborhood $\mathcal{N}'(\beta) := \{(x,y,s) \in S^0(P)\times S^0(D) : \lambda(x,s) \le \beta\}$ for $\beta \in (0,1)$. Note that $(x,y,s) \in S^0(P)\times S^0(D)$ lies in this neighborhood if $\bar\lambda_2(x,s,\mu) = \| s/\mu + F'(x) \|_x \le \beta$ for some $\mu > 0$ by (4.30). We could then try to decrease the value of μ and find new iterates for which this inequality remains true. However, it is easy to see that the inequality implies that μ is very close to $\mu(x,s)$:

LEMMA 6.1. *If $\| s/\mu + F'(x) \|_x \le \beta$ holds for some $\mu > 0$, then*

$$(1-\beta/\sqrt{\nu})\mu \le \mu(x,s) \le (1+\beta/\sqrt{\nu})\mu. \tag{6.3}$$

Proof. To prove the upper bound, note that

$$\begin{aligned}\langle s,x\rangle &= \mu(\langle s/\mu + F'(x), x\rangle - \langle F'(x), x\rangle)\\ &\le \mu(\| s/\mu + F'(x) \|_x \| x \|_x + \nu)\\ &\le (\nu + \beta\sqrt{\nu})\mu,\end{aligned}$$

where we have used (2.9) and (2.10). Now divide by ν. The lower bound is proved similarly. ∎

If we can find, for every such $(x,y,s) \in \mathcal{N}(\beta)$, some $(x_+,y_+,s_+) \in \mathcal{N}(\beta)$ with $\mu(x_+,s_+) \le (1-\kappa/\sqrt{\nu})\mu(x,s)$, for some positive constant κ, then clearly we can construct an algorithm which, given $(x_0,y_0,s_0) \in \mathcal{N}(\beta)$, achieves $(x_k,y_k,s_k) \in \mathcal{N}(\beta)$ with $\langle s_k,x_k\rangle \le \epsilon$ within $O(\sqrt{\nu}\ln(\langle s_0,x_0\rangle/\epsilon))$ iterations. All the algorithms of this section are of this type.

Now let us denote

$$\mu := \mu(x,s)$$

and examine the effect on (6.2) if we replace μ by

$$\mu_+ := (1-\frac{\kappa}{\sqrt{\nu}})\mu \tag{6.4}$$

for some fixed constant κ.

LEMMA 6.2. *Suppose (6.2) holds and μ_+ is given by (6.4). Then*

$$\delta_+ :=| s/\mu_+ + F'(x) |_x \leq \epsilon_+ :=\| s/\mu_+ + F'(x) \|_x \tag{6.5}$$

satisfy

$$\delta_+ \leq \frac{\delta+\kappa}{1-\kappa}, \quad \epsilon_+ \leq \frac{\epsilon+\kappa}{1-\kappa}. \tag{6.6}$$

Proof. Using (2.10), we have

$$\begin{aligned}\epsilon_+ =\| s/\mu_+ + F'(x) \|_x &\leq \frac{\mu}{\mu_+}[\| \frac{s}{\mu} + F'(x) \|_x +(1-\frac{\mu_+}{\mu}) \| F'(x) \|_x] \\ &\leq \frac{1}{1-\kappa}[\epsilon + \frac{\kappa}{\sqrt{\nu}}\sqrt{\nu}]\end{aligned}$$

as desired. We proceed similarly for δ_+, using $| F'(x) |_x \leq \| F'(x) \|_x$. ∎

Having decreased μ, we now need to update x, y, and s so that (6.2) remains true at the new iterate. Hence, with $w \in \operatorname{int} K$ such that $F''(w)x = s$, let q_x, q_s, and q_y solve

$$\begin{array}{rcccl} & F''(w)q_x & + & q_s & = & s + \mu_+ F'(x), \\ & Aq_x & & & = & 0, \\ A^* q_y & & + & q_s & = & 0. \end{array} \tag{6.7}$$

Note that the first equation can alternatively be written as $q_x + F''_*(t)q_s = x + \mu_+ F'_*(s)$, where $t := -F'(w) \in \operatorname{int} K^*$ satisfies $F''_*(t)s = x$, so that this is a symmetric system. Also note that (q_x, q_y, q_s) can be written in terms of the affine-scaling and centering directions of Section 5. From Definitions 5.1 and 5.5 we have

$$(q_x, q_y, q_s) = \frac{\mu - \mu_+}{\mu}(p_x, p_y, p_s) + \frac{\mu_+}{\mu}(d_x, d_y, d_s). \tag{6.8}$$

