INFEASIBLE-START PRIMAL-DUAL METHODS AND INFEASIBILITY DETECTORS FOR NONLINEAR PROGRAMMING PROBLEMS

Yu. Nesterov [*], M. J. Todd [†] and Y. Ye [‡]

May 28, 1996

[*]CORE, Catholic University of Louvain, Louvain-Neuve, Belgium. E-mail: nesterov@core.ucl.ac.be.

[†]School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853. E-mail: miketodd@cs.cornell.edu. Research supported in part by NSF through grant DMS-9505155 and ONR through grant N00014-96-1-0050.

[‡]Department of Management Science, The University of Iowa, Iowa City, IA 52242. E-mail: yyye@dollar.biz.uiowa.edu. Research supported in part by NSF Grants DDM-9207347 and DMI-9522507.

Abstract

In this paper we present several "infeasible-start" path-following and potential-reduction primal-dual interior-point methods for nonlinear conic problems. These methods try to find a recession direction of the feasible set of a self-dual homogeneous primal-dual problem. The methods under consideration generate an ϵ-solution for an ϵ-perturbation of an initial strictly (primal and dual) feasible problem in $O(\sqrt{\nu}\ln\frac{\nu}{\epsilon\rho_f})$ iterations, where ν is the parameter of a self-concordant barrier for the cone, ϵ is a relative accuracy and ρ_f is a feasibility measure.

We also discuss the behavior of path-following methods as applied to infeasible problems. We prove that strict infeasibility (primal or dual) can be detected in $O(\sqrt{\nu}\ln\frac{\nu}{\rho_\cdot})$ iterations, where $\rho_\cdot$ is a primal or dual infeasibility measure.

1 Introduction

Nesterov and Nemirovskii [9] first developed and investigated extensions of several classes of interior-point algorithms for linear programming to general convex programming. Many others have also studied methods of this kind: see, for example, den Hertog [3] and Jarre [4] and the references therein. More recently, Nesterov and Todd [10, 11] have developed a theoretical foundation for the extension of a popular symmetric primal-dual algorithm to solving "self-scaled" convex problems. Most of these results assume that the initial problem and its dual have strictly interior feasible points. One paper that relaxes this assumption is Renegar [15], which studies the complexity of using primal barrier methods and a two-phase approach to first determine feasibility and then, if feasible, obtain approximate optimality.

The aim of this paper is twofold. First, we extend the homogeneous and self-dual linear programming model [18] to convex programming problems in conic form. In this way, we are able to start an interior-point algorithm without any prior knowledge of the feasibility (primal or dual) of the initial problem. (Earlier work on this topic includes Nesterov [8], de Klerk, Roos and Terlaky [5], Luo, Sturm and Zhang [6], Potra and Sheng [12], and Andersen and Ye [1].) We present several path-following and potential-reduction primal-dual interior-point methods, which try to find a recession direction of the feasible set of a "projective version" of the model. Such a direction usually either constitutes an optimal solution for the original problem and its dual or establishes a certificate that the problem problem or its dual is not feasible. (In certain "ill-posed" instances, a recession direction may not give either of these indications.)

Second, we analyze and develop several complexity results according to a classification of the original problem and its dual. For example, our algorithms generate an ϵ-solution for an ϵ-perturbation of a strictly (primal and dual) feasible initial problem in $O(\sqrt{\nu}\ln\frac{\nu}{\epsilon\rho_f})$ iterations, where ν is the parameter of a self-concordant barrier for the convex cone, ϵ is a relative accuracy and ρ_f is a feasibility measure. We can actually do more. We can obtain the result above even if the problem and its dual just have optimal solutions with no duality gap; they need not be strictly feasible. If the pair is "strongly" primal or dual infeasible, we can obtain a certificate of infeasibility for an ϵ-perturbation of the original problem or its dual. Moreover, under suitable conditions we obtain these indications in $O(\sqrt{\nu}\ln\frac{\nu}{\epsilon})$ iterations of certain algorithms. Thus, our complexity bounds hold even for some "ill-posed" problems, a term used in Renegar [13, 14]. Furthermore, if the primal-dual pair is strictly primal or dual infeasible, we can obtain an "exact" certificate of strict primal or dual infeasibility in $O(\sqrt{\nu}\ln\frac{\nu}{\rho.})$ iterations, where $\rho.$ is a primal or dual infeasibility measure.

The paper is organized as follows. In Section 2, we review some general results on convex cones and barrier functions, construct the homogeneous and self-dual model, and study its recession directions and central path. We develop a primal-dual barrier and several proximity measures to the central path of this model in Section 3. Then, we propose and analyze various primal-dual methods for generating a recession direction in Sections 4. In Section 5, we discuss more primal-dual methods if the convex cone of the original problem is "self-scaled." As a consequence, we can derive symmetric primal-dual methods that do not require the solution of linear systems of twice the usual dimension. In Section 6, we establish a number of complexity results according to the classification of the original problem pair. We devote Section 7 to generating strong detectors for strictly infeasible problems. An appendix proves a general technical result which is used in Section 7.

In what follows we often refer to different statements of [9]. 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 [9].

2 The shifted primal-dual problem and the central path

Consider the following minimization problem:

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

where $c \in R^n$, $b \in R^m$, $A \in R^{m\times n}$ is an $m \times n$-matrix and K is a closed convex cone in R^n with nonempty interior. In what follows we assume that cone K is *pointed* (contains no straight line).

We also make the following technical assumption.

Assumption 1 *The matrix* $\begin{pmatrix} A \\ c^T \end{pmatrix}$ *has full row rank and* $b \neq 0$.

Note that this assumption is not restrictive: if the rank of the matrix A is equal to k, $k < m$, then either it is possible to eliminate $(m-k)$ linearly-dependent equality constraints, or the system $Ax = b$ is not consistent. If $c = A^T y$ for some $y \in R^m$, then the objective function in (P) is constant over the feasible set. Therefore in the latter case the problem (P) is a problem of finding a feasible solution. If $b = 0$, either $x = 0$ is optimal in (P), or the problem is unbounded.

We call the convex problem (P) *conic* in accordance with [9], Section 4.1.1. Note that the cone K in (P) could be nonpolyhedral. Several important examples of nonlinear conic problems are presented in [9], Chapter 6.

Let $F(x)$ be a ν-self-concordant logarithmically homogeneous barrier for the cone K (see D^* 2.3.2). Recall that by definition, $F(x)$ must be a self-concordant barrier for K (see D^* 2.3.1) and all $x \in \operatorname{int} K$ and $\lambda > 0$ must satisfy the identity:

$$F(\lambda x) \equiv F(x) - \nu \ln \lambda. \tag{2.2}$$

Note that, by C^* 2.3.3, we have $\nu \geq 1$.

We will often use the following straightforward consequences of (2.2):

$$F'(\lambda x) = \frac{1}{\lambda} F'(x), \quad F''(\lambda x) = \frac{1}{\lambda^2} F''(x), \tag{2.3}$$

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

$$\langle F'(x), x\rangle = -\nu, \tag{2.5}$$

$$\langle F''(x)x, x\rangle = \nu, \quad \langle F'(x), [F''(x)]^{-1}F'(x)\rangle = \nu \tag{2.6}$$

(see P^* 2.3.4).

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

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

and let the function $F_*(s)$ be dual to $F(x)$, namely:

$$F_*(s) := \max\{\langle -s, x\rangle - F(x) : x \in K\}. \tag{2.7}$$

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

$$-F'(x) \in \operatorname{int} K^*, \quad -F'_*(s) \in \operatorname{int} K, \tag{2.8}$$

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

(we used (2.5)),

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

$$F'_*(-F'(x)) = -x, \quad F'(-F'_*(s)) = -s, \tag{2.11}$$

$$F''_*(-F'(x)) = [F''(x)]^{-1}, \quad F''(-F'_*(s)) = [F''_*(s)]^{-1}, \tag{2.12}$$

$$F(x) + F_*(s) \geq -\nu + \nu\ln\nu - \nu\ln\langle s, x\rangle, \tag{2.13}$$

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

In the case of the linear programming (LP), $K = K^* = R^n_+$ and we can take

$$F(x) = -\sum_{i=1}^{n} \ln x^{(i)}, \quad F_*(s) = -n - \sum_{i=1}^{n} \ln s^{(i)}.$$

The value of the parameter ν for both barriers is equal to n. Note that in the LP case all relations (2.3)–(2.6) and (2.8)–(2.13) are trivial.

For semidefinite programming $K = K^*$ is the cone of symmetric positive semidefinite matrices of a certain order, say n. We can take

$$F(x) = -\ln\det(x), \quad F_*(s) = -n - \ln\det(s),$$

and the parameter ν equals the order of the matrices (here much smaller than the dimension of the cones). Other important examples of nonlinear cones and the corresponding barriers can be found in [9].

Let us write out the problem dual to (P):

$$\text{(D)} \quad \begin{array}{ll} \max\limits_{y,s} & \langle b, y\rangle \\ \text{s.t.} & A^T y + s = c \\ & s \in K^*. \end{array} \tag{2.14}$$

Usually the problems (P) and (D) are studied under the following assumption.

Assumption 2 *There exist points $\hat{x} \in \operatorname{int} K$, $\hat{s} \in \operatorname{int} K^*$, and $\hat{y} \in R^m$, which satisfy the following linear system:*

$$A\hat{x} = b,$$
$$A^T\hat{y} + \hat{s} = c.$$

In this case it can be proved that the optimal set of the primal-dual problem

$$\text{(PD)} \quad \begin{array}{ll} \min\limits_{x,y,s} & \langle c,x\rangle - \langle b,y\rangle \\ \text{s.t.} & Ax = b, \\ & A^Ty + s = c, \\ & x \in K,\ s \in K^*, \end{array} \tag{2.15}$$

is nonempty and bounded and that its optimal value is zero (see [9], [7]). Whether Assumption 2 holds or not, it is easy to see that if x and (y,s) are feasible in (P) and (D) respectively, then $\langle c,x\rangle \geq \langle b,y\rangle$, so if equality holds both are optimal, and the pair is optimal in (PD).

Let us fix arbitrary points

$$x_0 \in \operatorname{int} K, \quad s_0 \in \operatorname{int} K^*, \quad y_0 \in R^m, \quad \tau_0 > 0, \quad \kappa_0 > 0.$$

Using this data, we can form the *shifted* feasible set for the primal-dual problem (P), (D):

$$\begin{aligned} \mathcal{F} := \{(x,y,s,\tau,\kappa) : \quad & x \in K,\ y \in R^m,\ s \in K^*, \tau \geq 0,\ \kappa \geq 0, \\ & Ax - \tau b = \check{b}, \\ & A^Ty - \tau c + s = \check{c}, \\ & \langle c,x\rangle - \langle b,y\rangle + \kappa = \check{g}\}, \end{aligned}$$

where

$$\check{b} := Ax_0 - \tau_0 b, \quad \check{c} := A^Ty_0 + s_0 - \tau_0 c, \quad \check{g} := \langle c,x_0\rangle - \langle b,y_0\rangle + \kappa_0.$$

(This set is a "projective version" of the self-dual feasible set used in [18] for constructing infeasible-start methods for LP.)

There is another very convenient way to write the equality constraints of $\mathcal{F}$. Let us write

$$\bar{x} := (x;\tau), \quad \bar{s} := (s;\kappa), \quad \bar{c} := (\check{c};\check{g}), \tag{2.16}$$

where the Matlab-inspired notation $(v;w)$ denotes the vector that is the concatenation of v and w (we will use similar notation for $\bar{x}_0$, $\bar{s}_k$, etc.), and

$$G := \begin{bmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{bmatrix}, \tag{2.17}$$

a skew-symmetric matrix. Then the equality constraints of $\mathcal{F}$ can be rewritten as

$$G\begin{pmatrix} y \\ \bar{x} \end{pmatrix} - \begin{pmatrix} 0 \\ \bar{s} \end{pmatrix} = \begin{pmatrix} \check{b} \\ -\bar{c} \end{pmatrix}. \tag{2.18}$$

We will usually write definitions and theorems first in terms of the original variables (x, y, s, τ, κ), but our derivations and proofs are often in terms of the concatenated variables $\bar{x}$ and $\bar{s}$. This level of aggregation is very useful, since it exhibits most of the structure, but note that the "primal" variable $\bar{x}$ includes the linking variable τ, and similarly the "dual" variable $\bar{s}$ includes κ. We will also use a higher level of aggregation, with z or u denoting the full vector (x, y, s, τ, κ). In the following, we will feel free to write equations such as that above in the displayed style shown or in the form

$$G(y; \bar{x}) - (0; \bar{s}) = (\check{b}; -\bar{c}).$$

Note that our initial point $(x_0, y_0, s_0, \tau_0, \kappa_0)$ (which we will also write as $(\bar{x}_0, y_0, \bar{s}_0)$, etc.) lies in $\mathcal{F}$ with $x_0 \in \operatorname{int} K$, $s_0 \in \operatorname{int} K^*$, $\tau_0 > 0$, and $\kappa_0 > 0$. We call such points *strictly feasible* for $\mathcal{F}$. The set of all such points then forms the relative interior

$$Q := \operatorname{rint} \mathcal{F}$$

of $\mathcal{F}$. We will also need the following simple consequences of the skew-symmetry of G.

Lemma 1 *If $(x, y, s, \tau, \kappa) \in \mathcal{F}$, then we have*

$$\begin{aligned} \langle s, x\rangle + \kappa\tau + \langle s_0, x_0\rangle + \kappa_0\tau_0 &= \langle \bar{s}, \bar{x}\rangle + \langle \bar{s}_0, \bar{x}_0\rangle \\ &= \langle \bar{s}_0, \bar{x}\rangle + \langle \bar{s}, \bar{x}_0\rangle \\ &= \langle s_0, x\rangle + \kappa_0\tau + \langle s, x_0\rangle + \kappa\tau_0 \end{aligned} \tag{2.19}$$

and

$$\langle s, x\rangle + \kappa\tau = \langle \bar{s}, \bar{x}\rangle = \langle \check{c}, x\rangle - \langle \check{b}, y\rangle + \check{g}\tau. \tag{2.20}$$

If in addition the direction $(\Delta x, \Delta y, \Delta s, \Delta\tau, \Delta\kappa)$ satisfies

$$\begin{aligned} &A\Delta x - \Delta\tau b = 0, \\ &A^T\Delta y - \Delta\tau c + \Delta s = 0, \\ &\langle c, \Delta x\rangle - \langle b, \Delta y\rangle + \Delta\kappa = 0. \end{aligned} \tag{2.21}$$

then

$$\langle \Delta s, \Delta x\rangle + \Delta\kappa\Delta\tau = \langle \Delta\bar{s}, \Delta\bar{x}\rangle = 0 \tag{2.22}$$

and

$$\begin{aligned} \langle s, \Delta x\rangle + \langle \Delta s, x\rangle + \kappa\Delta\tau + \Delta\kappa\tau &= \langle \bar{s}, \Delta\bar{x}\rangle + \langle \Delta\bar{s}, \bar{x}\rangle \\ &= \langle \bar{s}_0, \Delta\bar{x}\rangle + \langle \Delta\bar{s}, \bar{x}_0\rangle \\ &= \langle s_0, \Delta x\rangle + \langle \Delta s, x_0\rangle + \kappa_0\Delta\tau + \Delta\kappa\tau_0. \end{aligned} \tag{2.23}$$