Now set

$$x_+ := x - q_x, \quad y_+ := y - q_y, \quad s_+ := s - q_s. \tag{6.9}$$

Clearly, $Ax_+ = b$ and $A^* y_+ + s_+ = c$. Also, since $\langle q_s, q_x \rangle = 0$, we have

$$\begin{aligned}\langle s_+, x_+ \rangle &= \langle s, x \rangle - \langle q_s, x \rangle - \langle s, q_x \rangle \\ &= \langle s, x \rangle - \langle F''(w)q_x + q_s, x \rangle \\ &= \langle s, x \rangle - \langle s + \mu_+ F'(x), x \rangle \\ &= \nu \mu_+.\end{aligned}$$

This also follows from (6.8) and Lemmas 5.2 and 5.6. Hence

$$\mu_+ = \mu(x_+, s_+). \tag{6.10}$$

LEMMA 6.3. *Suppose*

$$\eta := \frac{\epsilon_+}{1-\delta} \tag{6.11}$$

satisfies $\eta < 1$ (so that $\delta_+ \leq \epsilon_+ < 1$). Then

$$\begin{gathered}\| q_x \|_x \leq \eta, \| q_s \|_s \leq \eta, \\ x_+ \in S^0(P), (y_+, s_+) \in S^0(D),\end{gathered}$$

and for any $v \in E$, $u \in E^*$,

$$\| v \|_{x_+} \leq \frac{1}{1-\eta} \| v \|_x, \quad \| u \|_{x_+} \leq (1+\eta) \| u \|_x . \tag{6.12}$$

Proof. We have

$$\begin{aligned} \| q_x \|_x^2 &\leq [\mu(1-\delta)]^{-1} \| q_x \|_w^2 \text{ (by Lemma 3.13)} \\ &= [\mu(1-\delta)]^{-1} \langle F''(w) q_x + q_s, q_x \rangle \\ &= [\mu(1-\delta)]^{-1} \langle s + \mu_+ F'(x), q_x \rangle \\ &\leq (\mu_+/\mu)(1-\delta)^{-1} \| s/\mu_+ + F'(x) \|_x \| q_x \|_x \\ &\leq \eta \| q_x \|_x, \end{aligned}$$

which gives the first inequality, and similarly $\| q_s \|_s \leq \eta$. If $\eta < 1$, $x - q_x \in \operatorname{int} K$, so $x_+ \in S^0(P)$, and similarly $(y_+, s_+) \in S^0(D)$. Then (6.12) follows from Theorem 3.6. ☐

Note that it does not seem to be easy to get a good bound (depending on δ's, not ϵ's) on $| q_x |_x$ or $| q_s |_s$.

We can now assemble these results to get

THEOREM 6.4. *Let μ_+, x_+, y_+, s_+, δ_+, ϵ_+, and η be defined as above, and let $\eta < 1$. Then*

$$\lambda_2(x_+, s_+) = \bar{\lambda}_2(x_+, s_+, \mu_+) = \| s_+/\mu_+ + F'(x_+) \|_{x_+} \leq \frac{(2+\delta-\kappa)(\epsilon+\kappa)^2}{(1-\delta)^2(1-\kappa)^3}. \tag{6.13}$$

Proof. The two equations follow from the definitions and (6.10). For the inequality, we write

$$\begin{aligned} s_+ + \mu_+ F'(x_+) &= s - q_s + \mu_+ F'(x - q_x) \\ &= s - q_s + \mu_+ F'(x) + \mu_+ (F'(x - q_x) - F'(x)) \\ &= [s + \mu_+ F'(x) - (q_s + F''(w) q_x)] + [(F''(w) - \mu_+ F''(x)) q_x] \\ &\quad + \mu_+ [F'(x - q_x) - F'(x) - F''(x)(-q_x)]. \end{aligned}$$

The first term vanishes by the definition of our steps. Hence

$$\begin{aligned} \| s_+/\mu_+ + F'(x_+) \|_{x_+} \leq &\| [F''(w)/\mu_+ - F''(x)] q_x \|_{x_+} \\ &+ \| F'(x - q_x) - F'(x) - F''(x)(-q_x) \|_{x_+} . \end{aligned}$$

Now applying Lemma 6.3 and Corollary 3.14 (with μ_+ and δ_+ replacing μ and δ) to the first term, we find

$$\begin{aligned} \| [F''(w)/\mu_+ - F''(x)] q_x \|_{x_+} &\leq (1+\eta) \| [F''(w)/\mu_+ - F''(x)] q_x \|_x \\ &\leq (1+\eta)\delta_+ \| q_x \|_x \leq (1+\delta_+)\eta^2. \end{aligned}$$

Also, T_* 4.3 bounds the second term by $| q_x |_x \| q_x \|_x \leq \eta^2$, so we conclude that

$$\| s_+/\mu_+ + F'(x_+) \|_{x_+} \leq (2+\delta_+)\eta^2 = \frac{2+\delta_+}{(1-\delta)^2} \epsilon_+^2. \tag{6.14}$$

Substituting the upper bounds on δ_+ and ϵ_+ from Lemma 6.2 yields the conclusion of the theorem. ☐

We can now analyse some algorithms. First suppose we choose $\beta = 1/10$ and $\kappa = 1/15$. Then $\delta \leq \epsilon \leq 1/10$, $\delta_+ \leq \epsilon_+ \leq 5/28$, $\eta \leq 50/252 < 1$, and $(2+\delta-\kappa)(\epsilon+\kappa)^2/[(1-\delta)^2(1-\kappa)^3] < 1/10$. Thus these choices ensure that all iterates lie in $\mathcal{N}(1/10)$ and that μ decreases by a factor of $[1 - (15\sqrt{\nu})^{-1}]$ at each iteration. In summary, we have the following method:

ALGORITHM 6.1. *(Short-step method) Let $\beta = 1/10$ and $\kappa = 1/15$, and choose $\zeta > 0$. Given $(x_0, s_0) \in \mathcal{N}(\beta)$, set $k = 0$;*

While $\langle s_k, x_k \rangle > \zeta$ do *begin*

set $(x, y, s) := (x_k, y_k, s_k)$ *and* $\mu := \mu(x, s)$*;*

compute μ_+ *from (6.4);*

compute the directions (q_x, q_y, q_s) *from (6.7);*

set $(x_{k+1}, y_{k+1}, s_{k+1}) := (x_+, y_+, s_+)$ *from (6.9);*

$k := k + 1$*;*

end.

Thus we have extended the short-step algorithm of Monteiro and Adler [MA89] and Kojima, Mizuno, and Yoshise [KMY89] to the setting of self-scaled cones.

If our strategy is, as above, to first decrease μ to μ_+ so that ϵ_+ remains bounded, and then to take one or more steps to get close again with the same value of μ, then we cannot achieve much more than the reduction in μ given above. Indeed, Lemma 6.1 shows that if $\bar{\lambda}_2(x, s, \mu_i) \le \beta$ for two values μ_1 and μ_2, then we must have $\mu_2 \ge (1 - 2\beta/\sqrt{\nu})\mu_1$.

However, better progress is possible using adaptive steps. First, suppose we calculate the steps q_x, q_y, and q_s parametrically in μ_+ and choose the *smallest* μ_+ so that (x_+, y_+, s_+) remains in $\mathcal{N}(1/10)$. By the analysis above, we can choose $\mu_+ \le [1 - (15\sqrt{\nu})^{-1}]\mu$ at each iteration, and thus achieve the same complexity estimate as for the short-step method:

ALGORITHM 6.2. *(Adaptive-step method I) Let* $\beta = 1/10$ *and choose* $\zeta > 0$*. Given* $(x_0, s_0) \in \mathcal{N}(\beta)$*, set* $k = 0$*;*

While $\langle s_k, x_k \rangle > \zeta$ *do*

begin

set $(x, y, s) := (x_k, y_k, s_k)$ *and* $\mu := \mu(x, s)$*;*

compute the directions (q_x, q_y, q_s) *parametrically in* μ_+ *from (6.7);*

choose μ_+ *as small as possible so that* (x_+, y_+, s_+) *defined by (6.9) lies in* $\mathcal{N}(\beta)$*;*

set $(x_{k+1}, y_{k+1}, s_{k+1}) := (x_+, y_+, s_+)$*;*

$k := k + 1$*;*

end.

This algorithm extends the "largest-step path-following" algorithm which was suggested by Monteiro and Adler and made precise and implemented by Mizuno, Yoshise, and Kikuchi [MYK89].

The second adaptive-step algorithm we consider, and the last in this section, is an extension of the predictor-corrector method of Mizuno, Todd, and Ye [MTY90]. In this method, each step consists of two substeps. In the first, we strive to reduce the duality gap while not straying too far from the central path (as measured by the proximity measure λ_2) by taking a step along the affine-scaling direction, while in the second, we take a single step in the centering direction so that the duality gap remains constant and the iterates return to close proximity with the path. In the next section, we will consider a much more sophisticated version of this strategy, where we use the global proximity measure γ_F so that much longer steps can be taken, and correct by a sequence of centering steps as in Section 5.2.