Proof:
The first two and last equations of (2.19) follow from the notations, as do those of (2.23) and the first equation of (2.22). The middle equation of (2.19) follows from

$$\begin{aligned} \langle \bar{s} - \bar{s}_0, \bar{x} - \bar{x}_0\rangle &= \langle (0; \bar{s} - \bar{s}_0), (y - y_0; \bar{x} - \bar{x}_0)\rangle \\ &= \langle G(y - y_0; \bar{x} - \bar{x}_0), (y - y_0; \bar{x} - \bar{x}_0)\rangle = 0 \end{aligned}$$

from the skew-symmetry of G. Next, (2.20) comes from

$$\langle \bar{s}, \bar{x} \rangle = \langle (0; \bar{s}), (y; \bar{x}) \rangle = \langle G(y; \bar{x}) - (\check{b}; -\bar{c}), (y; \bar{x}) \rangle = -\langle (\check{b}; -\bar{c}), (y; \bar{x}) \rangle$$

which holds for the same reason. For the remaining equations, note first that (2.21) can be written as

$$G(\Delta y; \Delta \bar{x}) - (0; \Delta \bar{s}) = 0. \tag{2.24}$$

Then the second equality in (2.22) follows from the skew-symmetry of G. Finally, to obtain the middle equation of (2.23), observe that

$$\begin{aligned}
\langle (0; \bar{s} - \bar{s}_0), (\Delta y; \Delta \bar{x}) \rangle &= \langle G(y - y_0; \bar{x} - \bar{x}_0), (\Delta y; \Delta \bar{x}) \rangle \\
&= -\langle G(\Delta y; \Delta \bar{x}), (y - y_0; \bar{x} - \bar{x}_0) \rangle \\
&= -\langle (0; \Delta \bar{s}), (y - y_0; \bar{x} - \bar{x}_0) \rangle.
\end{aligned}$$

□

Any *recession direction* $(x^*, y^*, s^*, \tau^*, \kappa^*)$ of the set $\mathcal{F}$ satisfies the following relations:

$$x^* \in K,\ y^* \in R^m,\ s^* \in K^*, \tau^* \geq 0,\ \kappa^* \geq 0,$$

$$Ax^* - \tau^* b = 0,$$

$$A^T y^* - \tau^* c + s^* = 0,$$

$$\langle c, x^* \rangle - \langle b, y^* \rangle + \kappa^* = 0,$$

$$\langle s^*, x^* \rangle + \tau^* \kappa^* = 0.$$

(Note that the last equation follows easily from the first three, which can also be written as

$$G(y^*; \bar{x}^*) - (0; \bar{s}^*) = 0.$$

Using the skew-symmetry of G in this equation also provides a simple proof.) This implies that if we manage to find such a direction with $\tau^* > 0$ and $\kappa^* = 0$, then $(x^*/\tau^*, y^*/\tau^*, s^*/\tau^*)$ is an optimal primal-dual pair of our problem (PD). Alternatively, if $\tau^* = 0$ and $\kappa^* > 0$ for such a direction, then either the primal problem (P) or the dual problem (D) is shown to be infeasible. For example, if $A^T y^* + s^* = 0$, $\langle b, y^* \rangle > 0$, and $s^* \in K^*$, then there can be no feasible solution x to (P) (else $0 \leq \langle s^*, x \rangle = -\langle A^T y^*, x \rangle = -\langle b, y^* \rangle < 0$). We will analyze in detail how we can make similar conclusions using an approximation to a recession direction of the set $\mathcal{F}$. But first of all let us discuss the possibilities for finding such an approximation.

The most natural way to find a recession direction of a convex set is to minimize a self-concordant barrier for this set. (Note that nontrivial recession directions exist only for an unbounded set and that is exactly the case when a self-concordant barrier does not achieve its minimum.) This idea was used first by de Ghellink and Vial [2] for solving a system of linear inequalities. It was also used in [7] for constructing adaptive-step path-following methods for NLP and in [8] for developing infeasible-start methods for NLP. However, the basic primal-dual set in [8] differs from $\mathcal{F}$ and this results in different possibilities for detecting infeasible cases (as we will discuss later).

As in [8], our analysis is based on the behavior of the *central path* of the set $\mathcal{F}$. Let us give the definition:

Definition 1 *The trajectory* $\{z(\mu) = (x(\mu), y(\mu), s(\mu), \tau(\mu), \kappa(\mu)) : \mu > 0\}$, *where* $z(\mu)$ *is a solution of the minimization problem:*

$$\min_{x,y,s,\tau,\kappa} \{\Psi_\mu(x, y, s, \tau, \kappa) : (x, y, s, \tau, \kappa) \in \mathcal{F}\} \tag{2.25}$$

and

$$\Psi_\mu(x, y, s, \tau, \kappa) := \frac{1}{\mu}[\langle s_0, x\rangle + \langle s, x_0\rangle + \kappa_0\tau + \tau_0\kappa] + F(x) + F_*(s) - \ln\tau - \ln\kappa - 1, \tag{2.26}$$

is called the central path *of the set* $\mathcal{F}$.

In case x_0 is feasible in (P) and (y_0, s_0) in (D), so that $\check{b} = 0$ and $\check{c} = 0$ and all points in $\mathcal{F}$ are scaled feasible points, this central path is closely related to the usual (feasible) central path: if the latter is denoted $\{(\hat{x}(\pi), \hat{y}(\pi), \hat{s}(\pi)) : \pi > 0\}$, then we find

$$z(\mu) = (\mu\hat{x}(1/\mu), \mu\hat{y}(1/\mu), \mu\hat{s}(1/\mu), \mu, 1).$$

Theorem 1 *The central path* $\{z(\mu) = (x(\mu), y(\mu), s(\mu), \tau(\mu), \kappa(\mu))\}$ *is well-defined. Moreover, for any* $\mu > 0$ *the following relations hold:*

$$s(\mu) = -\mu F'(x(\mu)), \quad x(\mu) = -\mu F_*'(s(\mu)), \quad \tau(\mu)\kappa(\mu) = \mu, \tag{2.27}$$

$$\langle s_0, x(\mu)\rangle + \langle s(\mu), x_0\rangle + \kappa_0\tau(\mu) + \tau_0\kappa(\mu) = (\nu + 1)\mu + \langle s_0, x_0\rangle + \kappa_0\tau_0, \tag{2.28}$$

$$F(x(\mu)) + F_*(s(\mu)) - \ln\tau(\mu) - \ln\kappa(\mu) - 1 = -(\nu + 1)(\ln\mu + 1). \tag{2.29}$$

Proof:
Let us prove first that $z(\mu)$ is well-defined. Let

$$\bar{K} := K \times R_+, \quad \bar{\nu} := \nu + 1. \tag{2.30}$$

Note that

$$\bar{F}(\bar{x}) = \bar{F}(x; \tau) := F(x) - \ln\tau \tag{2.31}$$

is a self-concordant barrier for the cone $\bar{K}$ with parameter $\bar{\nu}$, and its dual is

$$\bar{F}_*(\bar{s}) = \bar{F}_*(s; \kappa) = F_*(s) - \ln\kappa - 1 \tag{2.32}$$

defined on the dual cone $\bar{K}^* = K^* \times R_+$. Therefore, in view of (2.13) and (2.19), for any strictly feasible $z = (x, y, s, \tau, \kappa)$ we have:

$$\begin{aligned} \bar{F}(\bar{x}) + \bar{F}_*(\bar{s}) &\geq -\bar{\nu}(1 - \ln\bar{\nu} + \ln\langle\bar{s}, \bar{x}\rangle) \\ &= -\bar{\nu}(1 - \ln\bar{\nu} + \ln[\langle\bar{s}_0, \bar{x}\rangle + \langle\bar{s}, \bar{x}_0\rangle - \langle\bar{s}_0, \bar{x}_0\rangle]). \end{aligned}$$

Hence for any $\mu > 0$ the function $\Psi_\mu(z)$, which we can also write as

$$\Psi_\mu(\bar{x}, y, \bar{s}) = \frac{1}{\mu}[\langle\bar{s}_0, \bar{x}\rangle + \langle\bar{s}, \bar{x}_0\rangle] + \bar{F}(\bar{x}) + \bar{F}_*(\bar{s})$$

defined on $\bar{K} \times R^m \times \bar{K}^*$, is bounded below on $\mathcal{F}$:

$$\begin{aligned}
\Psi_\mu(z) &\geq \frac{1}{\mu}[\langle \bar{s}_0, \bar{x}\rangle + \langle \bar{s}, \bar{x}_0\rangle] - \bar{\nu}(1 - \ln \bar{\nu} + \ln[\langle \bar{s}_0, \bar{x}\rangle + \langle \bar{s}, \bar{x}_0\rangle - \langle \bar{s}_0, \bar{x}_0\rangle]) \\
&\geq \frac{1}{\mu}\langle \bar{s}_0, \bar{x}_0\rangle - \bar{\nu}\ln\mu \\
&= \frac{1}{\mu}[\langle s_0, x_0\rangle + \tau_0\kappa_0] - \bar{\nu}\ln\mu.
\end{aligned}$$

(The second inequality comes from minimizing the right-hand side of the first inequality as a function of $\lambda := \langle \bar{s}_0, \bar{x}\rangle + \langle \bar{s}, \bar{x}_0\rangle$.) Since $\Psi_\mu(z)$ is a self-concordant function, this implies that a solution $z(\mu)$ of the problem (2.25) exists (see T^* 2.2.3(i)). Further, the set $\mathcal{F}$ does not contain a straight line in view of Assumption 1. Therefore the point $z(\mu)$ is unique by C^* 2.1.1.

Let us prove the second part of the theorem. Note that from (2.19) and (2.20),

$$\langle s_0, x\rangle + \langle s, x_0\rangle + \kappa_0\tau + \kappa\tau_0 = \langle \check{c}, x\rangle - \langle \check{b}, y\rangle + \check{g}\tau + \langle s_0, x_0\rangle + \tau_0\kappa_0$$

for any feasible solution z. Let us write out the (scaled) Lagrangian of the problem (2.25), taking into account the above relation:

$$\begin{aligned}
\mathcal{L}(x, y, s, \tau, \kappa, \hat{y}, \hat{x}, \hat{\tau}) &= [\langle \check{c}, x\rangle - \langle \check{b}, y\rangle + \check{g}\tau] + \mu F(x) + \mu F_*(s) - \mu\ln\tau - \mu\ln\kappa \\
&\quad -\langle Ax - \check{b} - \tau b, \hat{y}\rangle + \langle A^T y + s - \check{c} - \tau c, \hat{x}\rangle \\
&\quad +\hat{\tau}[\langle c, x\rangle - \langle b, y\rangle + \kappa - \check{g}] \\
&= -\langle(\check{b}; -\bar{c}), (y; \bar{x})\rangle + \mu\bar{F}(\bar{x}) + \mu\bar{F}_*(\bar{s}) \\
&\quad -\langle G(y; \bar{x}) - (0; \bar{s}) - (\check{b}; -\bar{c}), (\hat{y}; \bar{\hat{x}})\rangle.
\end{aligned}$$

Therefore the first order optimality conditions lead to the following nonlinear system:

$$G(\hat{y}; \bar{\hat{x}}) - (0; \bar{\hat{s}}) = (\check{b}; -\bar{c}), \quad \bar{\hat{s}} = -\mu\bar{F}'(\bar{x}) \tag{2.33}$$

and

$$G(y; \bar{x}) - (0; \bar{s}) = (\check{b}; -\bar{c}), \quad \bar{\hat{x}} = -\mu\bar{F}'_*(\bar{s}). \tag{2.34}$$

Relation (2.33) is obtained by setting to zero the derivative of the Lagrangian with respect to $(y, \bar{x})$, using the skew-symmetry of G; the new variable $\bar{\hat{s}} := (\hat{s}; \hat{\kappa})$ is just introduced as convenient notation. Similarly, the two parts of relation (2.34) are obtained by setting to zero the derivatives of the Lagrangian with respect to $(\hat{y}, \bar{\hat{x}})$ and $\bar{s}$ respectively.

In view of the first part of the theorem this system has a unique feasible solution $z(\mu) = (x(\mu), y(\mu), s(\mu), \tau(\mu), \kappa(\mu))$. Assumption 1 shows that the remaining part of the variables,

$$\hat{z} = \hat{z}(\mu) = (\hat{x}(\mu), \hat{y}(\mu), \hat{s}(\mu), \hat{\tau}(\mu), \hat{\kappa}(\mu)),$$

is also uniquely defined.

Let us show that the system (2.33) – (2.34) is symmetric with respect to $z(\mu)$ and $\hat{z}(\mu)$. Indeed, (2.3) and (2.11) show that the second equation in (2.34) holds if and only if $\bar{s} = -\mu\bar{F}'(\bar{\hat{x}})$, and then the symmetry is evident. Thus, we conclude that the vector $(\hat{z}(\mu), z(\mu))$ also solves (2.33) – (2.34). Therefore, since the solution is unique, we have $\hat{z}(\mu) = z(\mu)$.

Thus, the relations (2.27) follow from (2.33) and (2.34). Relation (2.28) follows from (2.27), (2.19), and (2.5). Finally, relation (2.29) follows from (2.27), (2.2) and (2.10). $\square$

Corollary 1 *For any $\mu > 0$, the point $z(\mu) = (x(\mu), y(\mu), s(\mu), \tau(\mu), \kappa(\mu))$ (which we can also write as $(\bar{x}(\mu), y(\mu), \bar{s}(\mu))$) is the unique solution of the following problem:*

$$\min_{x,y,s,\tau,\kappa} \{\Phi(x,y,s,\tau,\kappa) : \langle s_0, x\rangle + \kappa_0\tau + \langle s, x_0\rangle + \kappa\tau_0 \leq (\nu+1)\mu + \langle s_0, x_0\rangle + \kappa_0\tau_0, (x,y,s,\tau,\kappa) \in \mathcal{F}\}, \tag{2.35}$$

where

$$\Phi(x,y,s,\tau,\kappa) := F(x) + F_*(s) - \ln\tau - \ln\kappa - 1. \tag{2.36}$$

(Note that we can also write the objective function as

$$\Phi(\bar{x}, y, \bar{s}) = \bar{F}(\bar{x}) + \bar{F}_*(\bar{s})$$

and the constraints as

$$\langle \bar{s}_0, \bar{x}\rangle + \langle \bar{s}, \bar{x}_0\rangle \leq \bar{\nu}\mu + \langle \bar{s}_0, \bar{x}_0\rangle, (\bar{x}, y, \bar{s}) \in \mathcal{F}.)$$

Proof:
This statement immediately follows from the uniqueness of $z(\mu)$ and (2.28). $\square$

3 Primal-dual barrier and proximity measure

In this paper we will consider several infeasible-start primal-dual interior-point methods that approximate a recession direction of the set $\mathcal{F}$. To this end they are trying to find a point $z = (x, y, s, \tau, \kappa) \in Q$ with a large enough value of the linear function

$$\begin{aligned} \mu(z) \;&:=\; \tfrac{1}{\nu+1}[\langle s_0, x\rangle + \langle s, x_0\rangle + \kappa_0\tau + \tau_0\kappa - \langle s_0, x_0\rangle - \kappa_0\tau_0] = \tfrac{1}{\nu+1}[\langle s, x\rangle + \tau\kappa] \\ &=\; \tfrac{1}{\bar{\nu}}[\langle \bar{s_0}, \bar{x}\rangle + \langle \bar{s}, \bar{x_0}\rangle - \langle \bar{s_0}, \bar{x_0}\rangle] = \tfrac{1}{\bar{\nu}}\langle \bar{s}, \bar{x}\rangle. \end{aligned} \tag{3.1}$$

(see (2.19)). Let us present a result demonstrating that such a point can be found by minimizing the convex barrier Φ of (2.36).

Proposition 1 *Let $z \in Q$. Then*

$$\mu(z) \geq \exp\left(-\frac{1}{\nu+1}\Phi(z) - 1\right). \tag{3.2}$$

Proof:
Indeed, in view of (2.13) and (2.19) we have:

$$\frac{1}{\nu+1}\Phi(z) + 1 \geq \ln(\nu+1) - \ln[\langle s, x\rangle + \tau\kappa] = -\ln\mu(z).$$

$\square$

Thus, we are interested in efficient strategies for minimizing the barrier $\Phi(z)$. In fact, we shall present two classes of such methods: the first ignores the primal-dual symmetry in the vector z and the barrier Φ, while the second exploits these features and the skew-symmetry of G, but requires that the barrier F and the cone K be self-scaled [10, 11], so that $\bar{F}$ and $\bar{K}$ are also. We will give the definition of this concept later, when we introduce these structured methods. For now, we consider only the general methods. As in [7], [8] the efficiency estimates of these strategies will be derived from the fast decrease of the barrier $\Phi(z)$ along the central path.

Let us start from the following simple result.