We first consider the predictor step. Here the direction is (p_x, p_y, p_s) from Definition 5.1; so it satisfies (6.7) with $\mu_+ = 0$. We still assume that our iterate (x, y, s) satisfies (6.2). As in Lemma 6.3 (see also the proof of Lemma 5.2), we find

$$\begin{aligned}\| p_x \|_x^2 &\le [\mu(1-\delta)]^{-1} \| p_x \|_w^2 \\ &= [\mu(1-\delta)]^{-1} \langle F''(w) p_x + p_s, p_x \rangle \\ &= [\mu(1-\delta)]^{-1} \langle s, p_x \rangle \\ &\le [\mu(1-\delta)]^{-1} \| s \|_x \| p_x \|_x,\end{aligned}$$

so

$$\begin{aligned}\| p_x \|_x &\le (1-\delta)^{-1} \| s/\mu + F'(x) - F'(x) \|_x \\ &\le (1-\delta)^{-1}(\epsilon + \sqrt{\nu}) \le \frac{1+\epsilon}{1-\delta}\sqrt{\nu}.\end{aligned}$$

We choose the step size α as large as possible so that the next iterates remain approximately centered as measured by λ_2. We will show that, for a certain value of the constant κ,

$$\alpha := \frac{\kappa}{\sqrt{\nu}}$$

suffices. Note that then

$$\| \alpha p_x \|_x \leq \frac{1+\epsilon}{1-\delta}\kappa =: \eta,$$

and we choose κ so that $\eta < 1$. Similarly, $\| \alpha p_s \|_s \leq \eta$. So if we set

$$x^+ := x - \alpha p_x, \quad y^+ := y - \alpha p_y, \quad s^+ := s - \alpha p_s, \tag{6.15}$$

we will have $x^+ \in S^0(P)$ and $(y^+, s^+) \in S^0(D)$. Let us continue to assume that α is given as above. Note that from (5.9)

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

So we set $\mu^+ := (1-\alpha)\mu = \mu(x^+, s^+)$. Then we find

$$\begin{aligned}
\lambda_2(x^+, s^+) &= \| s^+/\mu^+ + F'(x^+) \|_{x^+} \\
&= \| (s - \alpha p_s)/[(1-\alpha)\mu] + F'(x - \alpha p_x) \|_{x^+} \\
&= \| [(1-\alpha)s + \alpha(s - p_s)]/[(1-\alpha)\mu] + F'(x - \alpha p_x) \|_{x^+} \\
&\leq \| s/\mu + F'(x) \|_{x^+} + \| F'(x - \alpha p_x) - F'(x) - F''(x)(-\alpha p_x) \|_{x^+} \\
&\quad + \| \alpha[\frac{F''(w)}{(1-\alpha)\mu} - F''(x)]p_x \|_{x^+} .
\end{aligned}$$

The first term above is at most $(1+\eta)\epsilon$ using Theorem 3.6. By T_* 4.3, the second term is bounded by η^2. Now (3.22) gives two-sided bounds for $F''(w)/\mu$. Dividing by $(1-\alpha)$, we see that the inequalities remain true with μ replaced with $(1-\alpha)\mu$ and δ replaced with

$$\delta' := \frac{\delta + \alpha}{1-\alpha}.$$

Then, following the argument in the proof of Corollary 3.14, we discover that the third term is bounded by $(1+\eta)\delta'\eta$. Using our form of α, we conclude that

$$\| s^+/\mu^+ + F'(x^+) \|_{x^+} \leq (1+\eta)\epsilon + \eta^2 + \frac{\epsilon+\kappa}{1-\kappa}(\eta + \eta^2).$$

Note that if $\epsilon \leq 1/10$ and we choose $\kappa := 1/10$, then $\eta \leq 1/8$ and the bound above is less than $1/6$. Thus we have proved:

THEOREM 6.5. *If (x, y, s) satisfies (6.2) with $\beta = 1/10$ and if we choose as our step size α^* the largest α so that the point defined by (6.15) lies in $\mathcal{N}(1/6)$, then $\alpha^* \geq (10\sqrt{\nu})^{-1}$, and*

$$\mu^+ = [1-\alpha^*]\mu \leq [1 - (10\sqrt{\nu})^{-1}]\mu. \qquad \square$$

We next apply a corrector step to the resulting iterate (x^+, y^+, s^+) to return to $\mathcal{N}(1/10)$. For this, note that we now have a point satisfying (6.2) with $\beta = 1/6$. We proceed as in the first part of this section, but choosing $\kappa = 0$, so that we take a unit step in the centering direction (d_x, d_y, d_s) of Definition 5.5 (so that we apply one step of the centering process (5.25) with a unit step). We find that the corresponding value of η is at most $1/5 < 1$, and, from Theorem 6.4, that the result (x_+, y_+, s_+) of the corrector step lies in $\mathcal{N}(1/10)$ with $\mu_+ := \mu(x_+, s_+) = \mu^+$. This corrector step completes the iteration; we are back in the desired neighborhood with an appropriate reduction of μ. In summary, we have

ALGORITHM 6.3. *(Predictor-corrector method I) Let $\beta = 1/10$ and $\bar{\beta} = 1/6$, and choose $\zeta > 0$. Given $(x_0, y_0, s_0) \in \mathcal{N}(\beta)$, set $k = 0$;*

While $\langle s_k, x_k\rangle > \zeta$ *do*
begin
  set $(x, y, s) := (x_k, y_k, s_k)$ *and* $\mu := \mu(x, s)$*;*
  compute the affine-scaling direction (p_x, p_y, p_s)*;*
  choose the largest stepsize α *so that* (x_+, y_+, s_+) *defined by (6.15) lies in* $\mathcal{N}(\bar{\beta})$*;*
  set $(x, y, s) := (x_+, y_+, s_+)$*;*
  compute the centering direction (p_x, p_y, p_s)*;*
  set $(x_{k+1}, y_{k+1}, s_{k+1}) := (x - d_x, y - d_y, s - d_s)$*;*
  $k := k + 1$*;*
end.

The analysis in this section has only used the local proximity measure λ_2 and the resulting narrow neighborhood $\mathcal{N}(\beta)$. In the next section we will consider a predictor-corrector algorithm based on a possibly much wider neighborhood that uses the proximity measure γ_F. However, for small values of λ_2 and γ_F, there is a close relationship between these measures as shown by Theorem 4.2. To conclude this section, we use this relationship to obtain a lower bound on the step size α^* in the predictor-corrector algorithm above; this will allow us to apply the results of the next section to get more information about α^*.

Suppose that (x, y, s) satisfies (6.2) with $\beta = 1/10$. Then $\lambda_\infty(x, s) \le \lambda_2(x, s) \le 1/10$, so that (4.23) implies

$$\gamma_F(x, s) \le (1/10)^2/[1 - 1/10] = 1/90 = .011\cdots.$$

Let $\tilde{\alpha}$ be the largest α so that the point defined by (6.15) satisfies

$$\gamma_F(x^+, s^+) \le 1/6 - \ln(1 + 1/6) = .012\cdots.$$

Then by (4.19) we see that $\lambda_2(x^+, s^+) \le 1/6$, and hence

$$\alpha^* \ge \tilde{\alpha}.$$

Thus, in addition to $\alpha^* \ge (10\sqrt{\nu})^{-1}$ from Theorem 6.5, we know that α^* is at least the step size $\tilde{\alpha}$ needed to increase γ_F from $1/90$ to $1/6 - \ln(1 + 1/6)$ along the affine-scaling direction. The next section gives some lower bounds for $\tilde{\alpha}$, which give a fortiori lower bounds on α^*.

We have indicated above that we could have used λ instead of λ_2 to define our neighborhood. The argument of the previous paragraph shows that similar results hold when γ_F is employed. Indeed, Theorem 4.2 suggests that, for algorithms for which every iterate (including intermediate ones) lies in a narrow neighborhood of the central path, similar results and estimates are possible whichever of λ, λ_2, γ_F, or γ_G is used to define the neighborhood.

7. Functional Proximity Path-Following Scheme. In this section we consider an algorithm which generates the main sequence of points in the following neighborhood of the primal-dual central path:

$$\mathcal{F}(\beta) := \{(x, y, s) \in S^0(P) \times S^0(D) : \gamma_F(x, s) \le \beta\},$$

where $0 < \beta < 1 - \ln 2$. For that purpose it forms first a predicted point in the neighborhood $\mathcal{F}(\beta + \Delta)$, $\Delta > 0$. Then it computes a new point in the initial neighborhood $\mathcal{F}(\beta)$ using the Newton process (5.25). Note that the predictor step size parameter Δ can be arbitrarily large, which contrasts strongly with the predictor-corrector method of the previous section.

Suppose we are given a positive constant β, $0 < \beta < 1 - \ln 2$, and a strictly feasible point $(x_0, y_0, s_0) \in \mathcal{F}(\beta)$. Consider the following predictor-corrector scheme.

ALGORITHM 7.1. *(Predictor-corrector method II) Choose* $\beta > 0$, $\Delta > 0$, *and* $\zeta > 0$. *Given* $(x_0, s_0) \in \mathcal{F}(\beta)$, *set* $k = 0$*;*
While $\langle s_k, x_k\rangle > \zeta$ *do*
begin
  set $(x, y, s) := (x_k, y_k, s_k)$*;*

compute the affine-scaling direction (p_x, p_y, p_s)*;*
choose the largest stepsize $\alpha := \alpha_k$ *so that* $(x^+, y^+, s^+) := (x - \alpha p_x, x - \alpha p_x, x - \alpha p_x)$ *satisfies* $\gamma_F(x^+, s^+) = \beta + \Delta$*;*
set $(x, y, s) := (x^+, y^+, s^+)$*;*
compute $(x_{k+1}, y_{k+1}, s_{k+1})$ *using the Newton method (5.25) starting from* (x, y, s) *and terminating as soon as a point in* $\mathcal{F}(\beta)$ *is found;*
$k := k + 1$*;*

end.