Lemma 2 *Let the direction $\Delta z = (\Delta x, \Delta y, \Delta s, \Delta\tau, \Delta\kappa)$ satisfy the linear system (2.21). Then for any $\mu > 0$ we have:*

$$\langle F'(x(\mu)), \Delta x\rangle + \langle \Delta s, F_*'(s(\mu))\rangle - \frac{\Delta\tau}{\tau(\mu)} - \frac{\Delta\kappa}{\kappa(\mu)} = -\frac{1}{\mu}[\langle s_0, \Delta x\rangle + \langle \Delta s, x_0\rangle + \kappa_0\Delta\tau + \tau_0\Delta\kappa].$$

Proof:
Indeed, Lemma 1 implies that (2.23) holds with $(x, y, s, \tau, \kappa) = z(\mu)$. Using Theorem 1, we can write

$$\begin{aligned}
\mu[\langle F'(x(\mu)), \Delta x\rangle + \langle \Delta s, F_*'(s(\mu))\rangle - \frac{\Delta\tau}{\tau(\mu)} - \frac{\Delta\kappa}{\kappa(\mu)}] &= \mu[\langle \bar{F}'(\bar{x}(\mu)), \bar{\Delta x}\rangle + \langle \bar{\Delta s}, \bar{F}_*'(\bar{s}(\mu))\rangle] \\
&= -[\langle \bar{s}(\mu), \bar{\Delta x}\rangle + \langle \bar{\Delta s}, \bar{x}(\mu)\rangle] \\
&= -[\langle \bar{s}_0, \bar{\Delta x}\rangle + \langle \bar{\Delta s}, \bar{x}_0\rangle].
\end{aligned}$$

as desired. □

The analysis from here on is similar to that in [8].

In order to simplify our discussion, let us introduce some notation. Define the following functions on Q (these functions are defined on larger sets, but we restrict them to Q): the barrier

$$\Phi(u) := F(x) + F_*(s) - \ln\tau - \ln\kappa - 1 = \bar{F}(\bar{x}) + \bar{F}_*(\bar{s}),$$

and the linear functions

$$\langle d, u\rangle := \langle s_0, x\rangle + \langle s, x_0\rangle + \kappa_0\tau + \tau_0\kappa = \langle \bar{s}_0, \bar{x}\rangle + \langle \bar{s}, \bar{x}_0\rangle$$

and

$$\mu(u) := \frac{\langle d, u\rangle}{\nu + 1} - \mu_0,$$

where $\mu_0 := (\langle s_0, x_0\rangle + \kappa_0\tau_0)/(\nu + 1)$. Finally, denote by $\{u(\mu)\}$ the central trajectory $\{(x(\mu), y(\mu), s(\mu), \tau(\mu), \kappa(\mu))\}$. We remark that Φ is a $2(\nu+1)$-self-concordant barrier for Q, and that, from (2.28), $\mu(u(\mu)) = \mu$. Note that u coincides with z earlier: we use z when we want to stress that this vector is composed of the primal and dual solutions, and u when we want to argue using general properties following from those below. In this notation, we have proved the following results:

$$\Phi'(u(\mu)) + \frac{1}{\mu}d = 0 \tag{3.3}$$

(Lemma 2),

$$\langle d, u(\mu)\rangle = (\nu+1)(\mu+\mu_0), \tag{3.4}$$

$$\Phi(u(\mu)) = -(\nu+1)(\ln\mu+1), \tag{3.5}$$

(Theorem 1), and

$$\Phi(u(\mu)) = \min\{\Phi(u) : \ \mu(u) \le \mu, \ u \in Q\} \tag{3.6}$$

(Corollary 1).

Let us prove the following important relation (compare with [8], Theorem 2).

Theorem 2 *For any $\mu > 0$ we have:*

$$\langle [\Phi''(u(\mu))]^{-1}\Phi'(u(\mu)), \Phi'(u(\mu))\rangle^{1/2} = \sqrt{\nu+1}. \tag{3.7}$$

Proof:
Indeed, in view of (3.3) we have

$$\Phi''(u(\mu))u'(\mu) = \frac{1}{\mu^2}d = -\frac{1}{\mu}\Phi'(u(\mu)).$$

On the other hand,

$$-\mu\langle \Phi'(u(\mu)), u'(\mu)\rangle = \langle d, u'(\mu)\rangle = \nu+1$$

by (3.4). Combining these two relations we obtain (3.7). □

Recall that $\Phi(u)$ is a $2(\nu+1)$-self-concordant barrier. Therefore for any u, strictly feasible for Q, the local norm of its gradient (that is, the left-hand side of (3.7)) does not exceed $\sqrt{2(\nu+1)}$. Hence, as in [8], the neighborhood of the central path $u(\mu)$ is a region of almost the fastest decrease of the barrier.

Let us introduce a functional proximity measure defined on Q. This measure can be used as an estimate for the distance to the central path $\{u(\mu)\}$.

Definition 2 *The function*

$$\Omega(u) := \Phi(u) + (\nu+1)(\ln\mu(u)+1)$$

is called the functional proximity measure *on Q.*

Note that in view of (3.5) and (3.2) this measure is correctly defined: it is nonnegative for any strictly feasible u and it is zero only along the central path $u(\mu)$.

It is important that we can use the value of the proximity measure $\Omega(u)$ as an estimate for the number of Newton steps necessary to get close to the central path.

Let us consider the following process:

$$\begin{array}{rcl} u_0 & \in & Q, \\ u_{k+1} & = & u_k - \frac{\lambda(u_k)}{1+\lambda(u_k)}\delta(u_k), \quad k = 0, 1, \dots, \end{array} \tag{3.8}$$

where $\delta(u)$ is the solution of the following problem:

$$\begin{array}{rl} \lambda(u) := \max\limits_{\delta} & \langle \Phi'(u), \delta \rangle \\ \text{s.t.} & \langle d, \delta \rangle = 0, \\ & \langle \Phi''(u)\delta, \delta \rangle \leq 1. \end{array} \tag{3.9}$$

Here δ is restricted to lie in the linear span of $Q - Q$.

It is easy to see that (3.8) is the damped Newton method as applied to minimizing the restriction of the proximity measure $\Omega(u)$ to the set $\{u \in Q : \mu(u) = \mu(u_0)\}$. Let us present a result on the convergence of this method (see P^* 2.2.2 and T^* 2.2.3).

Theorem 3 *Let $u_0 \in Q$. Then the sequence $\{u_k\}$ converges to the point $u(\mu(u_0))$ and*

$$\Omega(u_{k+1}) \leq \Omega(u_k) - \psi_*(\lambda(u_k)),$$

where $\psi_*(\zeta) := \zeta - \ln(1+\zeta)$. Moreover, if $\lambda(u_k) \leq 2 - \sqrt{3}$ then the process (3.8) converges quadratically:

$$\lambda(u_{k+1}) \leq \left(\frac{\lambda(u_k)}{1 - \lambda(u_k)} \right)^2.$$

4 Primal-dual infeasible-start methods, I

In this section we discuss different strategies for minimizing the barrier $\Phi(u)$ (in view of Proposition 1 this results in increasing the linear form $\mu(u)$). Here we confine ourselves to general methods that do not exploit the special form of Φ and Q. The next section describes another class of methods using this primal-dual symmetry. Our discussion is similar to that in [8], Section 4, since we have the same "raw material" for constructing primal-dual methods:

- a convex barrier $\Phi(u)$;
- a functional proximity measure $\Omega(u)$;
- a normalizing linear function $\mu(u)$ (in [7] $\mu(u) \equiv \tau$); and
- a reference point on the shifted central path: for any $u \in Q$ this is $u(\mu(u))$.

Therefore in this section we omit all proofs and outline only the basic ideas.

4.1 Penalty strategy

Suppose we want to minimize the barrier $\Phi(u)$ over the set Q. From Theorem 2 we suspect that the region of the fastest decrease of this potential is described by the proximity measure $\Omega(u)$. This measure is strictly positive at all points of Q except the points of the central path $u(\mu)$. Therefore it is possible to use $\Omega(u)$ as a *penalty* for going too far from the neighborhood of the central path. Namely, let us form the *penalty potential*

$$P_\gamma(u) := \frac{1}{1+\gamma}\Phi(u) + \frac{\gamma}{1+\gamma}\Omega(u) = \Phi(u) + \frac{\gamma(\nu+1)}{1+\gamma}(\ln \mu(u) + 1),$$

where $\gamma > 0$ is a penalty coefficient. In fact, we choose $\gamma := \sqrt{\nu+1} - 1$, and then we easily find

$$P_\gamma(u) = (\nu + 1 - \sqrt{\nu+1})(\ln \mu(u) + 1) + F(x) + F_*(s) - \ln \tau - \ln \kappa - 1.$$

This is very similar to the standard primal-dual potential function [16, 17], but with the expected coefficient $(\nu + 1 + \sqrt{\nu+1})$ of $\ln \mu(u)$ replaced by $(\nu + 1 - \sqrt{\nu+1})$; this change is appropriate because here we seek to increase, rather than decrease, $\mu(u)$.

Consider a Newton-type algorithm, which minimizes this potential:

$$\begin{array}{rcl} u_0 & \in & Q, \\ u_{k+1} & = & u_k - \frac{[\Phi''(u_k)]^{-1} P'_\gamma(u_k)}{1+\lambda_p(u_k)}, \quad k = 0, 1, \ldots, \end{array} \tag{4.1}$$

where $\lambda_p(u) := \langle [\Phi''(u)]^{-1} P'_\gamma(u), P'_\gamma(u)\rangle^{1/2}$.

Theorem 4 *If $\gamma = \sqrt{\nu+1} - 1$ then one step of the method (4.1) decreases the value of the penalty potential $P_\gamma(u)$ by an absolute constant:*

$$P_\gamma(u_{k+1}) \leq P_\gamma(u_k) - \theta, \quad \theta > 0.$$

(For a proof see [8], Sections 4.1, Lemma 4. The main property of the penalty potential used in the proof is that the local norm of its gradient is equal to one along the central path.)

Thus, with the choice $\gamma = \sqrt{\nu+1} - 1$, we have:

$$\begin{aligned} P_\gamma(u_0) - \theta k & = \tfrac{1}{1+\gamma}\Phi(u_0) + \tfrac{\gamma}{1+\gamma}\Omega(u_0) - \theta k \\ & \geq P_\gamma(u_k) = \tfrac{1}{1+\gamma}\Phi(u_k) + \tfrac{\gamma}{1+\gamma}\Omega(u_k) \geq \tfrac{1}{1+\gamma}\Phi(u_k) \\ & \geq -\tfrac{\nu+1}{1+\gamma}(\ln \mu(u_k) + 1) = -\sqrt{\nu+1}(\ln \mu(u_k) + 1) \end{aligned}$$

(see Proposition 1). Therefore

$$\mu(u_k) \geq \exp\left\{ \frac{1}{\sqrt{\nu+1}} \left(\theta k - \frac{1}{1+\gamma}\Phi(u_0) - \frac{\gamma}{1+\gamma}\Omega(u_0) \right) - 1 \right\}. \tag{4.2}$$

Note that this estimate is still true if we include in the method (4.1) an appropriate line-search procedure.

Observe that in infeasible-start algorithms we are free to choose the starting points $x_0 \in \operatorname{int} K$, $y_0 \in R^m$, $s_0 \in \operatorname{int} K^*$, τ_0, $\kappa_0 > 0$. Given choices for x_0 and τ_0, the estimate (4.2) implies that a reasonable choice for s_0 and κ_0 is as follows:

$$s_0 := -F'(x_0), \quad \kappa_0 := \frac{1}{\tau_0}.$$

(An alternative is to set $s_0 := -\mu_0 F'(x_0)$ and $\kappa_0 := \frac{\mu_0}{\tau_0}$ for some $\mu_0 > 0$. For example, μ_0 could be chosen so that (y_0, s_0) is nearly dual feasible for some y_0. The results below then hold with straightforward modifications. The same is true for the other methods to follow. For simplicity, we assume that μ_0 is taken as 1 henceforth.) With this choice, $\Omega(u_0) = 0$ (we are starting on the central path at $u(1)$),

$$\Phi(u_0) = F(x_0) + F_*(s_0) - \ln \tau_0 - \ln \kappa_0 - 1 = -\nu - 1,$$

and we have the following lower bound for $\mu(u_k)$:

$$\mu(u_k) \geq \exp\left\{ \frac{\theta}{\sqrt{\nu + 1}} k \right\}. \tag{4.3}$$

To conclude this subsection, let us write out the scheme of the method (4.1) in our initial notation:

1. Choose $x_0 \in \operatorname{int} K$ and $\tau_0 > 0$. Set $\gamma := \sqrt{\nu + 1} - 1$,

$$s_0 := -F'(x_0), \quad \kappa_0 := \frac{1}{\tau_0}, \quad z_0 := (x_0, y_0, s_0, \tau_0, \kappa_0).$$

2. Iterate the following process:

   - Compute the direction

$$\delta_p(z_k) := (\Delta x_p(z_k), \Delta y_p(z_k), \Delta s_p(z_k), \Delta \tau_p(z_k), \Delta \kappa_p(z_k))$$

   as a solution of the following problem:

$$\begin{array}{rll} \lambda_p(z) := & \displaystyle\max_{\Delta x, \Delta y, \Delta s, \Delta \tau, \Delta \kappa} & \langle F'(x) + \frac{\gamma s_0}{(1+\gamma)\mu(z_k)}, \Delta x\rangle + \langle \Delta s, F_*'(s) + \frac{\gamma x_0}{(1+\gamma)\mu(z_k)}\rangle \\ & & + \left(-\frac{1}{\tau} + \frac{\gamma \kappa_0}{(1+\gamma)\mu(z_k)}\right) \Delta \tau + \left(-\frac{1}{\kappa} + \frac{\gamma \tau_0}{(1+\gamma)\mu(z_k)}\right) \Delta \kappa \\ & \text{s.t.} & \text{equation (2.21)}, \\ & & \langle F''(x_k)\Delta x, \Delta x\rangle + \langle \Delta s, F_*''(s_k) \Delta s\rangle \\ & & + \left(\frac{\Delta \tau}{\tau}\right)^2 + \left(\frac{\Delta \kappa}{\kappa}\right)^2 \leq 1, \end{array}$$

   - Set

$$z_{k+1} := z_k - \frac{\lambda_p(z_k)}{1 + \lambda_p(z_k)} \delta_p(z_k), \quad k = 0, 1, \ldots,$$

Note that the subproblem of computing $\delta_p(z_k)$ in this method can be solved using standard linear algebra operations, like matrix multiplication and solution of linear systems. (However, observe that the linear system that arises from the optimality conditions of this problem is roughly twice the size of those that arise in usual primal-dual methods for linear programming. We will address this in the next section.)

In this method it is reasonable to apply an appropriate line-search strategy for the penalty potential, or for the function below, which differs by a constant:

$$(\nu + 1 - \sqrt{\nu + 1}) \ln[\langle s, x\rangle + \kappa \tau] + F(x) + F_*(s) - \ln \tau - \ln \kappa - 1. \tag{4.4}$$

4.2 Long-step path-following strategies

Since the neighborhood of the central path is a region of nearly fastest decrease of the potential, it seems reasonable to follow this path in an appropriate way. As in the previous subsection, we can use our freedom in choosing the initial points to start on the central path: for any $x_0 \in \operatorname{int} K$ and $\tau_0 > 0$, we can choose $s_0 := -F'(x_0)$, $\kappa_0 := 1/\tau_0$. Then our starting point $z_0 := (x_0, y_0, s_0, \tau_0, \kappa_0)$ (and consequently u_0) will be a point of the central path: $u_0 = z_0 = u(1) = z(1)$. Let us discuss now the possible ways to approximately follow this path.

Consider a typical situation: let $u(\mu)$ be a point of the central path and suppose we want to find the new point of the path corresponding to $\mu_+ > \mu$. To this end let us consider the tangent ray

$$u_\mu(\alpha) := u(\mu) + \alpha u'(\mu), \quad \alpha \geq 0, \tag{4.5}$$

where $u'(\mu)$ is a solution of the following system (see (3.3)):

$$\Phi''(u(\mu))u'(\mu) = \frac{1}{\mu^2} d. \tag{4.6}$$

Our strategy is as follows: we choose a feasible tangent step size $\alpha > 0$, then we fix a new value of the penalty parameter $\mu_+ = \mu + \alpha$ and find an approximation to $u(\mu_+)$ using the damped Newton Method (3.8) with $u_\mu(\alpha)$ as a starting point. Note first that, in view of (3.5), we know exactly the value of the potential $\Phi(u)$ at the point $u(\mu_+)$:

$$\Phi(u(\mu_+)) = -(\nu + 1)(\ln(\mu + \alpha) + 1).$$

Second, from Theorem 3 the value $\Omega(u_\mu(\alpha))$ provides us (up to a multiplicative constant) with an upper bound on the number of iterations in the auxiliary Newton process for computing a reasonable approximation to $u(\mu_+)$.