THEOREM 7.1. *Let* $\beta = \kappa - \ln(1 + \kappa)$ *for some* $\kappa \in (0, 1)$*. Then*

1. Algorithm 7.1 decreases the duality gap at a linear rate:

$$\langle s_{k+1}, x_{k+1} \rangle \le (1 - \alpha_k) \langle s_k, x_k \rangle, \tag{7.1}$$

where

$$\frac{\alpha_k^2}{1 - \alpha_k} \ge \frac{1}{\nu} \omega_1(\Delta, \kappa). \tag{7.2}$$

Moreover,

$$\frac{\alpha_k^2}{1 - \alpha_k} \ge \frac{\omega_2(\Delta, \kappa)}{|\, p_{x_k} \,|_{x_k} \cdot \|\, p_{s_k} \,\|_{s_k}} \ge \frac{\omega_2(\Delta, \kappa)}{\sqrt{\nu} \,|\, p_{x_k} \,|_{x_k} \cdot |\, p_{s_k} \,|_{s_k}}. \tag{7.3}$$

In the above inequalities the positive constants ω_i *depend on* Δ *and* κ *only.*

2. The number of Newton steps N_k *in the corrector phase* 1(c) *is bounded by an absolute constant:*

$$N_k \le \frac{\Delta}{\bar{\tau} - \ln(1 + \bar{\tau})}, \tag{7.4}$$

where

$$\bar{\tau} = \tfrac{1}{2} \sqrt{\frac{3\beta}{1 + \beta}}.$$

Proof.

In order to simplify the notation let us omit subscripts for the current iterate and use a subscript + for the next iterate; the superscript + denotes the intermediate iterates given by the predictor step.

From Theorem 5.3 we know that

$$\Delta = \gamma_F(x^+, s^+) - \gamma_F(x, s) = F(x + \xi p(x, s)) - F(x), \tag{7.5}$$

where $\xi := \alpha^2/(1 - \alpha)$. Since $\kappa < 1$, in view of Theorem 4.2 we have:

$$\lambda_\infty^+(x, s) \le \lambda_\infty(x, s) \le \lambda_2(x, s) \le \kappa,$$

$$\gamma_\infty(x, s) \le \frac{\kappa}{1 - \kappa}, \quad \gamma_G \le \frac{\kappa^2}{1 - \kappa}.$$

Therefore, from Lemma 5.4 we get the following inequalities:

$$\langle F'(x), p(x, s) \rangle \le \frac{2\kappa(1 + \kappa)}{\sqrt{1 - \kappa}} \,|\, p_x \,|_x \cdot \|\, p_s \,\|_s,$$

$$\langle F'(x), p(x, s) \rangle \le \frac{\nu \sqrt{\beta}}{1 - \kappa},$$

$$\| p(x,s) \|_x \leq | p_x |_x \cdot \| p_s \|_s \leq \frac{\nu}{2(1-\kappa)}.$$

In view of (7.5), T_* 4.2 and P_* 4.1 (iii),

$$\Delta = F(x+\xi p(x,s)) - F(x) \leq \xi \langle F'(x), p(x,s)\rangle - \xi \| p(x,s) \|_x - \ln(1 - \xi \| p(x,s) \|_x).$$

Hence

$$\Delta \leq \frac{\xi\nu\sqrt{\beta}}{1-\kappa} - \frac{\xi\nu}{2(1-\kappa)} - \ln\left(1 - \frac{\xi\nu}{2(1-\kappa)}\right).$$

Therefore

$$\frac{\nu\alpha^2}{1-\alpha} = \xi\nu \geq \omega_1(\Delta,\kappa),$$

where the positive constant ω_1 depends on Δ and κ only. Since from (5.20) and (5.4)

$$\langle s_+, x_+\rangle = \langle s^+, x^+\rangle = (1-\alpha)\langle s, x\rangle,$$

we get (7.1) with the estimate (7.2) for the predictor step size.

Similarly,

$$\Delta \leq \frac{2\kappa(1+\kappa)\bar{\xi}}{\sqrt{1-\kappa}} - \bar{\xi} - \ln(1-\bar{\xi}),$$

where

$$\bar{\xi} := \xi \mid p_x \mid_x \cdot \| p_s \|_s .$$

Therefore

$$\xi \geq \frac{\omega_2(\Delta,\kappa)}{\mid p_x \mid_x \cdot \| p_s \|_s},$$

where the positive constant ω_2 depends on Δ and κ only. Thus,

$$\frac{\alpha^2}{1-\alpha} \geq \frac{\omega_2(\Delta,\kappa)}{\mid p_x \mid_x \cdot \| p_s \|_s} \geq \frac{\omega_2(\Delta,\kappa)}{\sqrt{\nu} \mid p_x \mid_x \cdot \mid p_s \mid_s}.$$

Let us prove now that the number of the corrector steps is bounded by an absolute constant.

Let $(\tilde{x},\tilde{y},\tilde{s})$ be an intermediate point of the corrector process. In view of Theorem 5.7, the value of the proximity measure is reduced during this step by $\tilde{\tau} - \ln(1+\tilde{\tau})$, where

$$\tilde{\tau} = \sqrt{\frac{\gamma_G(\tilde{x},\tilde{s})}{1+\gamma_\infty(\tilde{x},\tilde{s})}}.$$

Note also that

$$\tilde{\tau} \geq \bar{\tau} := \sqrt{\frac{3\beta}{4(1+\beta)}}$$

since otherwise by (4.28)

$$\gamma_G(\tilde{x},\tilde{s}) = \tilde{\tau}^2(1+\gamma_\infty(\tilde{x},\tilde{s})) \leq \frac{4}{3}\tilde{\tau}^2(1+\gamma_G(\tilde{x},\tilde{s})) \leq \frac{\beta}{1+\beta}(1+\gamma_G(\tilde{x},\tilde{s}))$$

and therefore $\gamma_F(\tilde{x},\tilde{s})) \le \gamma_G(\tilde{x},\tilde{s})) \le \beta$. Thus, we conclude that the number of the corrector steps does not exceed $\Delta/(\bar{\tau} - \ln(1+\bar{\tau}))$. $\square$

Thus, we have proved three different inequalities for the size of the predictor step. The first inequality (7.2) provides us with the theoretical efficiency estimate of the method. It shows that after

$$O\left(\sqrt{\nu}\ln\frac{\langle s_0, x_0\rangle}{\epsilon}\right)$$

predictor steps we will generate a feasible primal-dual point with the duality gap less than ϵ. However, this inequality cannot describe the potential "adaptivity" of the scheme. We can get this information from the second inequality (7.3).

To explain this, let us suppose that at some iteration $\alpha_k \le .9$; if not, then we know that the duality gap decreases by a factor of 10, which is already very good progress. Then (7.3) gives

$$10\alpha_k^2 \ge \frac{\omega_2(\Delta,\kappa)}{|\, p_{x_k}\,|_{x_k}\cdot \|\, p_{s_k}\,\|_{s_k}} \ge \frac{\omega_2(\Delta,\kappa)}{\sqrt{\nu}\,|\, p_{x_k}\,|_{x_k}\cdot |\, p_{s_k}\,|_{s_k}}. \tag{7.6}$$

Since our algorithm is symmetric, we can also replace the denominator in the second term by $\|\, p_{x_k}\,\|_{x_k}\cdot |\, p_{s_k}\,|_{s_k}$. For definiteness, let us suppose that $\eta_k := \|\, p_{s_k}\,\|_{s_k} \ge \|\, p_{x_k}\,\|_{x_k}$ (otherwise x and s can be switched in what follows), and scale the affine-scaling directions by η_k, to get

$$\bar{p}_{x_k} = \frac{1}{\eta_k}p_{x_k}, \quad \bar{p}_{s_k} = \frac{1}{\eta_k}p_{s_k}, \quad \bar{p}_{y_k} = \frac{1}{\eta_k}p_{y_k},$$

The corresponding step size is

$$\bar{\alpha}_k := \alpha_k\eta_k.$$

Let us call

$$\bar{\alpha}_k^* := \max\{\alpha : x_k \pm \alpha\bar{p}_{x_k} \in K,\ s_k \pm \alpha\bar{p}_{s_k} \in K^*\} = 1/\max\{|\,\bar{p}_{x_k}\,|_{x_k}, |\,\bar{p}_{s_k}\,|_{s_k}\} \ge 1,$$

the *maximal feasible* step size, where the inequality follows from $|\,\bar{p}_{x_k}\,|_{x_k} = |\, p_{x_k}\,|_{x_k}/\eta_k \le 1$ and similarly $|\,\bar{p}_{s_k}\,|_{s_k} \le 1$.