Thus, the methods of the type under consideration are defined by the strategy of choosing the tangent step size α. As in [7] and [8], we can apply the following strategies.

Additive penalty strategy

This is a path-following variant of the penalty strategy. We fix a penalty coefficient $\gamma > 0$ and choose α to minimize the penalty potential $P_\gamma(u_\mu(\alpha))$.

Bounded proximity strategy

Let us fix a constant $\Delta > 0$. Then we choose α in (4.5) as the maximal value satisfying the inequality

$$\Omega(u_\mu(\alpha)) \leq \Delta.$$

Maximal utility strategy

Let us fix a constant $\chi > 0$. Note that we know exactly the "utility" of the update $\mu \to \mu_+$ in terms of the decrease of the potential:

$$\Phi(u(\mu)) - \Phi(u(\mu_+)) = (\nu + 1) \ln\left(1 + \frac{\alpha}{\mu}\right).$$

We also have an estimate for the complexity of this update (in view of Theorem 3 it is $O(\chi + \Omega(u_\mu(\alpha)))$; here we add the positive constant χ to avoid problems if $\Omega(u_\mu(\alpha))$ is zero or small, but any such value suffices). Therefore we can try to find α from the following maximization problem:

$$\max_{\alpha>0} \frac{\ln\left(1+\frac{\alpha}{\mu}\right)}{\chi + \Omega(u_\mu(\alpha))}. \tag{4.7}$$

The idea of strategy (4.7) is as follows: we are trying to maximize a lower bound on the average utility (per Newton step) of the auxiliary minimization process.

In the implementations of these path-following schemes we have to use an approximate tangent direction to the central path. Let us define such an approximation for any point $u \in Q$:

$$p(u) = \frac{1}{\mu^2(u)}[\Phi''(u)]^{-1}d \tag{4.8}$$

(compare with (4.6)).

Consider a bounded-proximity path-following scheme, based on the above definition of $p(u)$:

Choose $\quad u_0 = u(1), \quad \Delta_0 \in (0, 1 - \ln 2), \quad \Delta_2 > \Delta_1 > \Delta_0.$

$$\begin{array}{ll}
\textbf{if} & \Omega(u_k) > \Delta_0 \\
\textbf{then} & \\
\text{(A):} & u_{k+1} = u_k - \frac{\lambda(u_k)}{1+\lambda(u_k)}\delta(u_k), \\
& \text{where } \delta(\cdot) \text{ and } \lambda(\cdot) \text{ are defined in (3.9);} \\
\textbf{else} & \\
\text{(B):} & u_{k+1} = u_k + \alpha_k p(u_k), \\
\textbf{endif} &
\end{array} \tag{4.9}$$

where the step-size parameter α_k in step (B) is chosen in accordance with the following rule:

$$\Delta_1 \le \Omega(u_k + \alpha_k p(u_k)) \le \Delta_2. \tag{4.10}$$

Theorem 5 *There exists an absolute constant $\theta > 0$ depending only on Δ_0, Δ_1 and Δ_2, such that*

$$\mu(u_k) \ge \exp\left\{\frac{\theta}{\sqrt{\nu+1}}k\right\} \tag{4.11}$$

for any $k \ge 0$.

The proof is similar to that of Theorem 4 in [8].

5 Primal-dual infeasible-start methods, II

In this section we describe methods that use the primal-dual symmetry in Φ and Q but are based on the ideas of the previous section. In order to exploit the structure, we need to make additional assumptions on the cone K and the barrier F. We assume they are *self-scaled* [10, 11]. As a consequence, we can derive symmetric primal-dual methods that do not require the solution of linear systems of twice the usual dimension. These methods are penalty potential-reduction or short- or long-step path-following algorithms, and are simple modifications of the methods of [10, 11]. The changes are that here we are trying to increase, rather than decrease, the function $\mu(u) = \mu(\bar{x}, y, \bar{s})$, and that the constraints on allowable directions are different since the primal and dual parts are linked; however, the orthogonality relation (2.22) still holds because of the skew-symmetry of G, and this allows the proofs to go through with only minor modifications.

We start with the definition.

Definition 3 *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 v and x from* $\operatorname{int} K$,

$$F''(v)x \in \operatorname{int} K^* \tag{5.1}$$

and

$$F_*(F''(v)x) = F(x) - 2F(v) - \nu. \tag{5.2}$$

If K admits such a barrier, we call it a self-scaled cone.

In fact, self-scaled cones coincide with homogeneous self-dual cones (see [10]), but we will maintain the name self-scaled to emphasize our primary interest in the associated self-scaled barriers.

For the rest of this section, we assume that F is a self-scaled barrier for K. It follows from the results of [10] that $\bar{F}$ and $\bar{F}_*$ are also self-scaled, each with parameter $\bar{\nu} = \nu + 1$. A key property of self-scaled barriers is the existence of *scaling* points ([10], Theorem 3.2): for any $\bar{x} \in \operatorname{int} \bar{K}$ and $\bar{s} \in \operatorname{int} \bar{K}^*$, there is a unique $\bar{w} \in \operatorname{int} \bar{K}$ satisfying

$$\bar{F}''(\bar{w})\bar{x} = \bar{s}, \tag{5.3}$$

and this point also satisfies

$$\bar{F}''(\bar{w})\bar{F}'_*(s) = \bar{F}'(x).$$

Note also from (2.12) that

$$\bar{F}''_*(\bar{t})\bar{s} = \bar{x}$$

for $\bar{t} := -\bar{F}'(\bar{w})$, so we also have a scaling point when we interchange primal and dual points.

5.1 Initial points

Before we describe the methods, let us show how the self-scaled case allows a more natural symmetric choice of starting point. In the previous section, we often assumed that we were given an initial primal point $x_0 \in \operatorname{int} K$ and $\tau_0 > 0$, and then set $s_0 := -\mu_0 F'(x_0) \in \operatorname{int} K^*$ and $\kappa_0 := \mu_0/\tau_0$ for some $\mu_0 > 0$. Alternatively, if we were given $s_0 \in \operatorname{int} K^*$ we could set $x_0 = -\mu_0 F'_*(s_0) \in \operatorname{int} K$. (We chose μ_0 as 1 for simplicity.) But there was no way (without applying a subsidiary algorithm) to generate points on the central path using both x_0 and s_0. Here this is possible.

Given initial points $\bar{x}_0 \in \operatorname{int} \bar{K}$ and $\bar{s}_0 \in \operatorname{int} \bar{K}^*$, find the scaling point $\bar{w}_0 \in \operatorname{int} \bar{K}$ such that $\bar{F}''(\bar{w}_0)\bar{x}_0 = \bar{s}_0$, and set $\mu_0 := \langle \bar{s}_0, \bar{x}_0 \rangle / \bar{\nu}$. Then reset

$$\bar{x}_0 := \mu_0^{1/2} \bar{w}_0, \quad \bar{s}_0 := -\mu_0^{1/2} \bar{F}'(\bar{w}_0).$$

The result is a pair on the central path. Moreover, $\langle \bar{s}_0, \bar{x}_0 \rangle / \bar{\nu}$ remains equal to μ_0, and whenever the initial points were on the central path, they are unchanged by this procedure. For ease in comparing our results with those of the previous section, we assume that our starting point is $u(\mu_0) = z(\mu_0)$ for μ_0 equal to 1 henceforth.

5.2 Search directions

The search directions for all our methods come from solving a subproblem of the following type. Suppose the current point is $(\bar{x}, y, \bar{s}) := (\bar{x}_k, y_k, \bar{s}_k) \in Q$, and let $\bar{w} \in \operatorname{int} \bar{K}$ be such that $\bar{F}''(\bar{w})\bar{x} = \bar{s}$. Also suppose

$$\bar{F}''(\bar{w}) g_{\bar{s}} = g_{\bar{x}}. \tag{5.4}$$

Then consider the solution to

$$\begin{array}{ll} \min_{\Delta y, \Delta \bar{x}, \Delta \bar{s}} & \langle g_{\bar{x}}, \Delta \bar{x} \rangle + \langle \Delta \bar{s}, g_{\bar{s}} \rangle + \frac{1}{2} \langle \bar{F}''(\bar{w}) \Delta \bar{x}, \Delta \bar{x} \rangle + \frac{1}{2} \langle \Delta \bar{s}, [\bar{F}''(\bar{w})]^{-1} \Delta \bar{s} \rangle \\ & G(\Delta y; \Delta \bar{x}) - (0; \Delta \bar{s}) = 0. \end{array} \tag{5.5}$$

Note first that this direction can be thought of as the steepest-descent direction (in a suitable metric) for a function with gradient $(g_{\bar{s}}, 0, g_{\bar{s}})$ with respect to the variables $(\bar{x}, y, \bar{s})$, while staying in the set Q. Second, the direction is very similar to that arising from the subproblem at the end of Section 4.1, if we choose $g_{\bar{x}}$ and $g_{\bar{s}}$ appropriately. (It is also similar to that arising from the centering problem (3.9), except that the latter contains another linear constraint.) The two changes from that problem are: first, we have put the quadratic terms into the objective function here, instead of using them for a constraint. The effect of this is merely to change the length of the solution. Next, our quadratic terms are defined using the Hessian of the barrier function $\bar{F}$ at $\bar{w}$ and of the barrier function $\bar{F}_*$ at $-\bar{F}'(\bar{w})$, rather than at $\bar{x}$ and $\bar{s}$. This uses the primal-dual structure of the problem, and allows an easier solution, as we now show.

The solution to (5.5) is characterized as solving, together with some multiplier vector $(\lambda_y; \lambda_{\bar{x}})$, the system

$$\begin{array}{rcrcrcl} & & G(\Delta y; \Delta \bar{x}) & - & (0; \Delta \bar{s}) & = & 0, \\ (0; g_{\bar{x}}) & + & (0; \bar{F}''(\bar{w}) \Delta \bar{x}) & - & G^T(\lambda_y; \lambda_{\bar{x}}) & = & 0, \\ g_{\bar{s}} & + & \lambda_{\bar{x}} & + & [\bar{F}''(\bar{w})]^{-1} \Delta \bar{s} & = & 0. \end{array} \tag{5.6}$$

Suppose that we have instead a solution to

$$\begin{array}{rcrcl} G(\Delta y; \Delta \bar{x}) & - & (0; \Delta \bar{s}) & = & 0, \\ \bar{F}''(\bar{w}) \Delta \bar{x} & + & \Delta \bar{s} & = & -g_{\bar{x}}. \end{array} \tag{5.7}$$

Then this solution, together with $(\lambda_y; \lambda_{\bar{x}}) := (\Delta y, \Delta \bar{x})$ solves (5.6) above. Indeed, the first equations are identical. Since G is skew-symmetric, $-G^T = G$, and so (5.7) implies the second equation of (5.6). Finally, (5.4) and the second equation of (5.7) imply the last equation of (5.6).

Moreover, it is easy to see that the latter linear system has a unique solution: if $(\Delta\bar{x}, \Delta y, \Delta\bar{s})$ solves the corresponding homogeneous system, then the first equation shows that $\langle\Delta\bar{s}, \Delta\bar{x}\rangle = 0$, and then the second yields $\langle\bar{F}''(\bar{w})\Delta\bar{x}, \Delta\bar{x}\rangle = 0$. This implies that $\Delta\bar{x} = 0$, and then it easily follows that $\Delta\bar{s} = 0$ and $\Delta y = 0$ also.

The second equation in (5.7) can equivalently be written as

$$\Delta\bar{x} + \bar{F}''_*(\bar{t})\Delta\bar{s} = -g_{\bar{s}},$$

where $\bar{t} := -\bar{F}'(\bar{w})$, which shows that (except for the form of the first equation, which is an artifact of the way we have written the constraints) this system is symmetric between $\bar{x}$ and $\bar{s}$.

Finally, note that system (5.7) is comparable in size to the usual primal-dual system that arises in interior-point methods.

Two particular choices for $(g_{\bar{x}}, g_{\bar{s}})$ that satisfy (5.4) are $(\bar{s}, \bar{x})$ (this gives the *affine-scaling* directions), and $(\bar{s} + \mu\bar{F}'(\bar{x}), \bar{x} + \mu\bar{F}'_*(\bar{s}))$ for $\mu = \mu(\bar{x}, y, \bar{s})$ (these are the *centering* directions). Compare with Definitions 5.1 and 5.2 in [11].

Let us also compare these directions with those defined in the previous sections. The affine-scaling direction is very similar to the approximate tangent direction $p(u)$ defined in (4.8). The latter can be viewed as proportional to the solution to a problem like (5.5), with different Hessian matrices and with $(\bar{s}_0, \bar{x}_0)$ for $(g_{\bar{x}}, g_{\bar{s}})$. (Actually, the constant of proportionality is negative, because the affine-scaling direction points "down" the central path towards $\mu = 0$ while the tangent direction points "up" towards $\mu = \infty$.) But note that (2.23) implies that we could equivalently use $(\bar{s}, \bar{x})$ for $(g_{\bar{x}}, g_{\bar{s}})$ (thus satisfying (5.4)), and this is as in the affine-scaling direction.

To compare the centering directions, we first note that the results in [11] concerning the affine-scaling and centering directions remain true with the definition above. Indeed, the only use made there of the feasibility equations on the directions (which differ from those we employ) is that the resulting directions satisfy the orthogonality condition (2.22), which still holds by Lemma 1. There is a slight exception. In the proof of (5.24) in [11], a result is used which could be expressed in our present context as

$$\langle\Delta\bar{s}^a, \Delta\bar{x}^c\rangle = \langle\Delta\bar{s}^c, \Delta\bar{x}^a\rangle = 0,$$

where the superscript "a" (respectively, "c") refers to the affine-scaling (respectively, centering) directions. This may not hold true. However, all that is needed is that the sum of these two terms be zero, which holds since

$$\begin{aligned}\langle\Delta\bar{s}^a, \Delta\bar{x}^c)\rangle + \langle\Delta\bar{s}^c, \Delta\bar{x}^a)\rangle &= \langle(0; \Delta\bar{s}^a), (\Delta y^c; \Delta\bar{x}^c)\rangle + \langle(0; \Delta\bar{s}^c), (\Delta y^c; \Delta\bar{x}^a)\rangle \\ &= \langle G(\Delta y^a; \Delta\bar{x}^a), (\Delta y^c; \Delta\bar{x}^c)\rangle + \langle G(\Delta y^c; \Delta\bar{x}^c), (\Delta y^a; \Delta\bar{x}^a)\rangle \\ &= 0\end{aligned}$$

by the skew-symmetry of G. Hence Lemma 5.3 in [11] implies that the centering directions satisfy

$$\langle\bar{s}, \Delta\bar{x}\rangle + \langle\Delta\bar{s}, \bar{x}\rangle = 0. \tag{5.8}$$

Therefore, in the problem (5.5) for the centering directions, we could add (5.8) as an explicit constraint, which means that we could then replace $(\bar{s} + \mu \bar{F}'(\bar{x}), \bar{x} + \mu \bar{F}'_*(\bar{s}))$ in the objective function by $(\mu \bar{F}'(\bar{x}), \mu \bar{F}'_*(\bar{s}))$. Then again, except for the change in the Hessians and the length of the directions, we have the problem defining the centering direction at the end of Section 3.

We now proceed to describe methods using these directions, using the results of [10, 11].

5.3 Penalty strategy

Here our goal is to decrease the penalty potential P_γ of Section 4.1. Up to a constant, this function (for $\gamma := \sqrt{\bar{\nu}} - 1$) coincides with

$$\tilde{P}_\gamma(\bar{x}, y, \bar{s}) := (\bar{\nu} - \sqrt{\bar{\nu}}) \ln\langle \bar{s}, \bar{x} \rangle + \bar{F}(\bar{x}) + \bar{F}_*(\bar{s})$$

(see (4.4) and recall (3.1)). This is very similar to the usual primal-dual potential function; the only difference is that $(\bar{\nu} - \sqrt{\bar{\nu}})$ replaces $(\bar{\nu} + \sqrt{\bar{\nu}})$ as the coefficient of $\ln\langle \bar{s}, \bar{x} \rangle$, which is appropriate since we are trying to increase rather than decrease the latter.

To reduce $\tilde{P}_\gamma$, we employ the symmetric primal-dual method of Section 8 of [10], with $\rho := -\sqrt{\bar{\nu}}$ instead of $\rho = \sqrt{\bar{\nu}}$. The search directions corresponding to (8.1) in [10] are exactly the solution to (5.7), where

$$g_{\bar{x}} := \frac{\bar{\nu} - \sqrt{\bar{\nu}}}{\langle \bar{s}, \bar{x} \rangle} s + \bar{F}'(\bar{x}). \tag{5.9}$$

These directions can easily be seen to be equal to

$$\frac{\bar{\nu}}{\langle \bar{s}, \bar{x} \rangle} (\Delta \bar{x}^c, \Delta y^c, \Delta \bar{s}^c) - \frac{\sqrt{\bar{\nu}}}{\langle \bar{s}, \bar{x} \rangle} (\Delta \bar{x}^a, \Delta y^a, \Delta \bar{s}^a).$$

Next we choose a step size to approximately minimize $\tilde{P}_\gamma$ along these directions from the current point. Note that the proof of Theorem 8.1 in [10] only requires ρ to satisfy $\rho^2 \geq \bar{\nu}$, which holds for $\rho := -\sqrt{\bar{\nu}}$. Hence we conclude

Theorem 6 *If $\gamma = \sqrt{\bar{\nu}} - 1$ then a suitable step along the directions $(\Delta \bar{x}, \Delta y, \Delta \bar{s})$ that solve (5.7) where $g_{\bar{x}}$ is given by (5.9) yields new iterates $(\bar{x}_{k+1}, y_{k+1}, \bar{s}_{k+1})$ that decrease the value of the penalty potential $\tilde{P}_\gamma$ and hence P_γ by an absolute constant:*

$$P_\gamma(\bar{x}_{k+1}, y_{k+1}, \bar{s}_{k+1}) \leq P_\gamma(\bar{x}_k, y_k, \bar{s}_k) - \theta, \quad \theta > 0.$$

From this result, we can proceed exactly as in Section 4.1 to obtain, with a suitable starting point, the estimate

$$\mu(\bar{x}_k, y_k, \bar{s}_k) \geq \exp\left\{ \frac{\theta}{\sqrt{\nu + 1}} k \right\} \tag{5.10}$$

for all $k \geq 0$.

5.4 Short-step path-following methods

Here we show that the methods of Section 6 of [11] can be used to increase $\mu(\bar{x}, y, \bar{s})$ at each iteration. We suppose that our current iterate $(\bar{x}_k, y_k, \bar{s}_k)$ lies in the narrow neighborhood

$$\mathcal{N}(\beta) := \{(\bar{x}, y, \bar{s}) \in Q : \|\bar{s}/\mu(\bar{x}, y, \bar{s}) + \bar{F}'(\bar{x})\|_{\bar{x}} \le \beta\}$$

of the central path, where $\|\bar{v}\|_{\bar{x}} := \langle \bar{v}, [\bar{F}''(\bar{x})]^{-1}\bar{v}\rangle^{1/2}$ and $\beta \in (0, 1)$. We will also write $(\bar{x}, y, \bar{s})$ for our current iterate, and $\bar{w}$ for the corresponding scaling point.

In [11] we sought to decrease μ, whereas here we wish to increase it. So (6.4) in [11] becomes

$$\mu_+ := (1 + \frac{\lambda}{\sqrt{\nu}})\mu,$$

where λ is a positive constant and $\mu := \mu(\bar{x}, y, \bar{s})$. (We use λ instead of κ in [11] since κ is now a variable of our problem.) We now follow [11]. Our search directions $(\Delta\bar{x}, \Delta y, \Delta\bar{s})$ are the solution to (5.7), where

$$g_{\bar{x}} := \bar{s} + \mu_+ \bar{F}'(\bar{x}). \tag{5.11}$$

We can also write

$$(\Delta\bar{x}, \Delta y, \Delta\bar{s}) = \frac{\mu_+}{\mu}(\Delta\bar{x}^c, \Delta y^c, \Delta\bar{s}^c) - \frac{\mu_+ - \mu}{\mu}(\Delta\bar{x}^a, \Delta y^a, \Delta\bar{s}^a)$$

(compare with (6.7) and (6.8) in [11]).

Theorem 7 *If $\beta := 1/10$ and $\lambda := 1/15$, a full step along this direction from $(\bar{x}_k, y_k, \bar{s}_k)$ yields a new iterate $(\bar{x}_{k+1}, y_{k+1}, \bar{s}_{k+1})$ also in $\mathcal{N}(\beta)$ with*

$$\mu(\bar{x}_{k+1}, y_{k+1}, \bar{s}_{k+1}) \ge (1 + \frac{\lambda}{\sqrt{\nu}})\mu(\bar{x}_k, y_k, \bar{s}_k).$$

Hence, if we choose $(\bar{x}_0, y_0, \bar{s}_0) := (\bar{x}(1), y(1), \bar{s}(1))$, there exists an absolute constant $\theta > 0$ such that

$$\mu(\bar{x}_k, y_k, \bar{s}_k) \ge \exp\left\{\frac{\theta}{\sqrt{\nu + 1}}k\right\} \tag{5.12}$$

for all $k \ge 0$.

The second part of this result follows immediately from the first, which in turn follows from the analysis of [11], with suitable minor changes. In Lemma 6.2, the upper bounds on δ_+ and ϵ_+ become $\delta + \lambda$ and $\epsilon + \lambda$; in Lemma 6.3, η must be defined as $\epsilon_+(1+\lambda)/(1-\delta)$; and in Theorem 6.1, the bound becomes $(2 + \delta + \lambda)(\epsilon + \lambda)^2(1 + \lambda)^2/(1 - \delta)^2$.

The method and analysis above are for the basic short-step method. We can similarly adapt the "largest-step" and predictor-corrector methods of Section 6 of [11]; note that the predictor step should be along the negative of the affine-scaling directions, since we wish to increase μ. We will not give the details, but we observe that these methods lead to similar complexity estimates, while they are likely to be more efficient in practice.

All these algorithms confine the iterates to a narrow neighborhood of the central path of the form of $\mathcal{N}(\beta)$. In the next subsection we consider a long-step path-following strategy that uses a much wider neighborhood.

5.5 A long-step path-following method

Here we describe an analogue of the bounded-proximity method of Section 4.2 for the self-scaled case. This is a modification of the functional proximity path-following scheme of [11]; instead of moving "down" the central path we move "up."

The method has two kinds of steps: predictor steps, along the negative affine-scaling directions, and corrector steps, along the centering directions. Note that our parameter $\Omega(\bar{x}, y, \bar{s})$ coincides with the functional proximity measure $\gamma_F(\bar{x}, \bar{s})$ of [11]. Hence the centering process (5.25) of [11] corresponds to process (4.9) above, with different step sizes, and Theorem 5.2 of [11] gives a comparable bound to Theorem 5 above on the number of steps required in each centering phase.

The only change we make to the algorithm of (7.1) in [11] is that we choose α_k to be negative so that we are moving in the negative affine-centering directions, and we allow a little more flexibility in the step size.

Choose $(\bar{x}_0, y_0, \bar{s}_0) = (\bar{x}(1), y(1), \bar{s}(1)), \quad \Delta_0 \in (0, 1 - \ln 2), \quad \Delta_2 > \Delta_1 > \Delta_0.$
Set $\zeta := \sqrt{\frac{3\Delta_0}{4(1+\Delta_0)}}$.

if $\Omega(\bar{x}_k, y_k, \bar{s}_k) > \Delta_0$
then
(A): $(\bar{x}_{k+1}, y_{k+1}, \bar{s}_{k+1}) = (\bar{x}_k, y_k, \bar{s}_k) + \lambda(\Delta\bar{x}^c, \Delta y^c, \Delta\bar{s}^c)$,
where λ is chosen so that $\Omega(\bar{x}_{k+1}, y_{k+1}, \bar{s}_{k+1}) \leq \Omega(\bar{x}_k, y_k, \bar{s}_k) - [\zeta - \ln(1+\zeta)]$;
else
(B): $(\bar{x}_{k+1}, y_{k+1}, \bar{s}_{k+1}) = (\bar{x}_k, y_k, \bar{s}_k) + \alpha_k(\Delta\bar{x}^a, \Delta y^a, \Delta\bar{s}^a)$,
endif

(5.13)

where the negative step-size parameter α_k in step (B) is chosen in accordance with the following rule:

$$\Delta_1 \leq \Omega[(\bar{x}_k, y_k, \bar{s}_k) + \alpha_k(\Delta\bar{x}^a, \Delta y^a, \Delta\bar{s}^a)] \leq \Delta_2. \tag{5.14}$$

Theorem 8 *There exists an absolute constant $\theta > 0$ depending only on Δ_0, Δ_1 and Δ_2, such that*

$$\mu(\bar{x}_k, y_k, \bar{s}_k) \geq \exp\left\{\frac{\theta}{\sqrt{\nu+1}}k\right\} \tag{5.15}$$

for any $k \geq 0$.

The proof of this result follows that of Theorem 7.1 in [11]. First, the stated decrease in Ω at a corrector step (A) can be achieved, and the number of such steps between two successive predictor steps (B) is bounded by the constant $(\Delta_2 - \Delta_0)/[\zeta - \ln(1+\zeta)]$, exactly as in [11]; this follows from Theorem 5.2 there. Also, we can apply Theorem 5.1 in [11] for α negative, by using Theorem 3.3 with α and β reversed. Thus the proof of Theorem 7.1 there shows that a predictor step leads to

$$\langle \bar{s}_{k+1}, \bar{x}_{k+1} \rangle = (1 - \alpha_k)\langle \bar{s}_k, \bar{x}_k \rangle$$

with

$$\alpha_k \leq -\frac{\omega}{\sqrt{\bar{\nu}}}$$

for some positive constant ω depending only on Δ_0 and Δ_1. Together, these results imply (5.15).

6 Generating the output: weak detectors

We have discussed several efficient strategies for generating a point $z =: (x, y, s, \tau, \kappa) \in \mathcal{F}$ with a large value of $\mu(z)$. Note that this can be done for *any* initial problem (P), feasible or not. Intuitively, we can expect that a feasible point z with a large value of $\mu(z)$ will help in our post-run analysis. In order to describe the possibilities we should discuss first a classification of the "feasibility patterns" of a pair of dual conic problems.

The standard classification of these patterns looks as follows.

1. A primal-dual pair (P), (D) is called (primal-dual) *feasible* if there exist vectors $x^* \in K$, $s^* \in K^*$ and $y^* \in R^m$ such that

$$Ax^* = b, \quad A^T y^* + s^* = c;$$

it is *solvable* if (x^*, y^*, s^*) can be chosen so that also

$$\langle s^*, x^* \rangle = \langle c, x^* \rangle - \langle b, y^* \rangle = 0;$$

and it is *strictly* (primal-dual) *feasible* if (x^*, y^*, s^*) can be chosen so that $x^* \in \operatorname{int} K$ and $s^* \in \operatorname{int} K^*$, or equivalently if it satisfies Assumption 2. Note that strict feasibility implies solvability (see, for example, [7]). The converse statement is false: there are solvable conic pairs of problems with no strictly feasible primal-dual solutions, and there are also primal-dual feasible pairs that are not solvable.

2. A primal-dual pair (P), (D) is called *primal infeasible* if there is no feasible solution for the constraints of (P). It is *strongly primal infeasible* if there exist vectors $\hat{s} \in K^*$ and $\hat{y} \in R^m$ such that

$$A^T \hat{y} + \hat{s} = 0, \quad \langle b, \hat{y} \rangle = 1.$$

We call $(\hat{y}, \hat{s})$ a *certificate* of primal infeasibility. The pair is *strictly primal infeasible* if $\hat{s}$ can be chosen in $\operatorname{int} K^*$.

3. A primal-dual pair (P), (D) is called *dual infeasible* if there is no feasible solution for the constraints of (D). It is *strongly dual infeasible* if there exists a vector $\hat{x} \in K$ such that

$$A\hat{x} = 0, \quad \langle c, \hat{x} \rangle = -1.$$

We call $\hat{x}$ a *certificate* of dual infeasibility. The pair is *strictly dual infeasible* if $\hat{x}$ can be chosen in $\operatorname{int} K$.

The feasibility patterns that are not strict we call *ill-posed*, following Renegar [14]. Note that all ill-posed problems are unstable: an arbitrary small variation of the data (A, b, c) of such a problem can change its feasibility type. In particular, an arbitrarily small perturbation of the data of a pair that is both primal and dual infeasible makes it strictly primal (or if desired dual) infeasible, and then it is easily seen to be strictly dual (or primal) feasible. It follows that the set of data instances that are both primal and dual infeasible (which stops the classification above from being a partition of all possibilities) has empty interior, and such instances are ill-posed. On the other hand, all strictly (in)feasible problems are stable. (Note that Nesterov [8] uses the term *strictly ill-posed* for a different class of problems — we will see more of them in Subsection 6.4.)

The relations between the feasibility cases for nonlinear cones are rather complicated (see [9], Section 4.2.2 for a discussion). Our goal now is to describe our possibilities in detecting these cases. This is unfortunately not straightforward. In general, what we can obtain is the following: If the pair is strictly feasible, we can generate an ϵ-optimal pair of solutions feasible in a ϵ-perturbation of the original problem and its dual. For path-following algorithms, we can do more. We can obtain the result above even if the pair is just solvable. If the pair is strongly primal (dual) infeasible, we can obtain a certificate of infeasibility for an ϵ-perturbation of the original problem and its dual. We call the latter a weak detector of infeasibility. Moreover, under suitable conditions we obtain these indications in $O(\sqrt{\nu}\ln\frac{\nu}{\epsilon})$ iterations of one of the algorithms of the previous sections. If the primal-dual pair is strictly primal or dual infeasible, we can do more: we can obtain a certificate of strict primal or dual infeasibility from a path-following algorithm, with a similar bound on the number of iterations required.

In order to simplify our considerations, we always assume that $s_0 = -F'(x_0)$ and $\tau_0 = \kappa_0 = 1$ (this implies that our starting point is $z(1)$, a point on the central path of the set $\mathcal{F}$).

The next subsection deals with the most favorable situation, where both (P) and (D) are strictly feasible. We obtain results for both potential-reduction and path-following algorithms. Subsection 6.2 analyses the output of a path-following algorithm, and is followed by a subsection summarizing what can be achieved within what number of iterations for both kinds of algorithm. We conclude with a case in which potential-reduction algorithms can yield weak detectors of infeasibility in Subsection 6.4.

Section 7 describes how path-following algorithms can generate strong detectors of infeasibility.

6.1 Strictly feasible problems

Recall that we are able to generate a point $z =: (x, y, s, \tau, \kappa) \in \mathcal{F}$ with a large value of $\mu(z)$, using any one of the path-following or potential-reduction strategies discussed in the previous sections. However, in order to form a good approximation to a solution of the initial problem, we should have the τ-component of z large enough. Indeed, in this case we can form the output $(\tilde{x}, \tilde{y}, \tilde{s})$ as follows:

$$\tilde{x} = \frac{x}{\tau}, \quad \tilde{y} = \frac{y}{\tau}, \quad \tilde{s} = \frac{s}{\tau}.$$

It is easy to see that this output satisfies the system of the linear equations of the perturbed primal-dual pair of problems:

$$A\tilde{x} = \tilde{b}, \quad A^T\tilde{y} + \tilde{s} = \tilde{c},$$

where

$$\tilde{b} = b + \frac{1}{\tau}\check{b}, \quad \tilde{c} = c + \frac{1}{\tau}\check{c}. \tag{6.1}$$

Moreover, the duality gap in this perturbed pair is small:

$$\langle \tilde{c}, \tilde{x} \rangle - \langle \tilde{b}, \tilde{y} \rangle = \langle \tilde{s}, \tilde{x} \rangle = \frac{1}{\tau^2} \langle s, x \rangle \leq \frac{\nu + 1}{\tau^2} \mu(z) \tag{6.2}$$

(see (2.19) and (3.1)). We call such a point $(\tilde{x}, \tilde{y}, \tilde{s})$ an *ϵ-solution* to the primal-dual pair (P), (D) if

$$\max\{\frac{1}{\tau}, \frac{\nu+1}{\tau^2}\mu(z)\} \leq \epsilon. \tag{6.3}$$

(We assume that $\check{b}$ and $\check{c}$ have norms of order 1.) Thus, in order to form such an output with a balance between feasibility and optimality, τ must be proportional to $\mu(z)$. Let us show that for a strictly feasible pair of problems we can guarantee this property.

Let us assume that our problems (P) and (D) satisfy Assumption 2; namely, suppose there exist points $\hat{x} \in \operatorname{int} K$, $\hat{y} \in R^m$ and $\hat{s} \in \operatorname{int} K^*$ that satisfy the following linear system:

$$Ax = b, \quad A^T y + s = c. \tag{6.4}$$

Under our assumption, for primal-dual problem (PD) there exists a central trajectory $(\hat{x}(\mu),\ \hat{y}(\mu),\ \hat{s}(\mu))$, feasible for (6.4), such that $\hat{s}(\mu) = -\mu F'(\hat{x}(\mu))$, $\mu > 0$ (see [7]). So we can choose

$$\hat{x} = \hat{x}(1), \quad \hat{y} = \hat{y}(1), \quad \hat{s} = \hat{s}(1).$$

Then

$$\hat{s} = -F'(\hat{x}), \quad \langle \hat{s}, \hat{x} \rangle = \langle c, \hat{x} \rangle - \langle b, \hat{y} \rangle = \nu. \tag{6.5}$$

Let us introduce the *feasibility measure* ρ_f as follows:

$$\rho_f := \max\{\alpha : \ \hat{x} - \alpha x_0 \in K,\ \hat{s} - \alpha s_0 \in K^*\}. \tag{6.6}$$

It is clear that $\rho_f > 0$. (However, ρ_f can be very close to zero even for well-behaved primal dual pairs if x_0 and s_0 are poorly chosen.) Moreover,

$$\langle -F'(\hat{x}), \hat{x} - \rho_f x_0 \rangle \geq 0, \quad \langle -F'(\hat{x}) + \rho_f F'(x_0), x_0 \rangle \geq 0.$$

Therefore, using these relations and (2.5) we obtain:

$$\rho_f \leq 1 \tag{6.7}$$

(it can be shown that $\rho_f = 1$ if and only if $x_0 = \hat{x}$ or equivalently $s_0 = \hat{s}$).

We will also need another inequality for ρ_f:

$$0 \leq \langle \hat{s} - \rho_f s_0, \hat{x} - \rho_f x_0 \rangle = \langle \hat{s}, \hat{x} \rangle - \rho_f(\langle s_0, \hat{x} \rangle + \langle \hat{s}, x_0 \rangle) + \rho_f^2 \langle s_0, x_0 \rangle.$$

Thus,

$$\langle s_0, \hat{x} \rangle + \langle \hat{s}, x_0 \rangle \leq \nu \left(\frac{1}{\rho_f} + \rho_f \right). \tag{6.8}$$

Theorem 9 *For any $z = (x, y, s, \tau, \kappa) \in \mathcal{F}$ the following inequality holds:*

$$\tau \geq \frac{\nu + 1}{\nu + \rho_f} \rho_f \mu(z) - \frac{1 - \rho_f}{\rho_f}. \tag{6.9}$$

Proof:
Let us fix $z = (x, y, s, \tau, \kappa) \in \mathcal{F}$. Note that

$$\begin{aligned} \langle s, \hat{x} \rangle + \langle \hat{s}, x \rangle = &\ \langle s, \hat{x} - \rho_f x_0 \rangle + \langle \hat{s} - \rho_f s_0, x \rangle + \rho_f(\langle s, x_0 \rangle + \langle s_0, x \rangle) \\ \geq &\ \rho_f(\langle s, x_0 \rangle + \langle s_0, x \rangle). \end{aligned}$$

Therefore, applying (2.23) to $(\Delta x, \Delta y, \Delta s, \Delta \tau, \Delta \kappa) := (\hat{x}, \hat{y}, \hat{s}, 1, -\nu)$, we get the following:

$$\begin{aligned} \nu\tau = &\ (\langle s, \hat{x} \rangle + \langle \hat{s}, x \rangle + \kappa) - (\langle s_0, \hat{x} \rangle + \langle \hat{s}, x_0 \rangle + 1 - \nu) \\ \geq &\ \rho_f(\langle s, x_0 \rangle + \langle s_0, x \rangle + \kappa) - (\langle s_0, \hat{x} \rangle + \langle \hat{s}, x_0 \rangle + 1 - \nu). \end{aligned}$$

Hence,

$$(\nu + \rho_f)\tau \geq \rho_f(\langle s, x_0 \rangle + \langle s_0, x \rangle + \kappa + \tau) - (\langle s_0, \hat{x} \rangle + \langle \hat{s}, x_0 \rangle + 1 - \nu).$$

It remains to note that

$$\langle s, x_0 \rangle + \langle s_0, x \rangle + \kappa + \tau = (\nu + 1)(\mu(z) + 1)$$

in view of (3.1) and use inequality (6.8). □

Thus, we have proved that if $\mu(z)$ is large, the value of τ must also be large, provided that the feasibility measure ρ_f is not too small. This result can be used in the following strategy for generating the output from infeasible-start algorithms.

Suppose we want either to find an ϵ-solution ($\epsilon < 1$) or to determine that the feasibility measure ρ_f is too small. Our answer will be positive if we manage to find a point $z \in \mathcal{F}$ such that (6.3) holds. Note that from our efficiency estimates (4.3), (4.11), (5.10), (5.12), and (5.15) for potential-reduction and path-following methods, we can guarantee that after N iterations we will have a point z such that

$$\mu(z) \geq \exp\{\frac{\theta}{\sqrt{\nu + 1}} N\}.$$

On the other hand, if the termination criterion (6.3) is not satisfied, in view of Theorem 9 we have

$$\frac{\nu + 1}{\nu + \rho_f} \rho_f \mu(z) - \frac{1 - \rho_f}{\rho_f} \leq \max \left\{ \frac{1}{\epsilon}, \sqrt{(\nu + 1)\frac{\mu(z)}{\epsilon}} \right\}.$$

This means that we can guarantee that after $O(\sqrt{\nu} \ln \frac{\nu}{\epsilon})$ iterations of our method, when $\mu(z) \geq 2(\nu + 1)/\epsilon^3$, we either find an ϵ-solution of our problem or recognize that $\rho_f \leq \epsilon$.

Recognizing that $\rho_f \leq \epsilon$ in this way can be considered as a kind of *infeasibility detector*. We will call detectors of this type *very weak* infeasibility detectors since they cannot *prove* that our initial problem is infeasible, nor do they give even approximate certificates of infeasibility; they can only recognize that the feasible set (if nonempty) has a very small relative interior. We will call an approximate certificate of infeasibility a *weak* infeasibility detector, while an exact certificate of infeasibility will be called a *strong* infeasibility detector; Section 7 shows that in some cases we can obtain such strong infeasibility detectors.

Note that the above infeasibility detector can be used in both path-following and potential-reduction schemes. In the remaining part of this section we will show that *weak* infeasibility detectors can be obtained if we follow the central path approximately. For certain infeasibility cases, we can also obtain weak infeasibility detectors for potential-reduction methods.

6.2 Some results for path-following methods

Recall that we can generate a point $z =: (x, y, s, \tau, \kappa)$ with a large value of $\mu(z)$. We assume further here that the point z belongs to a certain neighborhood of the central path:

$$\tau\kappa \geq \omega\mu(z), \tag{6.10}$$

where $\omega \in (0, 1]$ is an absolute constant. In this case we can describe in more detail the behavior of τ and κ based on the classification of our primal-dual pair.

We did not describe path-following schemes dealing directly with this neighborhood, since their efficiency estimates are worse than those of the methods that exploit the functional proximity measure. However, let us demonstrate that all points from a neighborhood of the central path defined by a bound on the functional proximity satisfy the inequality (6.10) for a certain value of ω. (It is also easy to see that iterates in the narrow neighborhood $\mathcal{N}(\beta)$ of Subsection 5.4 satisfy (6.10) for $\omega = 1 - \beta$.)

Lemma 3 *Let $z = (x, y, s, \tau, \kappa) \in \mathcal{F}$ and $\Omega(z) \leq \Delta$. Then z satisfies inequality (6.10) with $\omega = \xi(\Delta) + 1$, where $\xi(\Delta)$ is the smaller root of the equation*

$$\xi - \ln(1 + \xi) = \Delta.$$

Proof:
In view of the definition of $\Omega(z)$ and inequality (2.13) we have:

$$\begin{aligned}
\Omega(z) &= F(x) + F_*(s) - \ln\tau - \ln\kappa - 1 + (\nu + 1)[\ln\mu(z) + 1] \\
&\geq -\nu + \nu\ln\nu - \nu\ln\langle s, x\rangle - \ln\tau - \ln\kappa - 1 + (\nu + 1)\,[\ln\mu(z) + 1] \\
&= \nu\ln\nu - \nu\ln\langle s, x\rangle - \ln\tau\kappa + (\nu + 1)\ln\mu(z).
\end{aligned}$$

Denote $\xi = \frac{\tau\kappa}{\mu(z)} - 1 > -1$. Then

$$\begin{aligned}
\mu(z) &= \tfrac{\tau\kappa}{1+\xi}, \\
\langle s, x\rangle &= \tfrac{(\nu+1)\tau\kappa}{1+\xi} - \tau\kappa = \tfrac{\nu-\xi}{1+\xi}\tau\kappa.
\end{aligned}$$

Therefore

$$\begin{aligned}
\Omega(z) &\geq \nu\ln\nu - \nu\ln\tfrac{\nu-\xi}{1+\xi} + (\nu + 1)\ln\tfrac{1}{1+\xi} \\
&= -\nu\ln\left(1 - \tfrac{1}{\nu}\xi\right) - \ln(1 + \xi) \geq \xi - \ln(1 + \xi).
\end{aligned}$$

It remains to note that the function $\xi - \ln(1+\xi)$ is decreasing and convex for negative ξ.
□

Theorem 10 *Suppose that the point $z = (x, y, s, \tau, \kappa) \in \mathcal{F}$ satisfies inequality (6.10).*

1. *If the primal-dual pair (P), (D) is solvable, then*

$$\kappa \leq \langle s_0, x^* \rangle + \langle s^*, x_0 \rangle + 1,$$

$$\tau \geq \tfrac{\omega\mu(z)}{\langle s_0, x^* \rangle + \langle s^*, x_0 \rangle + 1},$$

where (x^*, y^*, s^*) is an arbitrary solution of the pair.

2. *If the primal-dual pair (P), (D) is strongly primal infeasible, then*

$$\kappa \geq \tfrac{\omega\mu(z)}{\langle \hat{s}, x_0 \rangle + 1},$$

$$\tau \leq \langle \hat{s}, x \rangle + \tau = \langle \hat{s}, x_0 \rangle + 1,$$

where $(\hat{y}, \hat{s} \in K^*)$ is an arbitrary certificate of primal infeasibility.

3. *If the primal-dual pair (P), (D) is strongly dual infeasible, then*

$$\kappa \geq \tfrac{\omega\mu(z)}{\langle s_0, \hat{x} \rangle + 1},$$

$$\tau \leq \langle s, \hat{x} \rangle + \tau = \langle s_0, \hat{x} \rangle + 1,$$

where $\hat{x} \in K$ is an arbitrary certificate of dual infeasibility.

Proof:
1. Note that in this case the direction

$$(\Delta x, \Delta y, \Delta s, \Delta\tau \Delta\kappa) := (x^*, y^*, s^*, 1, 0)$$

is a recession direction of the set $\mathcal{F}$. Therefore, in view of equality (2.23) we have:

$$\langle s, x^* \rangle + \langle s^*, x \rangle + \kappa = \langle s_0, x^* \rangle + \langle s^*, x_0 \rangle + 1.$$

Since $x^* \in K$ and $s^* \in K^*$, we get the upper bound for κ. This inequality leads to the lower bound for τ in view of (6.10).

2. If the primal problem (2.1) is strongly infeasible, the direction

$$(\Delta x, \Delta y, \Delta s, \Delta\tau \Delta\kappa) := (0, \hat{y}, \hat{s}, 0, 1)$$

is a recession direction of the set $\mathcal{F}$. Therefore, in view of equality (2.23) we have:

$$\langle \hat{s}, x \rangle + \tau = \langle \hat{s}, x_0 \rangle + 1.$$

Since $\hat{s} \in K^*$, we get the upper bound for τ. This inequality leads to the lower bound for κ in view of (6.10).

3. The proof of the last statement is similar to the previous one. □

6.3 Complexity of obtaining weak detectors

We now use the theorem above to investigate how path-following methods can generate weak detectors.

Note that the first statement of the theorem can provide us with an alternative indirect (very weak) certificate of primal-dual infeasibility. Indeed, proceeding as in Subsection 6.1, we can guarantee that after $O(\sqrt{\nu}\ln\frac{\nu}{\epsilon})$ iterations of our method, when $\mu(z) \geq (\nu+1)^3/(\omega^2\epsilon^3)$, we either find an ϵ-solution of our primal-dual pair or recognize that every solution (x^*, y^*, s^*) must have $\langle s_0, x^*\rangle + \langle s^*, x_0\rangle + 1 > (\nu+1)/\epsilon$. This result is similar to that of Subsection 6.1. However, the very weak infeasibility detector described there can be applied also to a potential reduction scheme. Until the statement of the final result, all detectors presented in this subsection can be guaranteed only with a path-following approach.

Now suppose that the primal-dual pair is either strongly primal or strongly dual infeasible. Then the second or third part of the theorem above shows that any path-following method that generates an iterate $z \in \mathcal{F}$ with a large value of $\mu(z)$ also gives a large value for κ. If we then scale this solution by κ, we obtain a solution $(\tilde{x}, \tilde{y}, \tilde{s}) \in K \times R^m \times K^*$ to

$$\begin{array}{rcccl} & & A\tilde{x} & & = \frac{\tau}{\kappa}b + \frac{1}{\kappa}\check{b} \\ A^T\tilde{y} & & & + \;\tilde{s} & = \frac{\tau}{\kappa}c + \frac{1}{\kappa}\check{c} \\ -\langle b, \tilde{y}\rangle & + & \langle c, \tilde{x}\rangle & & = -1 + \frac{1}{\kappa}\check{g}. \end{array}$$

Let us assume that $\check{g}$ and the norms of b, c, $\check{b}$, and $\check{c}$ are of order 1. Then, for κ large enough (greater than $3\check{g}$), we have either

$$A^T\tilde{y} + \tilde{s} = \frac{\tau}{\kappa}b + \frac{1}{\kappa}\check{b}, \quad \langle b, \tilde{y}\rangle \geq 1/3, \quad \tilde{s} \in K^*, \tag{6.11}$$

or

$$A\tilde{x} = \frac{\tau}{\kappa}c + \frac{1}{\kappa}\check{c}, \quad \langle c, \tilde{x}\rangle \leq -1/3, \quad \tilde{x} \in K. \tag{6.12}$$

If

$$\kappa \geq \max\{1/\epsilon, \tau/\epsilon\} \tag{6.13}$$

so that τ/κ and $1/\kappa$ are at most ϵ, we call a solution to (6.11) an *ϵ-certificate* of primal infeasibility, and one to (6.11) an *ϵ-certificate* of dual infeasibility. (We assume that $\epsilon \leq (3\check{g})^{-1}$, so that (6.13) implies that κ is large enough.) Any ϵ-certificate of primal infeasibility gives an indication of approximate primal infeasibility. Indeed, in this case, $(\tilde{A}, b, c)$, with $\tilde{A} = A + bv^T$ for some v of order ϵ, is an infeasible instance with $(\tilde{y}, \tilde{s})$ an exact certificate of primal infeasibility. Moreover, if $\eta s_0 - A^T\tilde{y} - \tilde{s} \in K^*$, then if the primal problem has any feasible solution x, we must have $\langle s_0, x\rangle \geq (3\eta)^{-1}$. (This follows by applying strong duality to $\max\{\langle b, y\rangle : A^Ty + s = \eta s_0, s \in K^*\}$.) Similar statements hold in the second case. In either case, we say we have a weak infeasibility detector.

Suppose that we apply a path-following algorithm which generates iterates satisfying (6.10) and after N iterations guarantees that

$$\mu(z) \geq \exp\{\frac{\theta}{\sqrt{\nu+1}}N\}. \tag{6.14}$$

Then after $O(\sqrt{\nu}\ln\frac{\nu}{\epsilon})$ iterations we will have $\mu(z) \geq (\nu+1)^2/(\omega\epsilon^3)$. Hence, if there is a certificate $(\hat{y}, \hat{s})$ of primal infeasibility with $\langle \hat{s}, x_0\rangle + 1 \leq (\nu+1)/\epsilon$ or a certificate $\hat{x}$ of dual infeasibility with $\langle s_0, \hat{x}\rangle + 1 \leq (\nu+1)/\epsilon$, then the theorem implies that the termination criterion (6.13) will hold.

Let us summarize the results we have obtained here and in Subsection 6.1. In all cases, our algorithms are assumed to generate iterates satisfying (6.14) after N iterations; the iterates of path-following algorithms are required further to satisfy (6.10). Then we have:

Theorem 11 *1. Suppose the primal-dual pair is strictly feasible. Then a path-following or potential-reduction algorithm either generates an ϵ-solution to the pair within*

$$O(\sqrt{\nu}\ln\frac{\nu}{\epsilon})$$

iterations or demonstrates that $\rho_f \leq \epsilon$, where ρ_f is defined in (6.6).

2. Suppose the primal-dual pair is solvable. Then a path-following algorithm either generates an ϵ-solution to the pair within

$$O(\sqrt{\nu}\ln\frac{\nu}{\epsilon})$$

iterations or demonstrates that any solution (x^*, y^*, s^*) must have $\langle s_0, x^*\rangle + \langle s^*, x_0\rangle > (\nu+1)/\epsilon - 1$.

3. Suppose the primal-dual pair is strongly primal infeasible. Then a path-following algorithm either generates an ϵ-certificate of primal or dual infeasibility for the pair within

$$O(\sqrt{\nu}\ln\frac{\nu}{\epsilon})$$

iterations or demonstrates that any certificate $(\hat{y}, \hat{s})$ of primal infeasibility has $\langle \hat{s}, x_0\rangle \geq (\nu+1)/\epsilon - 1$.

4. Suppose the primal-dual pair is strongly dual infeasible. Then a path-following algorithm either generates an ϵ-certificate of primal or dual infeasibility for the pair within

$$O(\sqrt{\nu}\ln\frac{\nu}{\epsilon})$$

iterations or demonstrates that any certificate $\hat{x}$ of dual infeasibility has $\langle s_0, \hat{x}\rangle \geq (\nu+1)/\epsilon - 1$.

□

To conclude, we note that results of Renegar [13] indicate that if a primal-dual pair is solvable, and its data (A, b, c) is at least a distance η from ill-posedness (i.e., from primal or dual infeasible instances), then there are optimal solutions with norm at most $O(\eta^{-2})$ if the data is appropriately scaled; this relates to part (2) of the theorem above. Similarly, if the pair is primal (or dual) infeasible, and its data is at least a distance η from ill-posedness (i.e., from primal (or dual) feasible instances), then there is a certificate of infeasibility with norm at most $O(\eta^{-1})$ if the data is appropriately scaled; this relates to parts (3) and (4) of the theorem.

6.4 Negative-τ solutions

Note that there is an asymmetry between potential-reduction and path-following methods in the previous subsections. For the former, we can generate ϵ-solutions in the strictly feasible case, but when the primal or dual is infeasible, we may not be able to generate a weak infeasibility detector; for path-following methods, both indications are achievable. Here we briefly note a class of problems (these are exactly the problems termed *strictly ill-posed* in [8]) where any algorithm that generates $z \in \mathcal{F}$ with large $\mu(z)$ provides a weak infeasibility detector.

Suppose we have a solution to

$$\begin{array}{rcccccccl} & & A\tilde{x} & & & & & = & \tilde{\tau}b \\ A^T\tilde{y} & & & + & \tilde{s} & & & = & \tilde{\tau}c \\ -\langle b, \tilde{y}\rangle & + & \langle c, \tilde{x}\rangle & & & + & \tilde{\kappa} & = & 0 \end{array}$$

with $\tilde{x} \in \operatorname{int} K$, $y \in R^m$, $\tilde{s} \in \operatorname{int} K^*$, $\tilde{\tau} < 0$, and $\tilde{\kappa} > 0$. By scaling, we can assume that $\tilde{\tau} = -1$. We call this a *negative-τ* solution. Note that in this case the primal and dual problems cannot both have feasible solutions. For if $\hat{x}$ and $(\hat{y}, \hat{s})$ were primal and dual feasible, then $(x, y, s) := (\tilde{x}, \tilde{y}, \tilde{s}) + (\hat{x}, \hat{y}, \hat{s})$ would satisfy

$$Ax = 0, \quad A^T y + s = 0, \quad x \in \operatorname{int} K, \quad s \in \operatorname{int} K^*,$$

a contradiction (since $\langle s, x\rangle$ would have to be both positive and zero). Unfortunately, such a negative-τ solution does not yield a certificate of either primal or dual infeasibility. Nevertheless, we can show that if such a solution exists, then any $z \in \mathcal{F}$ with large $\mu(z)$ has a large value for κ and hence yields a weak infeasibility detector.

Suppose we have a negative-τ solution. Then the dual conic problems defined by replacing b and c by their negatives are strictly feasible, so the corresponding central path exists. Let us choose $(\hat{x}, \hat{y}, \hat{s})$ as the point on this path corresponding to the parameter value 1. Then $\hat{s} = -F'(\hat{x})$ and $(\Delta x, \Delta y, \Delta s, \Delta\tau, \Delta\kappa) := (\hat{x}, \hat{y}, \hat{s}, -1, \nu)$ satisfies (2.21) and we may therefore apply (2.23). Now proceeding as in the proof of Theorem 9 we find that any $z = (x, y, s, \tau, \kappa) \in \mathcal{F}$ satisfies

$$\kappa \geq \frac{\rho_n}{1 + \rho_n}(\nu + 1)\mu - \frac{\nu - \rho_n}{\rho_n},$$

where ρ_n is defined from $\hat{x}$, x_0, $\hat{s}$, and s_0 as is ρ_f in (6.6). (We use a subscript "n" to denote that $(\hat{x}, \hat{y}, \hat{s})$ is a negative-τ solution.) Hence generating a point in $\mathcal{F}$ with a large value of $\mu(z)$ guarantees that κ is large and hence that we can obtain a weak infeasibility detector.

7 Strong detectors for strictly infeasible problems

Here we show that path-following methods can generate strong infeasibility detectors.

7.1 Strict primal infeasibility

Let us assume that our primal problem (P) is strictly infeasible: there exist $\hat{s} \in \operatorname{int} K^*$ and $\hat{y} \in R^m$ such that

$$A^T\hat{y} + \hat{s} = 0, \quad \langle b, \hat{y}\rangle = 1. \tag{7.1}$$

In order to describe the behavior of path-following methods in this situation, it is useful to introduce the following auxiliary problem:

$$\begin{aligned} \min_{x,\tau} \quad & F(x) - \ln \tau \\ \text{s.t.} \quad & Ax - \tau b = \check{b}, \\ & x \in K, \ \tau \geq 0. \end{aligned} \tag{7.2}$$

Since $\hat{s} \in \operatorname{int} K^*$, the feasible set of the problem (7.2) is bounded (see Theorem 10). And since the function $F(x) - \ln \tau$ is strictly self-concordant on this set, the solution of the above problem exists and is unique. We denote it $(\hat{x}, \hat{\tau})$. Let $\hat{y}$ denote the Lagrange multiplier for the equality constraint in (7.2) and let $\hat{s} := -F'(\hat{x})$, $\hat{\kappa} := 1/\hat{\tau}$. Then $A^T\hat{y} + \hat{s} = 0$, $\hat{s} \in \operatorname{int} K^*$, and $\langle b, \hat{y}\rangle = \hat{\kappa} > 0$; thus we recover a certificate of strict primal infeasibility.

It is convenient to use the pair $(\hat{x}, \hat{\tau})$ in order to measure the level of infeasibility of our primal problem. Let us introduce the following *primal infeasibility* measure:

$$\rho_p := \max\{\alpha : \ x_0 - \alpha\hat{x} \in K, \ \tau_0 - \alpha\hat{\tau} \geq 0\}.$$

Intuitively, if the problem (P) is "almost" feasible, then $\hat{x}$ and $\hat{\tau}$ should be large. (Indeed, if $\bar{x} \in \operatorname{int} K$ satisfies $A\bar{x} = b + \sigma\check{b}$ for some small positive σ, then $\tau := 1/\sigma$ and $x := \tau\bar{x}$ are feasible in (7.2) and large.) Consequently, the feasible α's in the definition above are small. Thus, ρ_p should be small for problems that are close to being primal feasible.

In case F is a self-scaled barrier for K, we can use Lemma 3.3 of [11] (applied to $x = \bar{x}_0$, $s = \bar{\hat{s}}$, and $K := \bar{K}$) to get an alternative definition:

$$\rho_p := \max\{\alpha : \ \hat{s} - \alpha s_0 \in K^*, \ \hat{\kappa} - \alpha\kappa_0 \geq 0\}$$

(compare with (6.6)). Again, this provides intuition for ρ_p being small for problems that are close to being primal feasible, for then $\hat{s}$ should be close to the boundary of K^* or $\hat{\kappa}$ close to zero.

Note that in view of Proposition 2 $\rho_p \leq 1$.

In what follows we will need some estimates of the variation of a linear functional over the feasible set of (7.2) in terms of the primal infeasibility measure. It is natural to measure the norm of the linear functional using the Hessian of the barrier at (x_0, τ_0). For any pair $(p, \sigma) \in R^n \times R$ denote

$$\| (p, \sigma) \|^*_{(x_0,\tau_0)} := [\langle F''(x_0)^{-1}p, p\rangle + \tau_0^2\sigma^2]^{1/2}.$$

Lemma 4 *Let $(p, \sigma) \in R^n \times R$. Then for any (x, τ) feasible for (7.2) we have:*

$$| \langle p, x - \hat{x}\rangle + \sigma(\tau - \hat{\tau}) | \leq \frac{\nu(\nu + 1)}{\rho_p} \| (p, \sigma) \|^*_{(x_0,\tau_0)} . \tag{7.3}$$

Proof:
Since $(\hat{x}, \hat{\tau})$ is the minimum of the barrier $\bar{F}(x, \tau) := F(x) - \ln \tau$ over a linear subspace, we have:

$$\langle \bar{F}'(\hat{x}, \hat{\tau}), (x, \tau) - (\hat{x}, \hat{\tau})\rangle = 0.$$

Therefore (7.3) follows from Proposition 2 in the appendix since the parameter of the barrier $\bar{F}(x,\tau)$ is equal to $\nu+1$. $\square$

Now we can prove that in the strictly primal infeasible case the components $(x(\mu),\tau(\mu))$ of the central path converge to the point $(\hat{x},\hat{\tau})$.

Theorem 12 *If the primal problem (P) is strictly infeasible, then for any $\mu > 0$ the following inequality holds:*

$$F(x(\mu)) - \ln\tau(\mu) - [F(\hat{x}) - \ln\hat{\tau}] \le \frac{c_p(\nu,\rho_p)}{\mu}, \tag{7.4}$$

where

$$\begin{aligned} c_p(\nu,\rho_p) \quad &= \quad \frac{\nu(\nu+1)}{\rho_p^2}(\rho_p[\langle F''(x_0)^{-1}(A^T y_0 + s_0),(A^T y_0+s_0)\rangle + \tau_0^2(\kappa_0 - \langle b, y_0\rangle)^2]^{1/2} \\ &\quad +2\nu(\nu+1)\tau_0\langle [F''(x_0)]^{-1}c,c\rangle^{1/2}). \end{aligned}$$

Proof:
Recall that $\bar{F}(x,\tau)$ denotes the barrier $F(x)-\ln\tau$ with dual $\bar{F}_*(s,\kappa) = F_*(s) - \ln\kappa - 1$. Since the parameter of the barrier $\bar{F}$ is equal to $\bar{\nu} := \nu+1$, in view of (2.13) we have:

$$\begin{aligned} F_*(s(\mu)) - \ln\kappa(\mu) - 1 \quad &= \bar{F}_*(s,\kappa) \\ &\ge -\bar{\nu} - \bar{\nu}\ln\left(\tfrac{1}{\bar{\nu}}[\langle s(\mu),\hat{x}\rangle + \kappa(\mu)\hat{\tau}]\right) - F(\hat{x}) + \ln\hat{\tau}. \end{aligned}$$

On the other hand, (2.29) implies

$$-\bar{\nu}\ln\mu - \bar{\nu} = F(x(\mu)) - \ln\tau(\mu) + F_*(s(\mu)) - \ln\kappa(\mu) - 1.$$

Combining these two relations, we obtain:

$$\begin{aligned} F(x(\mu)) - \ln\tau(\mu) - [F(\hat{x}) - \ln\hat{\tau}] \quad &\le \quad \bar{\nu}\ln\tfrac{\langle s(\mu),\hat{x}\rangle + \kappa(\mu)\hat{\tau}}{\bar{\nu}\mu} \\ &= \quad \bar{\nu}\ln\left(1 + \tfrac{\langle s(\mu),\hat{x}-x(\mu)\rangle + \kappa(\mu)(\hat{\tau}-\tau(\mu))}{\bar{\nu}\mu}\right) \\ &\quad\quad \text{(we have used (2.27) and (2.5))} \\ &\le \quad \tfrac{1}{\mu}[\langle s(\mu),\hat{x}-x(\mu)\rangle + \kappa(\mu)(\hat{\tau}-\tau(\mu))]. \end{aligned}$$

Denote $\Delta x := \hat{x} - x(\mu)$, $\Delta\tau := \hat{\tau} - \tau(\mu)$. Clearly, $A\Delta x = \Delta\tau b$. Therefore

$$\begin{aligned} \langle s(\mu),\Delta x\rangle + \kappa(\mu)\Delta\tau \quad &= \langle \check{c} + \tau(\mu)c - A^T y(\mu),\Delta x\rangle + \kappa(\mu)\Delta\tau \\ &= \langle \check{c} + \tau(\mu)c,\Delta x\rangle - \langle A\Delta x, y(\mu)\rangle + \kappa(\mu)\Delta\tau \\ &= \langle \check{c} + \tau(\mu)c,\Delta x\rangle + \Delta\tau(\kappa(\mu) - \langle b, y(\mu)\rangle) \\ &= \langle \check{c} + \tau(\mu)c,\Delta x\rangle + \Delta\tau(\check{g} - \langle c, x(\mu)\rangle) \end{aligned}$$

Thus,

$$F(x(\mu)) - \ln \tau(\mu) - [F(\hat{x}) - \ln \hat{\tau}] \leq \frac{1}{\mu}[\langle \check{c} + \tau(\mu)c, \Delta x\rangle + \Delta\tau(\check{g} - \langle c, x(\mu)\rangle)].$$

It remains to estimate the quantity in the brackets.

$$\begin{aligned}
&\langle \check{c} + \tau(\mu)c, \hat{x} - x(\mu)\rangle + (\hat{\tau} - \tau(\mu))(\check{g} - \langle c, x(\mu)\rangle)\\
&= \langle A^T y_0 + s_0 + (\tau(\mu) - \tau_0)c, \hat{x} - x(\mu)\rangle + (\hat{\tau} - \tau(\mu))(\kappa_0 - \langle b, y_0\rangle + \langle c, x_0 - x(\mu)\rangle)\\
&= \langle A^T y_0 + s_0, \hat{x} - x(\mu)\rangle + (\kappa_0 - \langle b, y_0\rangle)(\hat{\tau} - \tau(\mu))\\
&\quad +(\tau(\mu) - \tau_0)\langle c, \hat{x} - x(\mu)\rangle + (\hat{\tau} - \tau(\mu))\langle c, x_0 - x(\mu)\rangle\\
&= \langle A^T y_0 + s_0, \hat{x} - x(\mu)\rangle + (\kappa_0 - \langle b, y_0\rangle)(\hat{\tau} - \tau(\mu)) + (\hat{\tau} - \tau_0)\langle c, \hat{x} - x(\mu)\rangle\\
&\quad +(\hat{\tau} - \tau(\mu))\langle c, x_0 - \hat{x}\rangle.
\end{aligned}$$

Note that both points (x_0, τ_0) and $(x(\mu), \tau(\mu))$ belong to the feasible set of problem (7.2). Therefore, using Lemma 4, we obtain:

$$\mid \langle A^T y_0 + s_0, \hat{x} - x(\mu)\rangle + (\kappa_0 - \langle b, y_0\rangle)(\hat{\tau} - \tau(\mu)) \mid \leq \tfrac{\nu(\nu+1)}{\rho_p} \parallel (A^T y_0 + s_0, \kappa_0 - \langle b, y_0\rangle) \parallel^*_{(x_0,\tau_0)},$$

$$\mid \hat{\tau} - \tau_0 \mid \leq \tfrac{\nu(\nu+1)}{\rho_p}\tau_0,$$

$$\mid \hat{\tau} - \tau(\mu) \mid \leq \tfrac{\nu(\nu+1)}{\rho_p}\tau_0,$$

$$\mid \langle c, \hat{x} - x(\mu)\rangle \mid \leq \tfrac{\nu(\nu+1)}{\rho_p}\langle [F''(x_0)]^{-1}c, c\rangle^{1/2},$$

$$\mid \langle c, \hat{x} - x_0\rangle \mid \leq \tfrac{\nu(\nu+1)}{\rho_p}\langle [F''(x_0)]^{-1}c, c\rangle^{1/2}.$$

Substituting these estimates in the previous inequality, we obtain (7.4). □

Note that the result of Theorem 12 can be used for constructing a *strong* infeasibility detector. Indeed, given the point $z = (x, y, s, \kappa, \tau) \in \mathcal{F}$, we can compute the value:

$$\begin{aligned}
\sigma_p(z) = \max_{\Delta x, \Delta\tau} \quad & \langle F'(x), \Delta x\rangle - \tfrac{1}{\tau}\Delta\tau\\
\text{s.t.} \quad & A\Delta x - \Delta\tau b = 0,\\
& \langle F''(x)\Delta x, \Delta x\rangle + \left(\tfrac{\Delta\tau}{\tau}\right)^2 \leq 1.
\end{aligned} \tag{7.5}$$

If $\sigma_p(z) < 1$, we conclude that the solution $(\hat{x}, \hat{\tau})$ of the problem (7.2) exists and therefore our primal problem (P) is strictly infeasible (see [9], Theorem 2.2.2). Indeed, we can easily generate a certificate of strict primal infeasibility from the Lagrange multipliers $\hat{y}$, $\hat{\kappa}$ for problem (7.5). We find that

$$A^T\hat{y} - F'(x) + 2\hat{\kappa}F''(x)\Delta x^* = 0,$$

$$-\langle b, \hat{y}\rangle + \frac{1}{\tau} + 2\hat{\kappa}\frac{\Delta\tau^*}{\tau^2} = 0.$$

From this we easily obtain $2\hat{\kappa} = \sigma_p(z) < 1$. Now the constraints of (7.5) imply that

$$|\frac{\Delta\tau^*}{\tau}| < 1, \quad \langle F''(x)\Delta x^*, [F''(x)]^{-1}F''(x)\Delta x^*\rangle < 1,$$

so we conclude that $\langle b, \hat{y}\rangle > 0$ and $\hat{s} := -F'(x) + 2\hat{\kappa}F''(x)\Delta x^* \in \operatorname{int} K^*$, so that $(\hat{y}, \hat{s})$ is a certificate of strict primal infeasibility.

Let us estimate the complexity of generating such a detector. We give only a sketch of the proof, based on the theory of self-concordant functions ([9]).

Note that $\sigma_p(z)$ is the local norm of the gradient of the restriction of the self-concordant barrier $\bar{F}(x,\tau)$ to the feasible set of the problem (7.2).

Let us follow the central path $z(\mu)$ in a tight neighborhood, for example, let $\Omega(z) \leq \Delta$ for Δ small enough. Then the distance between the point z and $z(\mu)$, $\mu = \mu(z)$, measured by the Hessian of the restriction of the barrier $F(x) + F_*(s) - \ln\tau - \ln\kappa$ at $z(\mu)$ to $\mathcal{F}$, is small. This implies that the distance between (x,τ) and $(x(\mu),\tau(\mu))$ measured by $\bar{F}''(x(\mu),\tau(\mu))$ is even smaller. Hence, the local norm of $\bar{F}'(x,\tau)$ is proportional to the local norm of $\bar{F}'(x(\mu),\tau(\mu))$. It remains to note that the local norm of $\bar{F}'(x(\mu),\tau(\mu))$ will be small if the right-hand side of (7.4) is less then a small absolute constant.

Thus, we conclude that, using a path-following approach, it is possible to obtain a certificate of strict primal infeasibility in finite time, namely, in $O(\sqrt{\nu}\ln\frac{\nu}{\rho_p})$ iterations.

7.2 Strict dual infeasibility

The analysis of the strict dual infeasibility is completely symmetric to the analysis of the previous subsection. Therefore we omit the proofs.

Let us assume that our dual problem (D) is strictly infeasible: there exists $\hat{x} \in \operatorname{int} K$ such that

$$A\hat{x} = 0, \quad \langle c, \hat{x}\rangle = -1. \tag{7.6}$$

Consider the following auxiliary problem:

$$\begin{aligned} \min_{y,s,\tau} \quad & F_*(s) - \ln\tau \\ \text{s.t.} \quad & A^T y + s - \tau c = \check{c}, \\ & y \in R^m,\ s \in K^*,\ \tau \geq 0. \end{aligned} \tag{7.7}$$

Since $\hat{x} \in \operatorname{int} K$, the feasible set of the problem (7.7) is bounded (see Theorem 10). And since the function $F_*(s) - \ln\tau$ is strictly self-concordant on this set, the solution of the above problem exists and is unique. Indeed, its Lagrange multiplier gives a certificate of strict dual infeasibility.

Let us use the pair $(\hat{s}, \hat{\tau})$ in order to measure the level of infeasibility of the dual problem. Define the following *dual infeasibility* measure:

$$\rho_d := \max\{\alpha :\ s_0 - \alpha\hat{s} \in K^*,\ \tau_0 - \alpha\hat{\tau} \geq 0\}.$$

As in the previous subsection, ρ_d should be small for problems that are close to being dual feasible. Note that in view of Proposition 2 in the appendix $\rho_d \leq 1$.

Theorem 13 *If the dual problem (D) is strictly infeasible, then for any $\mu > 0$ the following inequality holds:*

$$F_*(s(\mu)) - \ln\tau(\mu) - [F_*(\hat{s}) - \ln\hat{\tau}] \leq \frac{c_d(\nu,\rho_d)}{\mu}, \tag{7.8}$$

where $c_d(\nu,\rho_d) = O\left(\frac{\nu^4}{\rho_d^2}\right)$.

Thus, in the strictly dual infeasible case the components $(s(\mu),\tau(\mu))$ of the central path converge to the point $(\hat{s},\hat{\tau})$. Therefore, the strict dual infeasibility detector is as follows. We compute the value

$$\begin{array}{rll} \sigma_d(z) = & \max\limits_{\Delta y,\Delta s,\Delta\tau} & \langle \Delta s, F_*'(s)\rangle - \frac{1}{\tau}\Delta\tau \\ & \text{s.t.} & A^T\Delta y + \Delta s + \Delta\tau c = 0, \\ & & \langle \Delta s, F_*''(s)\Delta s\rangle + \left(\frac{\Delta\tau}{\tau}\right)^2 \leq 1. \end{array} \tag{7.9}$$

If $\sigma_d(z) < 1$, we conclude that the solution $(\hat{s},\hat{\tau})$ of the problem (7.7) exists and therefore our dual problem (D) is strictly infeasible. We can obtain a certificate of strict dual infeasibility from the Lagrange multipliers of (7.9). The complexity estimate of this detector is $O(\sqrt{\nu}\ln\frac{\nu}{\rho_d})$ iterations of a path-following scheme.

Appendix. A result in the theory of self-concordant functions

Let $H \subset R^q$ be a closed pointed cone with nonempty interior, and let $\Phi(x)$ be a ν-logarithmically homogeneous barrier for H. For $x \in \operatorname{int} H$ and $h,\, p \in R^q$ denote

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

We first prove a slight extension of Lemma 8 in [8]:

Lemma 5 *Let $x_0 \in \operatorname{int} H$ and suppose $\nu \geq 2$. Then if $x \in K$ satisfies $\langle \Phi'(x_0), x - x_0\rangle \geq 0$, we have*

$$\| x - x_0 \|_{x_0} \leq \sqrt{\nu(\nu-1)}.$$

Proof:
Any such x is a convex combination of 0 and a point in $\{x \in K : \langle \Phi'(x_0), x - x_0\rangle = 0\}$. For every point in the latter set the desired inequality holds by Lemma 8 in [8]. Also,

$$\| 0 - x_0 \|_{x_0} = \sqrt{\nu} \leq \sqrt{\nu(\nu-1)}$$

as long as $\nu \geq 2$. □

For $x_0,\, \hat{x} \in \operatorname{int} H$ denote

$$\rho(x_0,\hat{x}) := \max\{\alpha :\ x_0 - \alpha\hat{x} \in H\}.$$

Proposition 2 *Let $\langle \Phi'(\hat{x}), x_0 - \hat{x}\rangle = 0$. Then $\rho(x_0,\hat{x}) \le 1$ and for any $h \in R^q$ we have:*

$$\| h \|_{x_0} \le \frac{\sqrt{\nu(\nu-1)}}{\rho(x_0,\hat{x})} \| h \|_{\hat{x}} . \tag{7.1}$$

Moreover, for any $x \in K$ with $\langle \Phi'(\hat{x}), x - \hat{x}\rangle = 0$ and any $p \in R^q$ we have:

$$| \langle p, x - \hat{x}\rangle | \le \frac{\nu(\nu-1)}{\rho(x_0,\hat{x})} \| p \|^*_{x_0} . \tag{7.2}$$

Proof:
Denote $\rho := \rho(x_0,\hat{x})$. By the condition of the lemma we have

$$0 \le \langle -\Phi'(\hat{x}), x_0 - \rho\hat{x}\rangle = \nu(1-\rho).$$

Thus $\rho \le 1$. Assume temporarily that $\rho < 1$, and let $y := x_0 + \gamma(x_0 - \hat{x})$, with $\gamma := \rho/(1-\rho)$. Then $y \in H$ in view of the definition of ρ.

Consider the ellipsoids

$$W_\beta := \{u = y + \beta(\hat{x} - y) + h : \| h \|_{\hat{x}} \le \beta\}, \quad \beta \in [0,1].$$

Since $y \in H$ and $W_1 \subset H$, we conclude that $W_\beta \subset H$ for any $\beta \in [0,1]$.

Let us choose $\beta = \rho$. Then

$$y + \beta(\hat{x} - y) = x_0 + (\gamma - \beta(1+\gamma))(x_0 - \hat{x}) = x_0.$$

Thus, the ellipsoid $W_\rho \subset H$ is centered at x_0. Therefore, in view of Lemma 5, we have:

$$\{u = x_0 + h : \ \langle \Phi'(x_0), h\rangle \ge 0, \ \| h \|_{\hat{x}} \le \rho\} \subseteq \{u = x_0 + h : \| h \|_{x_0} \le \sqrt{\nu(\nu-1)}\}.$$

This implies (since we can replace h by its negative if necessary to satisfy $\langle \Phi'(x_0), h\rangle \ge 0$) that for any $h \in R^q$ we have:

$$\| h \|_{x_0} \le \frac{1}{\rho}\sqrt{\nu(\nu-1)} \| h \|_{\hat{x}}$$

and this is (7.1). If $\rho = 1$, then the argument above can be repeated with $\gamma := \rho'/(1-\rho')$, where $\rho' := \rho - \epsilon$ and ϵ is positive and arbitrarily small. This gives the inequality above with ρ' replacing ρ. Since ϵ was arbitrary, (7.1) follows.

Finally, if $x \in K$ and $\langle \Phi'(\hat{x}), x - \hat{x}\rangle = 0$ then

$$\| x - \hat{x} \|_{\hat{x}} \le \sqrt{\nu(\nu-1)}$$

(see [8], Lemma 8, or Lemma 5). Therefore, in view of (7.1), we have

$$\begin{aligned} \langle p, x - \hat{x}\rangle \quad & \le \| p \|^*_{x_0} \| x - \hat{x} \|_{x_0} \\ & \le \tfrac{1}{\rho}\sqrt{\nu(\nu-1)} \| x - \hat{x} \|_{\hat{x}} \| p \|^*_{x_0} \le \tfrac{\nu(\nu-1)}{\rho} \| p \|^*_{x_0} . \end{aligned}$$

□

References

[1] E. D. Andersen and Y. Ye, "On a homogeneous algorithm for the monotone complementarity problem," Working Paper, Department of Management Science, The University of Iowa, 1995.

[2] G. de Ghellink and J.-P. Vial, "A Polynomial Newton Method for Linear Programming," Algorithmica, 1 (1986), pp. 425-453.

[3] D. den Hertog, *Interior Point Approach to Linear, Quadratic and Convex Programming: Algorithms and Complexity.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[4] F. Jarre. Interior-point methods via self-concordance or relative Lipschitz condition. *Optimization Methods and Software*, 5(1):75–104, 1995.

[5] E. de Klerk, C. Roos, T. Terlaky, "Initialization in semidefinite programming via a self–dual skew–symmetric embedding," Report 96–10, Faculty of Technical Mathematics and Computer Science, Delft University of Technology, Delft, 1996.

[6] Z.-Q. Luo, J. Sturm, S. Zhang, "Duality And Self-Duality For Conic Convex Programming," Report 9620/A, Econometric Institute, Erasmus University Rotterdam, 1996.

[7] Yu. Nesterov, "Long-step strategies in interior point potential-reduction methods," Department SES COMIN, Geneva University, 1993, to appear *Mathematical Programming.*

[8] Yu. Nesterov, "Infeasible start interior point primal-dual methods in nonlinear programming," CORE Discussion Paper, 1995.

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

[10] Yu. Nesterov and M. Todd, "Self-scaled barriers and interior-point methods for convex programming," Discussion Paper 9462, CORE, Louvain-la-Neuve, 1994.

[11] Yu. Nesterov and M. Todd, "Primal-dual interior-point methods for self-scaled cones," Discussion Paper 9544, CORE, Louvain-la-Neuve, 1995.

[12] F. Potra and R. Sheng, "Homogeneous Interior–Point Algorithms for Semidefinite Programming," Reports On Computational Mathematics, No. 82/1995, Department Of Mathematics, The University Of Iowa, 1995.

[13] J. Renegar, "Some perturbation theory for linear programming," *Mathematical Programming* 65 (1994) 73–91.

[14] J. Renegar, "Incorporating condition measures into the complexity theory of linear programming," *SIAM Journal on Optimization* 5 (1995) 506–524.

[15] J. Renegar, "Linear programming, complexity theory, and elementary functional analysis," *Mathematical Programming* 70 (1995) 259–351.

[16] K. Tanabe, "Centered Newton method for mathematical programming," in: M. Iri and K. Yajima, eds., *System Modeling and Optimization : Proceedings of the 13th IFIP Conference, Tokyo, Japan, Aug./Sept. 1987*, volume 113 of *Lecture Notes in*

Control and Information Sciences (Springer Verlag, Berlin, Germany, 1988) pp. 197–206.

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

[18] Y. Ye, M. J. Todd, and S. Mizuno, "An $O(\sqrt{n}L)$–iteration homogeneous and self–dual linear programming algorithm," *Mathematics of Operations Research* 19 (1994) 53–67.