Now the first inequality in (7.6) gives

$$10\bar{\alpha}_k^2 \ge \frac{\omega_2(\Delta,\kappa)\eta_k^2}{|\, p_{x_k}\,|_{x_k}\cdot\|\, p_{s_k}\,\|_{s_k}} = \frac{\omega_2(\Delta,\kappa)\eta_k}{|\, p_{x_k}\,|_{x_k}} = \frac{\omega_2(\Delta,\kappa)}{|\,\bar{p}_{x_k}\,|_{x_k}} \ge \omega_2(\Delta,\kappa)\bar{\alpha}_k^*,$$

so $\bar{\alpha}_k = \Omega(\sqrt{\bar{\alpha}_k^*})$; the predictor step size is at least proportional to the square root of the maximal feasible step size in this normalization.

The second inequality in (7.6) gives

$$10\bar{\alpha}_k^2 \ge \frac{\omega_2(\Delta,\kappa)\eta_k^2}{\sqrt{\nu}\,|\, p_{x_k}\,|_{x_k}\cdot|\, p_{s_k}\,|_{s_k}} = \frac{\omega_2(\Delta,\kappa)}{\sqrt{\nu}\,|\,\bar{p}_{x_k}\,|_{x_k}\cdot|\,\bar{p}_{s_k}\,|_{s_k}} \ge \frac{\omega_2(\Delta,\kappa)}{\sqrt{\nu}}(\bar{\alpha}_k^*)^2,$$

so that $\bar{\alpha}_k = \Omega(\nu^{-1/4}\bar{\alpha}_k^*)$; the predictor step size is at least proportional to the maximal feasible step size divided by $\nu^{1/4}$.

These two results provide a strong indication that the functional neighborhood permits large steps.

Note that these bounds also apply to the step size in the λ_2-based predictor-corrector algorithm of the previous section. Indeed, we showed at the end of Section 6 that the step size α^* used there was at least the step size $\tilde{\alpha}$ needed to go from a point in $\mathcal{F}(\beta)$ with $\beta = 1/90 = .011\cdots < 1 - \ln 2$ along the affine-scaling direction to a point (x^+, y^+, s^+) with $\gamma_F(x^+, s^+) = \beta^+ := 1/6 - \ln(1 + 1/6) = .012\cdots$. Thus the bounds established above also hold for α^*, where now we use $\Delta := \beta^+ - \beta > 0$ and κ such that $\beta = \kappa - \ln(1+\kappa)$. However, the constants hidden in the Ω's above are much smaller for α^* due to the small value of Δ, while in the predictor-corrector method using the functional proximity neighborhood, Δ can be any positive constant.

REFERENCES

[ET76] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, North-Holland, Amsterdam, 1976.

[Gu96] O. Güler, Barrier Functions in Interior-Point Methods, *Math. Oper. Res.* 21 (1996), pp. 860–885.

[Ka84] N. K. Karmarkar, A New Polynomial-Time Algorithm for Linear Programming, *Combinatorica* 4 (1984), pp. 373–395.

[KMY89] M. Kojima, S. Mizuno, and A. Yoshise, A Polynomial-Time Algorithm for a Class of Linear Complementarity Problems, *Math. Programming* 44 (1989), pp. 1–26.

[MTY90] S. Mizuno, M. J. Todd, and Y. Ye, On Adaptive Step Primal-Dual Interior-Point Algorithms for Linear Programming, *Math. Oper. Res.* 18 (1993), pp. 964–981.

[MYK89] S. Mizuno, A. Yoshise, and T. Kikuchi, Practical Polynomial Time Algorithms for Linear Complementarity Problems, *Journal of the Operations Research Society of Japan* 32 (1989), pp. 75–92.

[MA89] R. D. C. Monteiro and I. Adler, Interior Path Following Primal-Dual Algorithms : Part I : Linear Programming, *Math. Programming* 44 (1989), pp. 27–41.

[Ne96] Yu. E. Nesterov, Long-Step Strategies in Interior-Point Primal-Dual Methods, *Math. Programming* 76 (1996), pp. 47–94.

[NN94] Yu. E. Nesterov and A. S. Nemirovskii, *Interior Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, 1994.

[NT97] Yu. E. Nesterov and M.J. Todd, Self-Scaled Barriers and Interior-Point Methods for Convex Programming, *Math. Oper. Res.* 22 (1997), pp. 1–42.

[Ro60] O. Rothaus, Domains of Positivity, *Abh. Math. Sem. Univ. Hamburg* 24 (1960), pp. 189–235.

[TTT96] M.J. Todd, K.-C. Toh, and R.H. Tütüncü, On the Nesterov-Todd Direction in Semidefinite Programming, TR–1156 (1996), School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY (to appear in *SIAM J. Optim.*, 8 (1998)).

[Tu96] L. Tunçel, Primal-Dual Symmetry and Scale Invariance of Interior-Point Algorithms for Convex Optimization, Research Report CORR 96–18, Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada.