



INTERNATIONAL ASSOCIATION FOR RESEARCH AND TEACHING
Economics, Finance, Operations Research, Econometrics and Statistics

++ research ++ teaching ++

ECORE DISCUSSION PAPER

2009/105

Local Quadratic Convergence of Polynomial-Time Interior-Point Methods for Conic Optimization Problems

Yurii NESTEROV
Levent TUNCEL

CORE DISCUSSION PAPER
2009/72

**Local quadratic convergence of polynomial-time
interior-point methods for conic optimization problems**

Yu. NESTEROV¹ and Levent TUNCEL²

November 2009

Abstract

In this paper, we establish a local quadratic convergence of polynomial-time interior-point methods for general conic optimization problems. The main structural property used in our analysis is the logarithmic homogeneity of self-concordant barrier functions. We propose new path-following predictor-corrector schemes which work only in the dual space. They are based on an easily computable gradient proximity measure, which ensures an automatic transformation of the global linear rate of convergence to the local quadratic one under some mild assumptions. Our step-size procedure for the predictor step is related to the maximum step size (the one that takes us to the boundary). It appears that in order to obtain local superlinear convergence, we need to tighten the neighborhood of the central path proportionally to the current duality gap.

Keywords: conic optimization problem, worst-case complexity analysis, self-concordant barriers, polynomial-time methods, predictor-corrector methods, local quadratic convergence.

¹ Université catholique de Louvain, CORE and INMA, B-1348 Louvain-la-Neuve, Belgium.
E-mail: yurii.nesterov@uclouvain.be

² Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Canada.
E-mail: ltuncel@math.uwaterloo.ca

The research results presented in this paper have been supported by a grant "Action de recherche concertée ARC 04/09-315" from the "Direction de la recherche scientifique – Communauté française de Belgique". The research of the second author was supported in part by Discovery Grants from NSERC.

This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

1 Introduction

Motivation. Local quadratic convergence is a natural and very desired property of many methods in Nonlinear Optimization. However, for interior-point methods the corresponding analysis is not trivial. The reason is that the barrier function is not defined in a neighborhood of the solution. Therefore, in order to study the behavior of the central path, we need to employ somehow the separable structure of the functional inequality constraints. From the very beginning [3], this analysis was based on the Implicit Function Theorem as applied to Karush-Kuhn-Tucker conditions.

This tradition explains, up to some extent, the delay in developing an appropriate framework for analyzing the local behavior of general polynomial-time interior-point methods [11]. Indeed, in the theory of self-concordant functions it is difficult to analyze the local structure of the solution since we have no access to the components of the barrier function. Moreover, in general, it is difficult to relate the self-concordant barrier with functional inequality constraints of the initial optimization problem. Therefore, up to now, the local superlinear convergence for polynomial-time path-following methods was proved only for Linear Programming [15, 8] and for Semidefinite Programming problems [6, 13, 7, 5]. In both cases, the authors use in their analysis the special boundary structure of the feasible regions and of the set of optimal solutions.

In this paper, we establish the local quadratic convergence of interior-point path-following methods by employing some geometric properties of the general conic optimization problem. The main structural property used in our analysis is the logarithmic homogeneity of self-concordant barrier functions. We propose new path-following predictor-corrector schemes which work only in the dual space. They are based on an easily computable gradient proximity measure, which ensures an automatic transformation of the global linear rate of convergence to the local quadratic rate (under a mild assumption). Our step-size procedure for the predictor step is related to the maximum step size to stay feasible. It appears that in order to attain local superlinear convergence (by an algorithm that follows the central path), we need to tighten the neighborhood of the central path proportionally to the current duality gap.

Contents. The paper is organized as follows. In Section 2 we introduce the conic primal-dual problem and define the central path. After that, we pass to a small full-dimensional dual problem and define the prediction operator. In order to achieve local quadratic convergence, we introduce two assumptions. One is on the strict dual maximum, and the second one is on the boundedness of the vector $\nabla^2 F_*(s)s_*$ along the central path. The main result of this section is Theorem 2 which demonstrates the quadratic decrease of the distance to the optimum for the prediction point, measured in an appropriately chosen fixed Euclidean norm.

In Section 3 we estimate efficiency of the predictor step measured in a local norm defined by the dual barrier function. Also, we show that the local quadratic convergence can be achieved by a feasible predictor step.

In Section 4 we prepare for the analysis of polynomial-time predictor-corrector strategies. For that, we introduce a new characteristic of self-concordant barriers, the *recession coefficient*. This coefficient bounds the growth of the Hessian of the barrier function along recession directions. We argue that in many practical situations this coefficient is a small

absolute constant. We study an important class of barriers with unit recession coefficient (we call them the *barriers with negative curvature*). This class includes at least self-scaled barriers [12] and hyperbolic barriers [4, 1, 14].

In Section 5 we establish some bounds on the growth of a variant of the gradient proximity measure. We show that we can achieve a local quadratic rate of convergence. It is important that the decrease of the parameter of the central path along the predictor direction be related to the *distance to the boundary* of the feasible solution set. We show that for local quadratic convergence the centering condition must be satisfied with increasing accuracy.

In Section 6 we show that the local quadratic convergence can be combined with the global polynomial-time complexity. We present two methods of this type. One of them uses the recession coefficient, but it has a cheap computation of the predictor step. For the second one, the recession coefficient is not needed, but the recession step is more expensive. Finally, in Section 7 we discuss the results and study two 2D-examples, which demonstrate that our assumptions are quite natural.

Notation and generalities. In what follows, we denote by \mathbb{E} a finite-dimensional linear space (other variants: \mathbb{H} , \mathbb{V}), and by \mathbb{E}^* its *dual* space, composed by linear functions on \mathbb{E} . The value of function $s \in \mathbb{E}^*$ at point $x \in \mathbb{E}$ is denoted by $\langle s, x \rangle$. This notation is the same for all spaces in use.

For an operator $A : \mathbb{E} \rightarrow \mathbb{H}^*$ we denote by A^* the corresponding *adjoint* operator:

$$\langle Ax, y \rangle = \langle A^*y, x \rangle, \quad x \in \mathbb{E}, \quad y \in \mathbb{H}.$$

Thus, $A^* : \mathbb{H} \rightarrow \mathbb{E}^*$. A self-adjoint positive-definite operator $B : \mathbb{E} \rightarrow \mathbb{E}^*$ (notation $B \succ 0$) defines the Euclidean norms for the primal and dual spaces:

$$\|x\|_B = \langle Bx, x \rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|s\|_B = \langle s, B^{-1}s \rangle^{1/2}, \quad s \in \mathbb{E}^*.$$

The sense of this notation is determined by the space of arguments. We use the following notation for ellipsoids in \mathbb{E} :

$$\mathcal{E}_B(x, r) = \{u \in \mathbb{E} : \|u - x\|_B \leq r\}.$$

If in this notation parameter r is missing, then $r = 1$.

For the future references, let us recall some facts from the theory of self-concordant functions. Most of these results can be found in Section 4 in [9]. We use the following notation for *gradient* and *Hessian* of function Φ :

$$\nabla\Phi(x) \in \mathbb{E}^*, \quad \nabla^2\Phi(x) \cdot h \in \mathbb{E}^*, \quad x, h \in \mathbb{E}.$$

Let Φ be a self-concordant function defined on the interior of a convex set $Q \subset \mathbb{E}$:

$$\nabla^3\Phi(x)[h, h, h] \leq 2\langle \nabla^2\Phi(x)h, h \rangle^{3/2}, \quad x \in \text{int } Q, \quad h \in \mathbb{E}, \quad (1.1)$$

where $\nabla^3\Phi(x)[h_1, h_2, h_3]$ is the third differential of function Φ at point x along the corresponding directions h_1, h_2, h_3 . Note that $\nabla^3\Phi(x)[h_1, h_2, h_3]$ is a trilinear symmetric form. Thus,

$$\nabla^3\Phi(x)[h_1, h_2] = \nabla^3\Phi(x)[h_2, h_1] \in \mathbb{E}^*,$$

and $\nabla^3\Phi(x)[h]$ is a self-adjoint linear operator from \mathbb{E} to \mathbb{E}^* .

Assume that Q contains no straight line. Then $\nabla^2\Phi(u)$ is nondegenerate for any $u \in \text{int } Q$. Self-concordant function Φ is called ν -self-concordant barrier if

$$\langle \nabla\Phi(u), [\nabla^2\Phi(u)]^{-1}\nabla\Phi(u) \rangle \leq \nu. \quad (1.2)$$

For local norms related to self-concordant functions we use the following concise notation:

$$\|h\|_u = \langle \nabla^2\Phi(u)h, h \rangle^{1/2}, \quad h \in \mathbb{E},$$

$$\|s\|_u = \langle s, [\nabla^2\Phi(u)]^{-1}s \rangle^{1/2}, \quad s \in \mathbb{E}^*.$$

Thus, inequality (1.2) can be written as $\|\nabla\Phi(u)\|_u^2 \leq \nu$.

For $u \in \text{int } Q$ define the Dikin ellipsoid $W_r(u) \stackrel{\text{def}}{=} \mathcal{E}_{\nabla^2\Phi(u)}(u, r)$. Then $W_r(u) \subset Q$ for all $r \in [0, 1)$.

Theorem 1 (*Theorem on Recession Direction; see Section 4 in [9] for the proof.*) *If h is a recession direction of the set Q , then*

$$\|h\|_u \leq \langle -\nabla\Phi(u), h \rangle. \quad (1.3)$$

If $v \in W_r(u)$, then

$$\langle \nabla\Phi(v) - \nabla\Phi(u), v - u \rangle \geq \frac{r^2}{1+r}, \quad r \geq 0. \quad (1.4)$$

For $r \in [0, 1)$ we have

$$(1-r)^2\nabla^2\Phi(u) \preceq \nabla^2\Phi(v) \preceq \frac{1}{(1-r)^2}\nabla^2\Phi(u), \quad (1.5)$$

$$\|\nabla\Phi(v) - \nabla\Phi(u)\|_u \leq \frac{r}{1-r}, \quad (1.6)$$

$$\|\nabla\Phi(v) - \nabla\Phi(u) - \nabla^2\Phi(u)(v-u)\|_u \leq \frac{r^2}{1-r}. \quad (1.7)$$

Finally, we need several statements on barriers for *convex cones*. We call cone $K \subset \mathbb{E}$ *regular* if it is a closed, convex, and pointed cone with nonempty interior. Sometimes it is convenient to write inclusion $x \in K$ in the form $x \succeq_K 0$.

If K is regular, then the *dual cone*

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

is also regular. For cone K , we assume available a ν -normal barrier $F(x)$. This means that F is self-concordant and ν -logarithmically homogeneous:

$$F(\tau x) = F(x) - \nu \ln \tau, \quad x \in \text{int } K, \tau > 0. \quad (1.8)$$

Note that $-\nabla F(x) \in \text{int } K^*$ for every $x \in \text{int } K$. Equality (1.8) leads to many interesting identities:

$$\nabla F(\tau x) = \tau^{-1} \cdot \nabla F(x), \quad (1.9)$$

$$\nabla^2 F(\tau x) = \tau^{-2} \cdot \nabla^2 F(x), \quad (1.10)$$

$$\langle \nabla F(x), x \rangle = -\nu, \quad (1.11)$$

$$\nabla^2 F(x) \cdot x = -\nabla F(x), \quad (1.12)$$

$$\nabla^3 F(x)[x] = -2\nabla^2 F(x), \quad (1.13)$$

$$\|\nabla F(x)\|_x^2 = \nu, \quad (1.14)$$

where $x \in \text{int } K$ and $\tau > 0$. Note that the *dual barrier*

$$F_*(s) = \max_{x \in \text{int } K} \{ -\langle s, x \rangle - F(x) \}$$

is a ν -normal barrier for cone K^* . The differential characteristics of the primal and dual barriers are related as follows:

$$\begin{aligned} \nabla F(-\nabla F_*(s)) &= -s, & \nabla^2 F(-\nabla F_*(s)) &= [\nabla^2 F_*(s)]^{-1}, \\ \nabla F_*(-\nabla F(x)) &= -x, & \nabla^2 F_*(-\nabla F(x)) &= [\nabla^2 F(x)]^{-1}, \end{aligned} \quad (1.15)$$

where $x \in \text{int } K$ and $s \in \text{int } K^*$.

For normal barriers, the Theorem on Recession Direction (1.3) can be written both in primal and dual forms:

$$\|u\|_x \leq \langle -\nabla F(x), u \rangle, \quad x \in \text{int } K, u \in K, \quad (1.16)$$

$$\|s\|_x \leq \langle s, x \rangle, \quad x \in \text{int } K, s \in K^*. \quad (1.17)$$

The following statement is very useful.

Lemma 1 *Let F be a ν -normal barrier for K and $B \succ 0$. Assume that $\mathcal{E}_B(u) \subset K$ and for $x \in \text{int } K$ we have*

$$\langle \nabla F(x), u - x \rangle \geq 0.$$

Then $B \succeq \frac{1}{4\nu^2} \nabla^2 F(x)$.

Proof:

Let us fix an arbitrary direction $h \in \mathbb{E}^*$. We can assume that

$$\langle \nabla F(x), B^{-1}h \rangle \geq 0, \quad (1.18)$$

(otherwise, multiply h by -1). Denote $y = u + \frac{B^{-1}h}{\|h\|_B}$. Then $y \in K$. Therefore,

$$\begin{aligned} \frac{\|B^{-1}h\|_x}{\|h\|_B} &\leq \|u\|_x + \|y\|_x \stackrel{(1.16)}{\leq} \langle -\nabla F(x), u \rangle + \langle -\nabla F(x), y \rangle \\ &\stackrel{(1.18)}{\leq} 2\langle -\nabla F(x), u \rangle \stackrel{(1.11)}{\leq} 2\nu. \end{aligned}$$

Thus, $B^{-1}\nabla^2 F(x)B^{-1} \preceq 4\nu^2 B^{-1}$. \square

Corollary 1 *Let $x, u \in \text{int } K$ and $\langle \nabla F(x), u - x \rangle \geq 0$. Then $\nabla^2 F(u) \succeq \frac{1}{4\nu^2} \nabla^2 F(x)$.*

Corollary 2 *Let $x \in \text{int } K$ and $u \in K$. Then $\nabla^2 F(x + u) \preceq 4\nu^2 \nabla^2 F(x)$.*

Proof:

Denote $y = x + u \in \text{int } K$. Then $\langle \nabla F(y), x - y \rangle = \langle -\nabla F(y), u \rangle \geq 0$. Hence, we can apply Corollary 1. \square

To conclude with notation, let us introduce the following relative measure for directions in \mathbb{E} :

$$\sigma_x(h) = \min_{\rho \geq 0} \{ \rho : \rho \cdot x - h \in K \} \leq \|h\|_x, \quad x \in \text{int } K, h \in \mathbb{E}. \quad (1.19)$$

2 Prediction from neighborhood of central path

Consider the standard conic optimization problem:

$$\min_{x \in K} \{ \langle c, x \rangle : Ax = b \}, \quad (2.1)$$

where $c \in \mathbb{E}^*$, $b \in \mathbb{H}^*$, A is a linear transformation from \mathbb{E} to \mathbb{H}^* , and $K \subset \mathbb{E}$ is a regular cone. The problem dual to (2.1) is then

$$\max_{s \in K^*, y \in \mathbb{H}} \{ \langle b, y \rangle : s + A^*y = c \}. \quad (2.2)$$

Note that the feasible points of the primal and dual problems move in the orthogonal subspaces:

$$\langle s_1 - s_2, x_1 - x_2 \rangle = 0 \quad (2.3)$$

for all $x_1, x_2 \in \mathcal{F}_p \stackrel{\text{def}}{=} \{x \in K : Ax = b\}$, and $s_1, s_2 \in \mathcal{F}_d \stackrel{\text{def}}{=} \{s \in K^* : s + A^*y = c\}$.

Under the *strict feasibility* assumption,

$$\exists x_0 \in \text{int } K, s_0 \in \text{int } K^*, y_0 \in \mathbb{H} : Ax_0 = b, s_0 + A^*y_0 = c, \quad (2.4)$$

the optimal sets of the primal and dual problems are nonempty and bounded, and there is no duality gap. Moreover, a *primal-dual central path* $z_\mu \stackrel{\text{def}}{=} (x_\mu, s_\mu, y_\mu)$:

$$\left. \begin{aligned} Ax_\mu &= b, \\ c + \mu \nabla F(x_\mu) &= A^*y_\mu, \\ s_\mu &= -\mu \nabla F(x_\mu) \end{aligned} \right\}, \quad \mu > 0, \quad (2.5)$$

is well defined. Note that

$$\langle c, x_\mu \rangle - \langle b, y_\mu \rangle = \langle s_\mu, x_\mu \rangle \stackrel{(2.5), (1.11)}{=} \nu \cdot \mu. \quad (2.6)$$

The majority of modern strategies for solving the primal-dual problem pair (2.1), (2.2) suggest to follow this trajectory as $\mu \rightarrow 0$. On the one hand, it is important that μ be decreased at a linear rate to attain a polynomial-time complexity. However, in a small neighborhood of the solution, it is highly desirable to switch on a super-linear rate. Such a possibility was already discovered for Linear Programming problems [15, 8]. There has also been significant progress in the case of Semidefinite Programming [6, 13, 7, 5]. In this paper, we study more general conic problems.

For a fast local convergence of a path-following scheme, we need to show that the predicted point

$$\hat{z}_\mu = z_\mu - z'_\mu \cdot \mu$$

enters a small neighborhood of the solution point

$$z_* = \lim_{\mu \rightarrow 0} z_\mu = (x_*, s_*, y_*).$$

It is more convenient to analyze this situation by looking at y -component of the central path.

Note that s -component of the dual problem (2.2) can be easily eliminated:

$$s = s(y) \stackrel{\text{def}}{=} c - A^*y.$$

Then, the remaining part of the dual problem can be written in a more concise full-dimensional form:

$$f^* \stackrel{\text{def}}{=} \max_{y \in \mathbb{H}} \{ \langle b, y \rangle : y \in Q \}, \quad (2.7)$$

$$Q \stackrel{\text{def}}{=} \{ y \in \mathbb{H} : c - A^*y \in K^* \}.$$

In view of Assumption (2.4), the interior of set Q is nonempty. Moreover, for this set we have a ν -self-concordant barrier

$$f(y) = F_*(c - A^*y), \quad y \in \text{int } Q.$$

Since the optimal set of problem (2.7) is bounded, Q contains no straight line. Thus, this barrier has nondegenerate Hessian at any strictly feasible point. Note that Assumption (2.4), also implies that the linear transformation A is surjective.

It is clear that y -component of the primal-dual central path z_μ coincides with the central path of the problem (2.7):

$$\begin{aligned} b &= \mu \nabla f(y_\mu) = -\mu A \nabla F_*(c - A^*y_\mu) \\ &= -\mu A \nabla F_*(s_\mu) \stackrel{(1.9),(1.15)}{=} Ax_\mu, \quad \mu > 0. \end{aligned} \quad (2.8)$$

Let us estimate the quality of the following prediction point:

$$\begin{aligned} p(y) &\stackrel{\text{def}}{=} y + v(y), \quad y \in \text{int } Q, \\ v(y) &\stackrel{\text{def}}{=} [\nabla^2 f(y)]^{-1} \nabla f(y), \quad s_p(y) \stackrel{\text{def}}{=} s(y) - A^*v(y). \end{aligned}$$

Indeed, in a neighborhood of a non-degenerate solution, the barrier function should be close to the barrier of a tangent cone centered at the solution. Hence, the relation (1.12) should be satisfied with a reasonably high accuracy.

For every $y \in \text{int } Q$, we have

$$\begin{aligned} p(y) &= [\nabla^2 f(y)]^{-1} \cdot [\nabla^2 f(y)y + \nabla f(y)] \\ &= [\nabla^2 f(y)]^{-1} \cdot [A \nabla^2 F_*(c - A^*y)A^*y - A \nabla F_*(c - A^*y)] \\ &\stackrel{(1.12)}{=} [\nabla^2 f(y)]^{-1} \cdot [A \nabla^2 F_*(c - A^*y)A^*y + A \nabla^2 F_*(c - A^*y)(c - A^*y)] \\ &= [\nabla^2 f(y)]^{-1} A \nabla^2 F_*(c - A^*y) \cdot c. \end{aligned}$$

Let us choose an arbitrary pair (s_*, y_*) from the optimal solution set of the problem (2.2). Then,

$$c = A^*y_* + s_*.$$

Thus, we have proved the following representation.

Lemma 2 For every $y \in \text{int } Q$ and every optimal pair (s_*, y_*) of dual problem (2.2) we have

$$p(y) = y_* + [\nabla^2 f(y)]^{-1} A \nabla^2 F_*(s(y)) \cdot s_*. \quad (2.9)$$

Remark 1 Note that the right-hand side of equation (2.9) has a gradient interpretation. Indeed, let us fix some $s \in K^*$ and define the function

$$\phi_s(y) = -\langle s, \nabla F_*(c - A^*y) \rangle, \quad y \in Q.$$

Then $\nabla \phi_s(y) = A \nabla^2 F(c - A^*y) \cdot s$, and, for self-scaled barriers $\phi_s(\cdot)$ is convex (as well as for the barriers with negative curvature, see Section 4). Thus, the representation (2.9) can be rewritten as follows:

$$p(y) = y_* + [\nabla^2 f(y)]^{-1} \nabla \phi_{s_*}(y). \quad (2.10)$$

Note that $[\nabla^2 f(y)]^{-1}$ in the limit acts as a projector onto the tangent subspace to the feasible set at the solution.

For some problems with simple structure, we can guarantee that the product of matrix $[\nabla^2 f(y)]^{-1}$ by the vector $A \nabla^2 F_*(c - A^*y) s_*$ is small in norm. However, in more general situations, we need to apply stronger assumptions. Namely, we are going to show that, under certain conditions, vector $\nabla^2 F_*(c - A^*y) s_*$ is bounded and matrix $[\nabla^2 f(y)]^{-1}$ becomes small in norm as y approaches y_* .

The global complexity analysis of interior-point methods is done in an affine-invariant framework. However, for analyzing the local convergence of these schemes, we need to fix some norms in the primal and dual spaces. For simplicity, let us choose them Euclidean. We recall the definitions of Euclidean norms based on B and using B , we define G below:

$$\|x\|_B = \langle Bx, x \rangle^{1/2}, \quad x \in \mathbb{E}, \quad \|s\|_B = \langle s, B^{-1}s \rangle^{1/2}, \quad s \in \mathbb{E}^*, \quad (2.11)$$

$$G \stackrel{\text{def}}{=} AB^{-1}A^*,$$

where the operator $B : \mathbb{E} \rightarrow \mathbb{E}^*$ is self-adjoint and positive definite. Thus, using a Schur complement argument and the fact that A is surjective, we have

$$A^*G^{-1}A \preceq B. \quad (2.12)$$

It is convenient to choose B related in a certain way to the primal central path. Let us define

$$B = \nabla^2 F(x_1) \quad (2.13)$$

and establish some natural bounds related to the points of the primal central path.

Lemma 3 If $\mu_1 \leq \mu_0$, then

$$\|x_{\mu_1}\|_{x_{\mu_0}} \leq \nu. \quad (2.14)$$

In particular, for any $\mu \leq 1$ we have:

$$\|x_\mu\|_B^2 \leq \nu^2. \quad (2.15)$$

Moreover,

$$\nabla^2 F(x_\mu) \succeq \frac{1}{4\nu^2} \nabla^2 F(x_1) \stackrel{(2.13)}{=} \frac{1}{4\nu^2} B. \quad (2.16)$$

Proof:

Indeed,

$$\begin{aligned} \|x_{\mu_1}\|_{x_{\mu_0}} &\stackrel{(1.16)}{\leq} \langle -\nabla F(x_{\mu_0}), x_{\mu_1} \rangle^2 = \frac{1}{\mu_0^2} \langle s_{\mu_0}, x_{\mu_1} \rangle^2 \\ &\stackrel{(2.5)}{=} \frac{1}{\mu_0^2} [\langle c, x_{\mu_1} \rangle - \langle b, y_{\mu_0} \rangle]^2 \stackrel{(2.6)}{\leq} \nu^2. \end{aligned}$$

The last inequality also uses the fact that $\langle c, x_{\mu_1} \rangle \leq \langle c, x_{\mu_0} \rangle$. Therefore, by (2.13) we obtain (2.15). Finally,

$$\langle \nabla F(x_1), x_{\mu} - x_1 \rangle \stackrel{(2.5)}{=} \langle c, x_1 - x_{\mu} \rangle \geq 0.$$

Therefore, applying Corollary 1, we get (2.16). \square

Now we can introduce our main assumptions.

Assumption 1 *There exists a constant $\gamma_d > 0$ such that*

$$f^* - \langle b, y \rangle = \langle s, x_* \rangle \geq \gamma_d \|s - s_*\|_B \equiv \gamma_d \|y - y_*\|_G, \quad (2.17)$$

for every $y \in Q$ (that is $s = s(y) \in \mathcal{F}_d$).

Thus, we assume that the dual problem (2.2) admits a *sharp* optimal solution. We need one more assumption.

Assumption 2 *There exists a constant σ_d such that for any $\mu \leq 1$ we have*

$$\|\nabla^2 F_*(s_{\mu})s_*\|_B \leq \sigma_d. \quad (2.18)$$

In what follows, we always suppose that these assumptions are valid. Let us look how they work in some important special cases.

Example 1 *Consider the nonnegative orthant $K = K^* = \mathbb{R}_+^n$ with barriers*

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

Denote by I_* the set of positive components of the optimal dual solution s_* . Then denoting by e the vector of all ones, we have

$$\langle e, \nabla^2 F_*(s_{\mu})s_* \rangle = \sum_{i \in I_*} \frac{s_*^{(i)}}{(s_{\mu}^{(i)})^2} = \sum_{i \in I_*} \frac{1}{s_*^{(i)}} \left(\frac{s_*^{(i)}}{s_{\mu}^{(i)}} \right)^2 \leq \|s_*\|_{s_{\mu}} \max_{i \in I_*} \frac{1}{s_*^{(i)}} \stackrel{(2.14)}{\leq} \max_{i \in I_*} \frac{\nu}{s_*^{(i)}}.$$

Since vector $\nabla^2 F_*(s_{\mu})s_*$ is nonnegative, we obtain for it an upper bound in terms of $\max_{i \in I_*} \frac{\nu}{s_*^{(i)}}$. Note that this bound is finite even for degenerate dual solution.

It is interesting, that we can find a bound for vector $F_*(s_{\mu})s_*$ based on the properties of the primal central path. Indeed, for all $i \in I_*$, we have

$$\left(\nabla^2 F_*(s_{\mu})s_* \right)^{(i)} = \frac{s_*^{(i)}}{(s_{\mu}^{(i)})^2} = s_*^{(i)} \cdot \frac{(x_{\mu}^{(i)})^2}{\mu^2} = s_*^{(i)} \cdot \left(\frac{x_{\mu}^{(i)} - x_*^{(i)}}{\mu} \right)^2.$$

Thus, assuming $\|x_\mu - x_*\| \leq O(\mu)$, we get a bound for $F_*(s_\mu)s_*$. Note that the latter assumption is weaker than assuming that the primal problem admits a sharp solution. It is even weaker than assuming differentiability of the primal trajectory x_μ at $\mu = 0$.

For the cone of positive-semidefinite matrices $K = K^* = \mathbb{S}_+^n$, we choose

$$F(X) = -\ln \det X, \quad F_*(S) = n - \ln \det S.$$

Then,

$$\langle I, \nabla^2 F_*(S_\mu)S_* \rangle = \langle I, S_\mu^{-1}S_*S_\mu^{-1} \rangle.$$

It seems difficult to get an upper bound for this value in terms of $\|S_*\|_{S_\mu} = \langle S_\mu^{-1}S_*S_\mu^{-1}, S_* \rangle$. However, the second approach also works here:

$$\langle I, S_\mu^{-1}S_*S_\mu^{-1} \rangle = \mu^{-2} \langle X_\mu^2, S_* \rangle = \mu^{-2} \langle (X_\mu - X_*)^2, S_* \rangle.$$

Thus, we get an upper bound for $\|\nabla^2 F_*(S_\mu)S_*\|$ assuming $\|X_\mu - X_*\| \leq O(\mu)$. Again, this condition is weaker than assuming that the primal problem admits a sharp solution. It is also weaker than assuming differentiability of the primal central path at $\mu = 0$. \square

Let us derive from Assumption 1 that $[\nabla^2 f(y)]^{-1}$ becomes small in norm as y approaches y_* .

Lemma 4 For every $y \in \text{int } Q$, we have

$$[\nabla^2 f(y)]^{-1} \preceq \frac{4}{\gamma_d} [f^* - \langle b, y \rangle]^2 \cdot G^{-1}. \quad (2.19)$$

Proof:

Let us fix some $y \in \text{int } Q$. Consider an arbitrary direction $h \in \mathbb{H}^*$. Without loss of generality, we may assume that $\langle b, [\nabla^2 f(y)]^{-1}h \rangle \geq 0$ (otherwise, we can consider direction $-h$). Since f is a self-concordant barrier, the point

$$y_h \stackrel{\text{def}}{=} y + \frac{[\nabla^2 f(y)]^{-1}h}{\langle h, [\nabla^2 f(y)]^{-1}h \rangle^{1/2}}$$

belongs to the set Q . Therefore, in view of inequality (2.17), we have

$$\gamma_d \|y_h - y_*\|_G \leq f^* - \langle b, y_h \rangle \leq f^* - \langle b, y \rangle.$$

Hence,

$$\begin{aligned} \frac{1}{\gamma_d} [f^* - \langle b, y \rangle] &\geq \frac{\|[\nabla^2 f(y)]^{-1}h\|_G}{\langle h, [\nabla^2 f(y)]^{-1}h \rangle^{1/2}} - \|y - y_*\|_G \\ &\stackrel{(2.17)}{\geq} \frac{\|[\nabla^2 f(y)]^{-1}h\|_G}{\langle h, [\nabla^2 f(y)]^{-1}h \rangle^{1/2}} - \frac{1}{\gamma_d} [f^* - \langle b, y \rangle]. \end{aligned}$$

Thus, for any $h \in \mathbb{H}^*$ we have

$$\|[\nabla^2 f(y)]^{-1}h\|_G^2 \leq \frac{4}{\gamma_d} [f^* - \langle b, y \rangle]^2 \cdot \langle h, [\nabla^2 f(y)]^{-1}h \rangle.$$

This means that

$$[\nabla^2 f(y)]^{-1}G[\nabla^2 f(y)]^{-1} \preceq \frac{4}{\gamma_d} [f^* - \langle b, y \rangle]^2 [\nabla^2 f(y)]^{-1},$$

and (2.19) follows. \square

Now, we can also estimate the size of the Hessian with respect to the norm induced by G . We define

$$\|[\nabla^2 f(y)]^{-1}\|_G \stackrel{\text{def}}{=} \max_{h \in \mathbb{H}^*} \{ \|[\nabla^2 f(y)]^{-1}h\|_G : \|h\|_G = 1 \}.$$

Corollary 3 *For every $y \in \text{int } Q$, we have*

$$\|[\nabla^2 f(y)]^{-1}\|_G \leq \frac{4}{\gamma_d^2} [f^* - \langle b, y \rangle]^2. \quad (2.20)$$

Therefore, $\|v(y)\|_G \leq \frac{2\nu^{1/2}}{\gamma_d} [f^* - \langle b, y \rangle]$.

Proof:

Note that

$$\|[\nabla^2 f(y)]^{-1}h\|_G^2 = \langle h, [\nabla^2 f(y)]^{-1}G[\nabla^2 f(y)]^{-1}h \rangle, \quad h \in \mathbb{H}^*.$$

Hence, (2.20) follows directly from (2.19). Further,

$$\begin{aligned} \|v(y)\|_G^2 &= \langle G[\nabla^2 f(y)]^{-1}\nabla f(y), [\nabla^2 f(y)]^{-1}\nabla f(y) \rangle \\ &\stackrel{(2.19)}{\leq} \frac{4}{\gamma_d^2} [f^* - \langle b, y \rangle]^2 \langle \nabla f(y), [\nabla^2 f(y)]^{-1}\nabla f(y) \rangle. \end{aligned}$$

It remains to use inequality (1.2). \square

Before proving the main result of this section, we need to estimate the norm of the initial data.

Lemma 5 *We have*

$$\begin{aligned} \|A\|_G &\stackrel{\text{def}}{=} \max_{h \in \mathbb{E}} \{ \|Ah\|_G : \|h\|_B = 1 \} \leq 1, \\ \|b\|_G^2 &\leq \nu. \end{aligned} \quad (2.21)$$

Proof:

Indeed, for any $h \in \mathbb{E}$, we have

$$\begin{aligned} \|Ah\|_G^2 &= \langle Ah, G^{-1}Ah \rangle = \max_{y \in \mathbb{H}} [2\langle Ah, y \rangle - \langle Gy, y \rangle] \\ &= \max_{y \in \mathbb{H}} [2\langle A^*y, h \rangle - \langle A^*y, B^{-1}A^*y \rangle] \\ &\leq \max_{s \in \mathbb{E}^*} [2\langle s, h \rangle - \langle s, B^{-1}s \rangle] = \|h\|_B^2. \end{aligned}$$

For justifying the second inequality, note that

$$\begin{aligned}
\|b\|_G^2 &= \langle b, G^{-1}b \rangle = \max_{y \in \mathbb{H}} [2\langle b, y \rangle - \langle A^*y, B^{-1}A^*y \rangle] \\
&= \max_{y \in \mathbb{H}} [2\langle A^*y, x_1 \rangle - \langle A^*y, B^{-1}A^*y \rangle] \\
&\leq \max_{s \in \mathbb{E}^*} [2\langle s, x_1 \rangle - \langle s, B^{-1}s \rangle] = \langle Bx_1, x_1 \rangle \\
&\stackrel{(2.13)}{=} \langle \nabla^2 F(x_1)x_1, x_1 \rangle \stackrel{(1.11), (1.12)}{=} \nu.
\end{aligned}$$

□

We will work with points in a small neighborhood of the central path defined by the *local gradient proximity measure*. Denote

$$\mathcal{N}(\mu, \beta) = \left\{ y \in \mathbb{H} : \gamma(y, \mu) \stackrel{\text{def}}{=} \|\nabla f(y) - \frac{1}{\mu}b\|_y \leq \beta \right\}, \quad \mu \in (0, 1], \beta \in (0, \frac{1}{2}). \quad (2.22)$$

This proximity measure has a very familiar interpretation in the special case of Linear Programming. Denoting by S the diagonal matrix made up from the slack variable $s = c - A^T y$, notice that Dikin's affine scaling direction in this case is given by $[AS^{-2}A^T]^{-1}b$. Our predictor step corresponds to the search direction $[AS^{-2}A^T]^{-1}AS^{-1}e$. Our proximity measure becomes

$$\left\| AS^{-1}e - \frac{1}{\mu}b \right\|_{AS^{-2}A^T}.$$

Lemma 6 *Let $y \in \mathcal{N}(\mu, \beta)$ with $\mu \in (0, 1]$ and $\beta \in (0, \frac{1}{2})$. Then*

$$\|\nabla^2 F_*(s(y))s_*\|_B \leq \frac{\sigma_d(1-\beta)^2}{(1-2\beta)^2}, \quad (2.23)$$

$$f^* - \langle b, y \rangle \leq \kappa_1 \cdot \mu, \quad (2.24)$$

where $\kappa_1 = \nu + \frac{\beta(\beta+\sqrt{\nu})}{1-\beta}$.

Proof:

Indeed,

$$\begin{aligned}
\|s(y) - s_\mu\|_{s(y)} &= \langle \nabla^2 F_*(s(y))A^*(y - y_\mu), A^*(y - y_\mu) \rangle^{1/2} \\
&= \|y - y_\mu\|_y \stackrel{\text{def}}{=} r \leq \frac{\beta}{1-\beta}.
\end{aligned}$$

(for the last inequality we used (1.4)). Therefore, by (1.5) we have

$$\nabla^2 F(s(y)) \preceq \frac{1}{(1-r)^2} \nabla^2 F(s_\mu) \preceq \frac{(1-\beta)^2}{(1-2\beta)^2} \nabla^2 F(s_\mu).$$

Thus,

$$\begin{aligned}
\|\nabla^2 F_*(s(y))s_*\|_B^2 &= \langle (B^{1/2}\nabla^2 F_*(s(y))B^{1/2})^2 B^{-1/2}s_*, B^{-1/2}s_* \rangle \\
&\leq \frac{(1-\beta)^4}{(1-2\beta)^4} \|\nabla^2 F_*(s_\mu)s_*\|_B^2.
\end{aligned}$$

Now, (2.23) follows from Assumption 2.

To establish (2.24), note that

$$\begin{aligned}
\frac{1}{\mu} [f^* - \langle b, y \rangle] &= \frac{1}{\mu} [\langle b, y^* - y_\mu \rangle - \langle b, y_\mu - y \rangle] \\
&\stackrel{(2.6)}{\leq} \nu + \langle \frac{b}{\mu}, y_\mu - y \rangle \\
&= \nu + \langle -\nabla f(y) + \frac{b}{\mu}, y_\mu - y \rangle + \langle \nabla f(y), y_\mu - y \rangle \\
&\leq \nu + \|\nabla f(y) - \frac{b}{\mu}\|_y \cdot \|y - y_\mu\|_y + \|\nabla f(y)\|_y \cdot \|y - y_\mu\|_y \\
&\leq \nu + \beta \frac{\beta}{1-\beta} + \sqrt{\nu} \frac{\beta}{1-\beta},
\end{aligned}$$

where the last inequality follows from the assumptions of the lemma and (1.2). \square

Now, we can put all our observations together.

Theorem 2 *Let dual problem (2.2) satisfy Assumptions 1 and 2. If for some $\mu \in (0, 1]$ and $\beta \in (0, \frac{1}{2})$ we have $y \in \mathcal{N}(\mu, \beta)$, then*

$$\|p(y) - y_*\|_G \leq \frac{4\sigma_d(1-\beta)^2}{\gamma_d^2(1-2\beta)^2} \langle b, y - y_* \rangle^2 \leq \frac{4\nu(1-\beta)^2\sigma_d}{\gamma_d^2(1-2\beta)^2} \cdot \|y - y_*\|_G^2. \quad (2.25)$$

Proof:

Indeed, in view of representation (2.9), we have

$$\|p(y) - y_*\|_G \leq \|[\nabla^2 f(y)]^{-1}\|_G \cdot \|A\|_G \cdot \|\nabla^2 F_*(s(y))s_*\|_B.$$

Now, we can use inequalities (2.20), (2.21), and (2.23). For justifying the second inequality, we apply the second bound in (2.21). \square

3 Efficiency of predictor step

Let us estimate now the efficiency of the predictor step with respect to the local norm.

Lemma 7 *If $y \in \mathcal{N}(\mu, \beta)$, then*

$$\|p(y) - y_*\|_y \leq \kappa_2 \cdot [f^* - \langle b, y \rangle] \leq \mu \cdot \kappa, \quad (3.1)$$

where $\kappa_2 = \frac{2\sigma_d(1-\beta)^2}{\gamma_d(1-2\beta)^2}$, and $\kappa = \kappa_1 \cdot \kappa_2$.

Proof:

Indeed,

$$\begin{aligned}
\|p(y) - y_*\|_y^2 &\stackrel{(2.9)}{=} \langle A\nabla^2 F_*(s(y))s_*, [\nabla^2 f(y)]^{-1} A\nabla^2 F_*(s(y))s_* \rangle \\
&\stackrel{(2.19)}{\leq} \frac{4}{\gamma_d^2} [f^* - \langle b, y \rangle]^2 \langle A\nabla^2 F_*(s(y))s_*, G^{-1} A\nabla^2 F_*(s(y))s_* \rangle \\
&\stackrel{(2.12)}{\leq} \frac{4}{\gamma_d^2} [f^* - \langle b, y \rangle]^2 \langle B\nabla^2 F_*(s(y))s_*, \nabla^2 F_*(s(y))s_* \rangle.
\end{aligned}$$

It remains to use the bounds (2.23) and (2.24). \square

Since $\|y_* - y\|_y \geq 1$, inequality (3.1) demonstrates a significant drop in the distance to the optimal point after a full predictor step. The following fact is also useful.

Lemma 8 *For every $y \in Q$ we have $A \cdot \nabla^2 F_*(s(y)) \cdot s_p(y) = 0$.*

Proof:

Indeed,

$$\begin{aligned}
A\nabla^2 F_*(s(y))s_p(y) &= A\nabla^2 F_*(s(y))(s(y) - A^*v(y)) \\
&\stackrel{(1.12)}{=} -A\nabla F(s(y)) - \nabla f(y) = 0.
\end{aligned}$$

\square

Corollary 4 *If \mathcal{F}_p is bounded, then the point $\nabla^2 F_*(s(y)) \cdot s_p(y) \notin K$ (therefore, it is infeasible for the primal problem (2.1)).*

We can show now that a large predictor step can still keep dual feasibility. Denote

$$y(\alpha) = y + \alpha v(y), \quad \alpha \in [0, 1].$$

Theorem 3 *Let $y \in \mathcal{N}(\mu, \beta)$ with $\mu \in (0, 1]$ and $\beta \in (0, \frac{1}{2})$. Then, for every $r \in (0, 1)$, the point $y(\hat{\alpha})$ with*

$$\hat{\alpha} \stackrel{\text{def}}{=} \frac{r}{r + \kappa_2 [f^* - \langle b, y \rangle]} \tag{3.2}$$

belongs to Q . Moreover,

$$f^* - \langle b, y(\hat{\alpha}) \rangle \leq \kappa_3 \cdot [f^* - \langle b, y \rangle]^2, \tag{3.3}$$

where $\kappa_3 = \kappa_2 \cdot \left(\frac{1}{r} + \frac{2\sqrt{\nu}}{\gamma_d}\right)$.

Proof:

Consider the Dikin ellipsoid $W_r(y) = \{u \in \mathbb{H} : \|u - y\|_y \leq r\}$. Since $W_r(y) \subset Q$, its convex combination with point y_* , defined as

$$Q(y) = \{u \in \mathbb{H} : \|u - (1-t)y - ty_*\|_y \leq r(1-t), t \in [0, 1]\},$$

is contained in Q . Note that

$$\begin{aligned} \|y(\hat{\alpha}) - (1 - \hat{\alpha})y - \hat{\alpha}y_*\|_y &= \hat{\alpha}\|p(y) - y_*\|_y \\ &\stackrel{(3.1)}{\leq} \kappa_2 \hat{\alpha}[f^* - \langle b, y \rangle] \stackrel{(3.2)}{=} r(1 - \hat{\alpha}). \end{aligned}$$

Hence, $y(\hat{\alpha}) \in Q$. Further,

$$\begin{aligned} f^* - \langle b, y(\hat{\alpha}) \rangle &= (1 - \hat{\alpha})[f^* - \langle b, y \rangle] + \hat{\alpha}\langle b, y_* - p(y) \rangle \\ &\leq \frac{\kappa_2}{r}[f^* - \langle b, y \rangle]^2 + \|b\|_G \cdot \|p(y) - y_*\|_G. \end{aligned}$$

Since $\|b\|_G \stackrel{(2.21)}{\leq} \sqrt{\nu}$ and

$$\|p(y) - y_*\|_G \stackrel{(2.25)}{\leq} \frac{2\kappa_2}{\gamma_d}[f^* - \langle b, y \rangle]^2,$$

we obtain the desired inequality (3.3). \square

Denote by $\bar{\alpha}(y)$, the maximal feasible step along direction $v(y)$:

$$\bar{\alpha}(y) = \max_{\alpha \geq 0} \{\alpha : y + \alpha v(y) \in Q\}.$$

Let us show that $\bar{\alpha} = \bar{\alpha}(y)$ is big enough. In general,

$$\bar{\alpha}(y) \geq \frac{1}{\|v(y)\|_y} \stackrel{(1.2)}{\geq} \frac{1}{\nu^{1/2}}. \quad (3.4)$$

However, in a small neighborhood of the solution, we can establish a better bound.

Corollary 5 *Let $y \in \mathcal{N}(\mu, \beta)$ with $\mu \in (0, 1]$ and $\beta \in (0, \frac{1}{2})$. Then*

$$1 - \bar{\alpha}(y) \leq \frac{\kappa\mu}{1 + \kappa\mu}, \quad (3.5)$$

$$\|y(\bar{\alpha}) - y^*\|_y \leq (1 + \sqrt{\nu})\kappa\mu. \quad (3.6)$$

Proof:

Since for any $r \in (0, 1)$

$$\bar{\alpha} \stackrel{(3.2)}{\geq} \hat{\alpha} = \frac{r}{r + \kappa_2[f^* - \langle b, y \rangle]},$$

we have $1 - \bar{\alpha} \leq \frac{\kappa_2[f^* - \langle b, y \rangle]}{1 + \kappa_2[f^* - \langle b, y \rangle]} \stackrel{(2.24)}{\leq} \frac{\kappa\mu}{1 + \kappa\mu}$. Further,

$$\begin{aligned} \|y(\bar{\alpha}) - y^*\|_y &\leq \|y(\bar{\alpha}) - p(y)\|_y + \|p(y) - y^*\|_y \stackrel{(3.1)}{\leq} (1 - \bar{\alpha})\|v(y)\|_y + \kappa \cdot \mu \\ &\stackrel{(1.2)}{\leq} (1 - \bar{\alpha})\sqrt{\nu} + \kappa \cdot \mu. \end{aligned}$$

It remains to apply the inequality (3.5). \square

Despite the extremely good progress in function value, we have to worry about the distance to the central path and we cannot yet appreciate the new point $y(\hat{\alpha})$. Indeed, for getting close again to the central path, we need to find an approximate solution to the auxiliary problem

$$\min_y \{f(y) : \langle b, y \rangle = \langle b, y(\hat{\alpha}) \rangle\}.$$

In order to estimate the complexity of this *corrector stage*, we need to develop some bounds on the growth of the gradient proximity measure.

4 Recession coefficient of barrier function

Definition 1 We call γ_F recession coefficient of the normal barrier F if it is the smallest positive constant such that for every $x \in \text{int } K$ and $u \in K$ we have

$$\nabla^2 F(x+u) \preceq \gamma_F \cdot \nabla^2 F(x). \quad (4.1)$$

Clearly, $\gamma_F \geq 1$. On the other hand, in view of Corollary 2, we have

$$\gamma_F \leq 4\nu^2. \quad (4.2)$$

However, very often this upper bound is very pessimistic. Note that the following main operations with convex cones *do not increase* this coefficient.

Theorem 4 1. Let F be a normal barrier for the cone K , and

$$K_A = \{x \in K : Ax = 0\}.$$

Denote by f the restriction of F onto the relative interior of K_A . Then $\gamma_f \leq \gamma_F$.

2. Let F_i , $i = 1, 2$, be normal barriers for cones $K_i \subset \mathbb{E}$. Denote $F = F_1 + F_2$. If $\text{int}(K_1 \cap K_2) \neq \emptyset$, then $\gamma_F \leq \max\{\gamma_{F_1}, \gamma_{F_2}\}$.
3. Let F_i , $i = 1, 2$, be normal barriers for cones $K_i \subset \mathbb{E}_i$. Denote $F(x, y) = F_1(x) + F_2(y)$. Then $\gamma_F \leq \max\{\gamma_{F_1}, \gamma_{F_2}\}$.

Thus, all barriers constructed as sums or direct products of small-dimensional cones have small recession coefficients. On the other hand, restriction of such barriers onto linear subspaces does not increase the recession coefficient. It remains to note that there exists an important family of normal barriers with *minimal* value of the recession coefficient.

Definition 2 Let F be a normal barrier for the regular cone K . We say that F has negative curvature if for every $x \in \text{int } K$ and $h \in K$ we have

$$\nabla^3 F(x)[h] \preceq 0. \quad (4.3)$$

Thus, for such a barrier $\gamma_F = 1$. It is clear that self-scaled barriers have negative curvature (see [12]). Some other important barriers, like the logarithms of *hyperbolic polynomials* (see [4]) also share this property.

Theorem 5 *Let K be a regular cone and F be a normal barrier for K . Then, the following statements are equivalent:*

1. F has negative curvature;
2. for every $x \in \text{int } K$ and $h \in \mathbb{E}$ we have

$$-\nabla^3 F(x)[h, h] \in K^*; \quad (4.4)$$

3. for every $x \in \text{int } K$ and for every $h \in \mathbb{E}$ such that $x + h \in \text{int } K$, we have

$$\nabla F(x + h) - \nabla F(x) \preceq_{K^*} \nabla^2 F(x)h. \quad (4.5)$$

Proof:

Let F have negative curvature. Then, for every $h \in \mathbb{E}$ and $u \in K$ we have

$$0 \geq \nabla^3 F(x)[h, h, u] = \langle \nabla^3 F(x)[h, h], u \rangle. \quad (4.6)$$

Clearly, this condition is equivalent to (4.4). On the other hand, from (4.4) we have

$$\nabla F(x + h) - \nabla F(x) - \nabla^2 F(x)h = \int_0^1 \nabla^3 F(x + \tau h)[h, h] d\tau \preceq_{K^*} 0.$$

Note that we can replace in (4.5) h by τh , divide everything by τ^2 , and take the limit as $\tau \rightarrow 0^+$. Then we arrive back at the inclusion (4.4). \square

Theorem 6 *Let the curvature of F be negative. Then for every $x \in K$, we have*

$$\nabla^2 F(x)h \succeq_{K^*} 0, \quad \forall h \in K, \quad (4.7)$$

and, consequently,

$$\nabla F(x + h) - \nabla F(x) \succeq_{K^*} 0. \quad (4.8)$$

Proof:

Let us prove that $\nabla^2 F(x)h \in K^*$ for $h \in K$. Assume first that $h \in \text{int } K$. Consider the following vector function:

$$s(t) = \nabla^2 F(x + th)h \in \mathbb{E}^*, \quad t \geq 0.$$

Note that $s'(t) = \nabla^3 F(x + th)[h, h] \stackrel{(4.4)}{\preceq_{K^*}} 0$. This means that

$$\nabla^2 F(x)h \succeq_{K^*} \nabla^2 F(x + th)h \stackrel{(1.10)}{=} \frac{1}{t^2} \nabla^2 F(h + \frac{1}{t}x)h.$$

Taking the limit as $t \rightarrow \infty$, we get $\nabla^2 F(x)h \in K^*$. By continuity arguments, we can extend this inclusion onto all $h \in K$. Therefore,

$$\nabla F(x + h) = \nabla F(x) + \int_0^1 \nabla^2 F(x + \tau h)h d\tau \succeq_{K^*} \nabla F(x).$$

\square

As we have proved, if F has negative curvature, then $\nabla^2 F(x)K \subseteq K^*$, for every $x \in \text{int } K$. This property implies that the situations when both F and F_* has negative curvature are very seldom.

Lemma 9 *Let both F and F^* have negative curvature. Then K is a symmetric cone.*

Proof:

Indeed, for every $x \in \text{int } K$ we have $\nabla^2 F(x)K \subseteq K^*$. Denote $s = -\nabla F(x)$. Since F_* has negative curvature, then $\nabla^2 F_*(s)K^* \subseteq K$. However, since $\nabla^2 F_*(s) \stackrel{(1.15)}{=} [\nabla^2 F(x)]^{-1}$, this means $K^* \subseteq \nabla^2 F(x)K$. Thus $K^* = \nabla^2 F(x)K$. Now, using the same arguments as in [12] it is easy to prove that for every pair $x \in \text{int } K$ and $s \in \text{int } K^*$ there exists a scaling point $w \in \text{int } K$ such that $s = \nabla^2 F(w)x$ (this w can be taken as the minimizer of the convex function $-\langle \nabla F(w), x \rangle + \langle s, w \rangle$). Thus, we have proved that K is homogeneous and self-dual. Hence, it is symmetric. \square

Remark 2 *Lemma 9 shows that the value $\gamma_F + \gamma_{F^*} - 2$ can be seen as a measure of the distance between the pair (F, F_*) and the family of self-scaled barriers.*

The following statement demonstrates the importance of the recession coefficient.

Theorem 7 *Let K be a regular cone and F be a normal barrier for K . Further let $x, x + h \in \text{int } K$. Then for every $\alpha \in [0, 1)$ we have*

$$\frac{1}{\gamma_F(1+\alpha\sigma_x(h))^2} \nabla^2 F(x) \preceq \nabla^2 F(x + \alpha h) \preceq \frac{\gamma_F}{(1-\alpha)^2} \nabla^2 F(x). \quad (4.9)$$

Proof:

Indeed,

$$\begin{aligned} \nabla^2 F(x + \alpha h) &= \nabla^2 F((1 - \alpha)x + \alpha(x + h)) \\ &\stackrel{(4.1)}{\preceq} \gamma_F \cdot \nabla^2 F((1 - \alpha)x) \stackrel{(1.10)}{=} \frac{\gamma_F}{(1-\alpha)^2} \nabla^2 F(x). \end{aligned}$$

Further, denote $\bar{x} = x - \frac{h}{\sigma_x(h)}$. By definition, $\bar{x} \in K$. Note that

$$x = (x + \alpha h) + \frac{\alpha\sigma_x(h)}{1+\alpha\sigma_x(h)}(\bar{x} - (x + \alpha h)).$$

Therefore, by the second inequality in (4.9), we have

$$\nabla^2 F(x) \preceq \gamma_F(1 + \alpha\sigma_x(h))^2 \nabla^2 F(x + \alpha h).$$

\square

5 Bounding the growth of the proximity measure

Let us analyze now our predictor step

$$y(\alpha) = y + \alpha v(y), \quad \alpha \in [0, \bar{\alpha}],$$

where $\bar{\alpha} = \bar{\alpha}(y)$. Denote $\bar{s} = s(y(\bar{\alpha})) \in K^*$.

Lemma 10 For every $\alpha \in [0, \bar{\alpha})$, we have

$$\delta_y(\alpha) \stackrel{\text{def}}{=} \|\nabla f(y(\alpha)) - \frac{\bar{\alpha}}{\bar{\alpha}-\alpha} \nabla f(y)\|_G \leq \frac{\alpha \gamma_{F_*}}{\bar{\alpha}-\alpha} \|\nabla^2 F_*(s(y)) \bar{s}\|_B. \quad (5.1)$$

Proof:

Indeed,

$$\begin{aligned} \delta_y^2(\alpha) &= \langle G^{-1}(\nabla f(y(\alpha)) - \frac{\bar{\alpha}}{\bar{\alpha}-\alpha} \nabla f(y)), \nabla f(y(\alpha)) - \frac{\bar{\alpha}}{\bar{\alpha}-\alpha} \nabla f(y) \rangle \\ &= \langle G^{-1}A^*(\nabla F_*(s(y(\alpha))) - \frac{\bar{\alpha}}{\bar{\alpha}-\alpha} \nabla F_*(s(y))), A^*(\nabla F_*(s(y(\alpha))) - \frac{\bar{\alpha}}{\bar{\alpha}-\alpha} \nabla F_*(s(y))) \rangle \\ &\stackrel{(2.12)}{\leq} \langle B(\nabla F_*(s(y(\alpha))) - \nabla F_*((1 - \frac{\alpha}{\bar{\alpha}})s(y))), \nabla F_*(s(y(\alpha))) - \nabla F_*((1 - \frac{\alpha}{\bar{\alpha}})s(y)) \rangle. \end{aligned}$$

Note that $y(\alpha) = y + \frac{\alpha}{\bar{\alpha}}(y(\bar{\alpha}) - y)$. Therefore

$$s(y(\alpha)) = (1 - \frac{\alpha}{\bar{\alpha}})s(y) + \frac{\alpha}{\bar{\alpha}}\bar{s}.$$

Denote $s' = (1 - \frac{\alpha}{\bar{\alpha}})s(y)$ and $d = \frac{\alpha}{\bar{\alpha}}\bar{s}$. Then

$$\nabla F_*(s(y(\alpha))) - \nabla F_*((1 - \frac{\alpha}{\bar{\alpha}})s(y)) = \left(\int_0^1 \nabla^2 F_*(s' + \tau d) d\tau \right) \cdot d \stackrel{\text{def}}{=} C \cdot d.$$

Note that $0 \leq C \stackrel{(4.1)}{\leq} \gamma_{F_*} \cdot \nabla^2 F_*(s')$. Therefore,

$$\begin{aligned} \delta_y^2(\alpha) &\leq \langle BCd, Cd \rangle \leq \gamma_{F_*}^2 \left(\frac{\alpha}{\bar{\alpha}}\right)^2 \langle B\nabla^2 F_*(s')\bar{s}, \nabla^2 F_*(s')\bar{s} \rangle \\ &\stackrel{(1.10)}{=} \gamma_{F_*}^2 \left(\frac{\alpha}{\bar{\alpha}-\alpha}\right)^2 \|\nabla^2 F_*(s(y))\bar{s}\|_B^2. \end{aligned}$$

□

Note that at the predictor stage, we need to choose the rate of decrease of the penalty parameter (central path parameter) μ as a function of the predictor step size α . Inequality (5.1) suggests the following dependence:

$$\mu(\alpha) \approx \left(1 - \frac{\alpha}{\bar{\alpha}}\right) \cdot \mu. \quad (5.2)$$

However, if $\bar{\alpha}$ is close to its lower limit (3.4), this strategy may be too aggressive. Indeed, in a small neighborhood of the point y we can guarantee only

$$\begin{aligned} \|\nabla f(y(\alpha)) - (1 + \alpha)\nabla f(y)\|_y &= \|\nabla f(y(\alpha)) - \nabla f(y) - \alpha \nabla^2 f(y)v(y)\|_y \\ &\stackrel{(1.7)}{\leq} \frac{\alpha^2 \|v(y)\|_y^2}{1 - \alpha \|v(y)\|_y}. \end{aligned} \quad (5.3)$$

In this situation, a more reasonable strategy for decreasing μ looks as follows:

$$\mu(\alpha) \approx \frac{\mu}{1+\alpha}. \quad (5.4)$$

It appears that it is possible to combine both strategies (5.2) and (5.4) in a single expression. Denote

$$\xi_{\bar{\alpha}}(\alpha) = 1 + \frac{\alpha\bar{\alpha}}{\bar{\alpha}-\alpha}, \quad \alpha \in [0, \bar{\alpha}).$$

Note that

$$\xi_{\bar{\alpha}}(\alpha) = 1 + \alpha + \frac{\alpha^2}{\bar{\alpha}-\alpha} = \frac{\bar{\alpha}}{\bar{\alpha}-\alpha} - \frac{\alpha(1-\bar{\alpha})}{\bar{\alpha}-\alpha}. \quad (5.5)$$

Let us prove an upper bound for the growth of the local gradient proximity measure along direction $v(y)$, when the penalty parameter is dropped by the factor $\xi_{\bar{\alpha}}(\alpha)$.

Theorem 8 *Let $y \in \mathcal{N}(\mu, \beta)$ with $\mu \in (0, 1]$ and $\beta \in (0, \frac{1}{2})$. Then for $y(\alpha) = y + \alpha v(y)$ with $\alpha \in (0, \bar{\alpha})$ we have*

$$\begin{aligned} \gamma\left(y(\alpha), \frac{\mu}{\xi_{\bar{\alpha}}(\alpha)}\right) &\leq \Gamma_{\mu}(y, \alpha) \stackrel{\text{def}}{=} \gamma_{F_*}^{1/2}(1 + \alpha \cdot \sigma_{s(y)}(-A^*v(y))) \|\nabla f(y(\alpha)) - \frac{\xi_{\bar{\alpha}}(\alpha)}{\mu} \cdot b\|_y \\ &\leq \gamma_{F_*}^{1/2}(1 + \alpha \cdot \sigma_{s(y)}(-A^*v(y))) \left[\gamma_1(\alpha) + \beta \cdot \left(1 + \frac{\alpha\bar{\alpha}}{\bar{\alpha}-\alpha}\right) \right], \\ \gamma_1(\alpha) &\stackrel{\text{def}}{=} \|\nabla f(y(\alpha)) - \xi_{\bar{\alpha}}(\alpha) \nabla f(y)\|_y \\ &\leq \frac{\alpha\mu}{\bar{\alpha}-\alpha} \cdot \left(\kappa\sqrt{\nu} + \frac{2\kappa_1\gamma_{F_*}}{\gamma_d} \cdot \left[\frac{\sigma_d(1-\beta)^2}{(1-2\beta)^2} + 2\kappa\nu(1 + \sqrt{\nu}) \frac{1-\beta}{1-2\beta} \right] \right). \end{aligned}$$

Proof:

Indeed,

$$\begin{aligned} \gamma\left(y(\alpha), \frac{\mu}{\xi_{\bar{\alpha}}(\alpha)}\right) &= \|\nabla f(y(\alpha)) - \frac{\xi_{\bar{\alpha}}(\alpha)}{\mu} \cdot b\|_{y(\alpha)} \\ &\stackrel{(4.9)}{\leq} \gamma_{F_*}^{1/2}(1 + \alpha\sigma_{s(y)}(-A^*v(y))) \cdot \|\nabla f(y(\alpha)) - \frac{\xi_{\bar{\alpha}}(\alpha)}{\mu} \cdot b\|_y. \end{aligned}$$

Further,

$$\|\nabla f(y(\alpha)) - \frac{\xi_{\bar{\alpha}}(\alpha)}{\mu} \cdot b\|_y \leq \gamma_1(\alpha) + \xi_{\bar{\alpha}}(\alpha) \|\nabla f(y) - \frac{1}{\mu} b\|_y.$$

Since $y \in \mathcal{N}(\mu, \beta)$, the last term does not exceed $\beta \cdot \xi_{\bar{\alpha}}(\alpha)$. Let us estimate now $\gamma_1(\alpha)$.

$$\begin{aligned} \gamma_1(\alpha) &\stackrel{(5.5)}{\leq} \|\nabla f(y(\alpha)) - \frac{\bar{\alpha}}{\bar{\alpha}-\alpha} \nabla f(y)\|_y + \frac{\alpha(1-\bar{\alpha})}{\bar{\alpha}-\alpha} \|\nabla f(y)\|_y \\ &\stackrel{(3.5)}{\leq} \|\nabla f(y(\alpha)) - \frac{\bar{\alpha}}{\bar{\alpha}-\alpha} \nabla f(y)\|_y + \frac{\alpha}{\bar{\alpha}-\alpha} \cdot \frac{\kappa\mu\sqrt{\nu}}{1+\kappa\mu}. \end{aligned}$$

For the second inequality above, we also used (1.2). Note that

$$\begin{aligned} &\|\nabla f(y(\alpha)) - \frac{\bar{\alpha}}{\bar{\alpha}-\alpha} \nabla f(y)\|_y^2 \\ &= \langle [\nabla^2 f(y)]^{-1} (\nabla f(y(\alpha)) - \frac{\bar{\alpha}}{\bar{\alpha}-\alpha} \nabla f(y)), \nabla f(y(\alpha)) - \frac{\bar{\alpha}}{\bar{\alpha}-\alpha} \nabla f(y) \rangle \\ &\stackrel{(2.19)}{\leq} \frac{4}{\gamma_d^2} [f^* - \langle b, y \rangle]^2 \cdot \|\nabla f(y(\alpha)) - \frac{\bar{\alpha}}{\bar{\alpha}-\alpha} \nabla f(y)\|_G^2 \stackrel{(2.24)}{\leq} \frac{4\kappa_1^2\mu^2}{\gamma_d^2} \cdot \delta_y^2(\alpha). \end{aligned}$$

Moreover,

$$\begin{aligned}
\delta_y(\alpha) &\stackrel{(5.1)}{\leq} \frac{\alpha\gamma_{F_*}}{\bar{\alpha}-\alpha} \|\nabla^2 F_*(s(y))\bar{s}\|_B \\
&\leq \frac{\alpha\gamma_{F_*}}{\bar{\alpha}-\alpha} [\|\nabla^2 F_*(s(y))s^*\|_B + \|\nabla^2 F_*(s(y))(\bar{s} - s^*)\|_B] \\
&\stackrel{(2.23)}{\leq} \frac{\alpha\gamma_{F_*}}{\bar{\alpha}-\alpha} \left[\frac{\sigma_d(1-\beta)^2}{(1-2\beta)^2} + \|\nabla^2 F_*(s(y))(\bar{s} - s^*)\|_B \right].
\end{aligned}$$

It remains to estimate the last term.

Denote $r = \|s(y) - s_\mu\|_{s(y)} \stackrel{(1.4)}{\leq} \frac{\beta}{1-\beta}$. Then

$$\begin{aligned}
B &\stackrel{(2.16)}{\leq} 4\nu^2 \nabla^2 F(x_\mu) \stackrel{(2.5)}{=} 4\nu^2 \nabla^2 F(-\mu \nabla F_*(s_\mu)) \\
&\stackrel{(1.10),(1.15)}{=} \frac{4\nu^2}{\mu^2} [\nabla^2 F_*(s_\mu)]^{-1} \stackrel{(1.5)}{\leq} \frac{4\nu^2}{\mu^2(1-r)^2} [\nabla^2 F_*(s(y))]^{-1}.
\end{aligned}$$

Therefore,

$$\|\nabla^2 F_*(s(y))(\bar{s} - s^*)\|_B \leq \frac{2\nu}{\mu(1-r)} \|y(\bar{\alpha}) - y^*\|_y \stackrel{(3.6)}{\leq} 2\kappa\nu(1 + \sqrt{\nu}) \cdot \frac{1-\beta}{1-2\beta}.$$

Putting all the estimates together, we obtain the claimed upper bound on $\gamma_1(\alpha)$. \square

Taking into account definition of $\xi_{\bar{\alpha}}(\alpha)$, we can see that our predictor-corrector scheme with centering parameter $\beta = O(\mu)$ has local quadratic convergence.

6 Polynomial-time path-following method

Let us show now that the predictor-corrector strategy described in Section 5 has polynomial-time complexity.

Lemma 11 *Let $y \in \mathcal{N}(\mu, \beta)$ with $\beta \leq \frac{1}{18\gamma_{F_*}^{1/2}}$. Then for all*

$$\alpha \in \left[0, \frac{1}{6\gamma_{F_*}^{1/2} \max\{1, \|v(y)\|_y\}} \right] \tag{6.1}$$

we have $\Gamma_\mu(y, \alpha) \leq \beta' \stackrel{\text{def}}{=} \frac{1}{6}$.

Proof:

Denote $r = \|v(y)\|_y$, and $\hat{r} = \max\{1, \|v(y)\|_y\}$. For any $\alpha \in [0, \frac{1}{\hat{r}}]$ we have

$$\begin{aligned}
\Gamma_\mu(y, \alpha) &\stackrel{(1.19)}{\leq} \gamma_{F_*}^{1/2} (1 + \alpha r) \cdot \|\nabla f(y(\alpha)) - \frac{\xi_{\bar{\alpha}}(\alpha)}{\mu} \cdot b\|_y \\
&\stackrel{(5.5)}{\leq} \gamma_{F_*}^{1/2} (1 + \alpha r) \cdot \left(\|\nabla f(y(\alpha)) - (1 + \alpha) \nabla f(y)\|_y + \frac{\alpha^2 r}{\bar{\alpha} - \alpha} + \xi_{\bar{\alpha}}(\alpha) \|\nabla f(y) - \frac{1}{\mu} \cdot b\|_y \right) \\
&\stackrel{(5.3),(3.4)}{\leq} \gamma_{F_*}^{1/2} (1 + \alpha r) \cdot \left(\frac{2\alpha^2 r^2}{1 - \alpha r} + \beta \cdot \left[1 + \frac{\alpha}{1 - \alpha r} \right] \right) \leq \gamma_{F_*}^{1/2} (1 + \alpha \hat{r}) \cdot \frac{2\alpha^2 \hat{r}^2 + \beta}{1 - \alpha \hat{r}}.
\end{aligned}$$

Hence, $\Gamma_\mu(y, \alpha) \stackrel{(6.1)}{\leq} \gamma_{F_*}^{1/2} \left(1 + \frac{1}{6} \gamma_{F_*}^{-1/2}\right) \cdot \frac{\frac{2}{36} \gamma_{F_*}^{-1} + \frac{1}{18} \gamma_{F_*}^{-1/2}}{1 - \frac{1}{6} \gamma_{F_*}^{-1/2}}$. Note that the maximum of the right-hand side of this inequality is attained for $\gamma_{F_*} = 1$. Therefore,

$$\Gamma_\mu(y, \alpha) \leq \left(1 + \frac{1}{6}\right) \cdot \frac{6}{5} \cdot \frac{1}{9} = \frac{7}{45} < \frac{1}{6}.$$

□

By Lemma 11, we can justify the polynomial-time complexity of the predictor-corrector methods. For the sake of simplicity, let us consider a simple short-step path-following method applied to a barrier with $\gamma_{F_*} = 1$. We set

$$\begin{aligned} \alpha_k &= \frac{1}{6 \cdot \max\{1, \|v(y_k)\|_{y_k}\}}, & \mu_{k+1} &= \frac{\mu_k}{\xi_{\bar{\alpha}(y_k)}(\alpha_k)}, \\ p_k &= y_k + \alpha_k [\nabla^2 f(y_k)]^{-1} \nabla f(y_k), \\ y_{k+1} &= p_k - [\nabla^2 f(p_k)]^{-1} \left[\nabla f(p_k) - \frac{b}{\mu_{k+1}} \right]. \end{aligned} \tag{6.2}$$

Theorem 9 *Let K be a regular cone and F_* be a normal barrier for K^* with negative curvature. Also let $y_0 \in \mathcal{N}(\mu_0, \frac{1}{18})$ for some $\mu_0 > 0$. Then, method (6.2) generates a sequence of feasible points such that*

$$f^* - \langle b, y_k \rangle \leq \mu_0 \kappa_1 \exp \left\{ -\frac{k}{1+6\nu^{1/2}} \right\}. \tag{6.3}$$

Proof:

In view of Lemma 11, we have $\gamma(p_k, \mu_{k+1}) \leq \beta'$. Therefore, a single Newton step decreases the local norm of the gradient as follows:

$$\gamma(y_{k+1}, \mu_{k+1}) \leq \left(\frac{\beta'}{1-\beta'} \right)^2 < \frac{1}{18}.$$

Thus, we have $y_k \in \mathcal{N}(\mu_k, \frac{1}{18})$ for all $k \geq 0$. Therefore,

$$f^* - \langle b, y_k \rangle \stackrel{(2.24)}{\leq} \kappa_1 \cdot \mu_k.$$

It remains to note that $\mu_{k+1} \stackrel{(5.5)}{\leq} \frac{\mu_k}{1+\alpha_k}$, and

$$1 + \alpha_k \stackrel{(1.2)}{\geq} 1 + \frac{1}{6\sqrt{\nu}} \geq \exp \left\{ \frac{1}{1+6\nu^{1/2}} \right\}.$$

□

In order to apply method (6.2) to barriers with $\gamma_{F_*} > 1$, we need to change the formulae for α_k in accordance with (6.1) and introduce several corrector steps ensuring $\gamma(y_{k+1}, \mu_{k+1}) \leq \frac{1}{18}$. Lemma 11 guarantees that p_k belongs to the region of quadratic convergence of the Newton method. Hence, this corrector stage cannot be long.

As we have seen in Theorem 8, method (6.2) can be accelerated. Define the following univariate function:

$$\eta_{\bar{\alpha}}(\alpha) = \begin{cases} 2\alpha, & \alpha \in [0, \frac{1}{3}\bar{\alpha}], \\ \frac{\alpha+\bar{\alpha}}{2}, & \alpha \in [\frac{1}{3}\bar{\alpha}, \bar{\alpha}]. \end{cases} \quad (6.4)$$

This function will be used for updating the length of our predictor step.

Lemma 12 *If $\alpha \geq 0$ and $\alpha_+ = \eta_{\bar{\alpha}}(\alpha)$, then $\xi_{\bar{\alpha}}(\alpha_+) \geq 2\xi_{\bar{\alpha}}(\alpha) - 1$. Hence, for the recurrence*

$$\alpha_{i+1} = \eta_{\bar{\alpha}}(\alpha_i), \quad i \geq 0,$$

we have $\xi_{\bar{\alpha}}(\alpha_i) \geq 1 + \alpha_0 \cdot 2^i$.

Proof:

If $\alpha_+ = 2\alpha$, then $\xi_{\bar{\alpha}}(\alpha_+) = 1 + \frac{2\alpha\bar{\alpha}}{\bar{\alpha}-2\alpha} \geq 1 + \frac{2\alpha\bar{\alpha}}{\bar{\alpha}-\alpha} = 2\xi_{\bar{\alpha}}(\alpha) - 1$. If $\alpha_+ = \frac{\alpha+\bar{\alpha}}{2}$, then

$$\xi_{\bar{\alpha}}(\alpha_+) = 1 + \frac{\bar{\alpha}(\bar{\alpha}+\alpha)}{\bar{\alpha}-\alpha} = \xi_{\bar{\alpha}}(\alpha) + \frac{\bar{\alpha}^2}{\bar{\alpha}-\alpha} \geq 2\xi_{\bar{\alpha}}(\alpha) - 1.$$

Therefore, $\xi_{\bar{\alpha}}(\alpha_i) \geq 1 + (\xi_{\bar{\alpha}}(\alpha_0) - 1) \cdot 2^i \stackrel{(5.5)}{\geq} 1 + \alpha_0 \cdot 2^i$. □

Consider the following predictor-corrector process.

Path-following method based on recession coefficient

1. Set $\mu_0 = 1$ and find point $y_0 \in \mathcal{N}\left(\mu_0, \frac{1}{18\gamma_{F^*}^{1/2}}\right)$.
 2. For $k \geq 0$ iterate:
 - a) Compute $\bar{\alpha}_k = \bar{\alpha}(y_k)$.
 - b) Using recurrence

$$\alpha_{k,0} = \frac{1}{6 \cdot \gamma_{F^*}^{1/2} \max\{1, \|v(y_k)\|_{y_k}\}}, \quad \alpha_{k,i+1} = \eta_{\bar{\alpha}_k}(\alpha_{k,i}),$$
 find the maximal $i \equiv i_k$ such that $\Gamma_{\mu_k}(y_k, \alpha_{k,i}) \leq \beta'$.
 - c) Set $\alpha_k = \alpha_{k,i_k}$, $p_k = y_k + \alpha_k v(y_k)$, $\mu_{k+1} = \frac{\mu_k}{\xi_{\bar{\alpha}_k}(\alpha_k)}$.
 - d) Starting from p_k , apply the Newton method for finding $y_{k+1} \in \mathcal{N}\left(\mu_{k+1}, \frac{\mu_{k+1}}{18\gamma_{F^*}^{1/2}}\right)$.
- (6.5)

As for method (6.2), we can prove for (6.5) the polynomial complexity bound:

$$f^* - \langle b, y_k \rangle \leq \mu_0 \kappa_1 \exp \left\{ -\frac{k}{1+6\gamma_{F^*}^{1/2}\nu^{1/2}} \right\}. \quad (6.6)$$

On the other hand, in a small neighborhood of the solution, (6.5) can accelerate up to local quadratic convergence.

In this scheme we have two search procedures. The recurrence at Step 2b is trying to maximize the predictor step. Number of iterations in this process cannot be too large. In fact, each successful iteration results in a significant decrease of the penalty parameter (see Lemma 12). Therefore, the rate of convergence (6.6) of method (6.5) can be related

to the total number of steps in this line search procedure. Note that computation of the estimate $\Gamma_{\mu_k}(y_k, \alpha)$ for different values of α is cheap since it does not require new matrix inversions. We pay for that by the presence of factor $\gamma_{F_*}^{1/2}$ in the rate of convergence (6.6). However, we have already argued at the beginning of Section 4, that in the majority of practical problems this factor is a small absolute constant.

The auxiliary minimization process on Step 2d cannot be too long either. Note that the penalty parameters μ_k are bounded from below by $\frac{\kappa_1}{\epsilon}$, where ϵ is the desired accuracy of the solution. On the other hand, the point p_k belongs to the region of quadratic convergence of the Newton method. Therefore, the number of iterations on Step 2d is bounded by $O(\ln \ln \frac{\kappa_1}{\epsilon})$. In Section 7 we will demonstrate on simple examples that the high accuracy in approximating the trajectory of central path is crucial for local quadratic convergence.

It is possible to eliminate both from the scheme (6.5) and the estimate (6.6) the recession coefficient of the barrier function. This can be achieved by increasing the complexity of predictor step.

Path-following method for general barriers

1. Set $\mu_0 = 1$ and find point $y_0 \in \mathcal{N}(\mu_0, \frac{1}{18})$.
2. For $k \geq 0$ iterate:
 - a) Compute $\bar{\alpha}_k = \bar{\alpha}(y_k)$.
 - b) Using recurrence

$$\alpha_{k,0} = \frac{1}{6 \cdot \max\{1, \|v(y_k)\|_{y_k}\}}, \quad \alpha_{k,i+1} = \eta_{\bar{\alpha}_k}(\alpha_{k,i}), \tag{6.7}$$
 find the maximal $i \equiv i_k$ such that $\gamma\left(y_k(\alpha_{k,i}), \frac{\mu_k}{\xi_{\bar{\alpha}_k}(\alpha_{k,i})}\right) \leq \beta'$.
 - c) Set $\alpha_k = \alpha_{k,i_k}$, $p_k = y_k + \alpha_k v(y_k)$, $\mu_{k+1} = \frac{\mu_k}{\xi_{\bar{\alpha}_k}(\alpha_k)}$.
 - d) Starting from p_k , apply the Newton method for finding $y_{k+1} \in \mathcal{N}(\mu_{k+1}, \frac{1}{18}\mu_{k+1})$.

In this scheme, for computing the value of gradient proximity measure at new points, we need to compute and invert the Hessian of barrier function. However, the step size in this procedure is rapidly increased. Therefore, it is easy to prove that the total number of auxiliary steps i_k , which is necessary for computing an ϵ -solution to our problem is bounded by $O(\nu^{1/2} \ln \frac{\nu}{\epsilon})$. As in method (6.5), the number of steps at the correction stage (Step 2d) cannot be large since p_k belongs to the region of quadratic convergence of the Newton method. In any case, if Assumptions 1 and 2 are satisfied, then method (6.7) is locally quadratically convergent.

7 Discussion

7.1 2D-examples

Let us look now at several 2D-examples illustrating different aspects of our approach. Let us start with the following problem:

$$\max_{y \in \mathbb{R}^2} \{\langle b, y \rangle : y_2 \geq 0, y_1 \geq y_2^2\}. \tag{7.1}$$

For this problem, we can use the following barrier function:

$$f(y) = -\ln(y_1 - y_2^2) - \ln y_2.$$

We are going to check our conditions for the optimal point $y_* = 0$.

Problem (7.1) can be seen as a restriction of the following conic problem:

$$\max_{s,y} \{ \langle b, y \rangle : s_1 = y_1, s_2 = y_2, s_3 = 1, s_4 = y_2, s_1 s_3 \geq s_2^2, s_4 \geq 0 \}, \quad (7.2)$$

endowed with the barrier $F_*(s) = -\ln(s_1 s_3 - s_2^2) - \ln s_4$. Note that

$$\nabla F_*(s) = \left(\frac{-s_3}{s_1 s_3 - s_2^2}, \frac{2s_2}{s_1 s_3 - s_2^2}, \frac{-s_1}{s_1 s_3 - s_2^2}, \frac{-1}{s_4} \right)^T.$$

Denote $\omega = s_1 s_3 - s_2^2$. Since in problem (7.2) $y_* = 0$ corresponds to $s_* = e_3$, we have the following representation:

$$\nabla^2 F_*(s) s_* = \nabla^2 F_*(s) e_3 = \frac{1}{\omega^2} \cdot (s_2^2, -2s_1 s_2, s_1^2, 0)^T.$$

Let us choose in the primal space the norm

$$\langle Bx, x \rangle = x_1^2 + \frac{1}{2}x_2^2 + x_3^2 + x_4^2.$$

Then $\|\nabla^2 F_*(s) s_*\|_B = [s_1^2 + s_2^2]/\omega^2$. Hence, the region $\|\nabla^2 F_*(s(y)) s_*\|_B \leq \sigma_d$ is formed by vectors $y = (y_1, y_2)$ satisfying the inequality

$$y_1^2 + y_2^2 \leq \sigma_d (y_1 - y_2^2)^2.$$

Thus, the boundary curve of this region is given by equation

$$y_1 = y_2^2 + \frac{1}{\sigma_d^{1/2}} \sqrt{y_1^2 + y_2^2},$$

which has a positive slope $[\sigma_d^{1/2} - 1]^{-1}$ at the origin (see Figure 1). Note that the central path corresponding to the vector $b = (-1, 0)^T$ can be found from the equations

$$\frac{1}{y_1 - y_2^2} = \frac{1}{\mu}, \quad \frac{1}{y_2} = \frac{2y_2}{y_1 - y_2^2}.$$

Thus, its characteristic equation is $y_1 = 3y_2^2$, and, for any value of σ_d , it leaves the region of quadratic convergence as $\mu \rightarrow 0$. It is interesting that in our example Assumption 2 is valid if and only if the problem (7.1) with $y^* = 0$ satisfies Assumption 1.

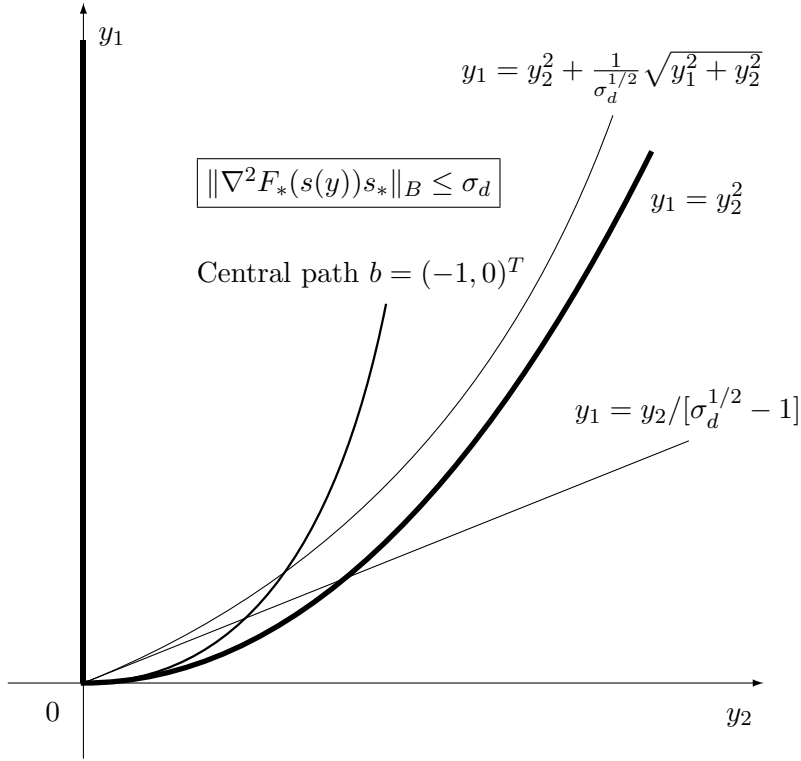


Figure 1. Behavior of $\|\nabla^2 F_*(s(y))s_*\|_B$.

In our second example we need the *maximal neighborhood* of the central path:

$$\begin{aligned} \mathcal{M}(\beta) &= \text{Cl} \left(\bigcup_{\mu \in R} \mathcal{N}(\mu, \beta) \right) \\ &= \left\{ y : \theta^2(y) \stackrel{\text{def}}{=} \|\nabla f(y)\|_y^2 - \frac{1}{\|b\|_y^2} \langle \nabla f(y), [\nabla^2 f(y)]^{-1} b \rangle^2 \leq \beta^2 \right\}. \end{aligned} \quad (7.3)$$

Note that $\theta(y) = \min_{t \in R} \|\nabla f(y) - tb\|_y$.

Consider the following problem:

$$\max_{y \in R^2} \{y_1 : \|y\| \leq 1\}. \quad (7.4)$$

where $\|\cdot\|$ is the standard Euclidean norm. Let us endow the feasible set of this problem with the standard barrier function $f(y) = -\ln(1 - \|y\|^2)$. Note that

$$\begin{aligned} \nabla f(y) &= \frac{2y}{1 - \|y\|^2}, \quad \nabla^2 f(y) = \frac{2I}{1 - \|y\|^2} + \frac{4yy^T}{(1 - \|y\|^2)^2}, \\ [\nabla^2 f(y)]^{-1} &= \frac{1 - \|y\|^2}{2} \left(I - \frac{2yy^T}{1 + \|y\|^2} \right), \quad [\nabla^2 f(y)]^{-1} \nabla f(y) = \frac{1 - \|y\|^2}{1 + \|y\|^2} \cdot y. \end{aligned}$$

Therefore,

$$\|\nabla f(y)\|_y^2 = \frac{2\|y\|^2}{1 + \|y\|^2},$$

and for $b = (1, 0)$ we have

$$\|b\|_y^2 = \frac{1 - \|y\|^2}{2} \cdot \frac{1 - y_1^2 + y_2^2}{1 + \|y\|^2}, \quad \langle \nabla f(y), [\nabla^2 f(y)]^{-1} b \rangle = \frac{1 - \|y\|^2}{1 + \|y\|^2} \cdot y_1.$$

Thus,

$$\begin{aligned}\theta^2(y) &= \frac{2\|y\|^2}{1+\|y\|^2} - \frac{2}{1-\|y\|^2} \cdot \frac{1+\|y\|^2}{1-y_1^2+y_2^2} \cdot \frac{(1-\|y\|^2)^2 y_1^2}{(1+\|y\|^2)^2} \\ &= \frac{2}{1+\|y\|^2} \left(\|y\|^2 - \frac{y_1^2(1-\|y\|^2)}{1-y_1^2+y_2^2} \right) = \frac{2y_2^2}{1-y_1^2+y_2^2}.\end{aligned}$$

We conclude that for problem (7.4) the maximal neighborhood of the central path has the following representation:

$$\mathcal{M}(\beta) = \left\{ y \in R^2 : y_1^2 + \frac{2-\beta^2}{\beta^2} \cdot y_2^2 \leq 1 \right\} \quad (7.5)$$

(see Figure 2).

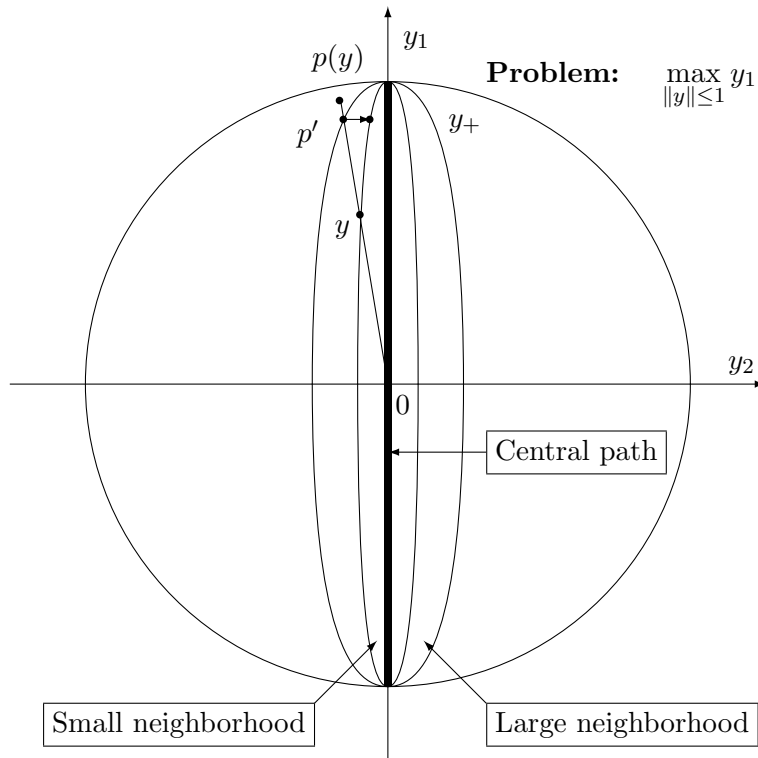


Figure 2. Prediction in the absence of strict maximum.

Note that $p(y) = \frac{2y}{1+\|y\|^2} \in \text{int } Q$. If the radii of the small and large neighborhoods of the central path are fixed, by straightforward computations we can see that the simple predictor-corrector update $y \rightarrow y_+$ shown at Figure 2 has local linear rate of convergence. In order to get a superlinear rate, we need to tighten the small neighborhood of the central path as $\mu \rightarrow 0$.

7.2 Examples of cones with negative curvature

In accordance with the definition (4.3), negative curvature of barrier functions is preserved by the following operations.

- If barriers F_i for cones $K_i \subset \mathbb{E}_i$, $i = 1, 2$, have negative curvature, then the curvature of the barrier $F_1 + F_2$ for the cone $K_1 \oplus K_2$ is negative.
- If barriers F_i for cones $K_i \subset \mathbb{E}$, $i = 1, 2$, have negative curvature, then the curvature of the barrier $F_1 + F_2$ for the cone $K_1 \cap K_2$ is negative.
- If barrier F for cone K has negative curvature, then the curvature of the barrier $f(y) = F(A^*y)$ for the cone $K_y = \{y \in \mathbb{H} : A^*y \in K\}$ is negative.
- If barrier $F(x)$ for cone K has negative curvature, then the curvature of its restriction onto the linear subspace $\{x \in \mathbb{E} : Ax = 0\}$ is negative.

At the same time, we know two important families of cones with negative curvature.

- Self-scaled barriers have negative curvature (see Corollary 3.2(i) in [12]).
- Let $p(x)$ be hyperbolic polynomial. Then the barrier $F(x) = -\ln p(x)$ has negative curvature (see [4]).

Thus, using above mentioned operations, we can construct barriers with negative curvature for many interesting cones. In some situations we can argue that currently, some nonsymmetric treatments of the primal-dual problem pair have better complexity bounds than the primal-dual symmetric treatments.

Example 2 Consider the cone of nonnegative polynomials:

$$K = \left\{ p \in \mathbb{R}^{2n+1} : \sum_{i=0}^{2n} p_i t^i \geq 0, \forall t \in \mathbb{R} \right\}.$$

The dual to this cone is the cone of positive semidefinite Hankel matrices. For $k = 0, \dots, 2n$, denote

$$H_k \in \mathbb{R}^{(n+1) \times (n+1)} : H_k^{(i,j)} = \begin{cases} 1, & \text{if } i+j = k+2 \\ 0, & \text{otherwise} \end{cases}, \quad i, j = 0, \dots, n.$$

For $s \in \mathbb{R}^{2n+1}$ we can define now the following linear operator:

$$H(s) = \sum_{i=0}^{2n} s_i \cdot H_i.$$

Then the cone dual to K can be represented as follows:

$$K^* = \{s \in \mathbb{R}^{2n+1} : H(s) \succeq 0\}.$$

The natural barrier for the dual cone is $f(s) = -\ln \det H(s)$. Clearly, it has negative curvature. Note that we can lift the primal cone to a higher dimensional space (see [10]):

$$K = \{p \in \mathbb{R}^{2n+1} : p_i = \langle H_i, Y \rangle, Y \succeq 0, i = 0, \dots, 2n\},$$

and use $F(Y) = -\ln \det Y$ as a barrier function for the extended feasible set. However, in this case we significantly increase the number of variables. Moreover, we need $O(n^3)$ operations for computing the value of the barrier $F(Y)$ and its gradient. On the other hand, in the dual space the cost of all necessary computations is very low ($O(n \ln^2 n)$ for the function value and $O(n^2 \ln^2 n)$ for solution of the Newton system, see [2]). On top of these advantages, for non-degenerate dual problems, now we have a locally quadratically convergent path-following scheme (6.5).

To conclude the paper, let us mention that the negative curvature seems to be a natural property of self-concordant barriers. Indeed, let us move from some point $x \in \text{int } K$ along the direction $h \in K$: $u = x + h$. Then the Dikin ellipsoid of barrier F at point x , moved to the new center u , still belongs to K :

$$u + (W_r(x) - x) = h + W_r(x) \subset K.$$

We should expect that the Dikin ellipsoid $W_r(u)$ becomes even larger (in any case, we should expect that it does not get smaller). This is exactly the negative curvature condition: $\nabla^2 F(x) \succeq \nabla^2 F(u)$. At this moment, it is not clear if it is an attribute of the barrier or of the cone, or both. In other words, is it possible to construct a barrier with negative curvature for any convex cone? However, we have already seen that for nonsymmetric cones this property is *not* dual-invariant. Another interesting question is related to existence of the barrier function which ensures a small recession coefficient for arbitrary regular cone. Up to now, we do not have examples of cones where the recession coefficient is indeed large (growing with dimension).

References

- [1] H. H. Bauschke, O. Güler, A. S. Lewis and H. Sendov, Hyperbolic polynomials and convex analysis, *Canad. J. Math.* 53 (2001) 470–488.
- [2] Y. Genin, Y. Hachez, Yu. Nesterov, P. Van Dooren, Optimization problems over positive pseudo-polynomial matrices, *SIAM J. Matrix Anal. Appl.* 25 (2003) 57–79.
- [3] A. V. Fiacco, G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, J.Wiley & Sons Inc., New York, 1968.
- [4] O. Güler, Hyperbolic polynomials and interior-point methods for convex programming, *Math. Oper. Res.* 22 (1997) 350–377.
- [5] J. Ji, F. A. Potra and R. Sheng, On the local convergence of a predictor-corrector method for semidefinite programming, *SIAM J. Optim.* 10 (1999) 195–210 .
- [6] M. Kojima, M. Shida and S. Shindoh, Local convergence of predictor-corrector infeasible-interior-point algorithms for SDPs and SDLCPs, *Math. Program.* 80 (1998) 129–160.
- [7] Z.-Q. Luo, J. F. Sturm and Zhang, Superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming, *SIAM J. Optim.* 8 (1998) 59–81.
- [8] S. Mehrotra, Quadratic convergence in a primal-dual method, *Math. Oper. Res.* 18 (1993) 741–751.
- [9] Yu. Nesterov, *Introductory Lectures on Convex Optimization*, Kluwer, Boston, 2004.
- [10] Yu. Nesterov, Squared functional systems and optimization problems, In *High Performance Optimization*, H.Frenk, T.Terlaky and S.Zhang (Eds.), Kluwer, 1999. pp.405–439.
- [11] Yu. Nesterov, A. Nemirovskii, *Interior Point Polynomial Methods in Convex Programming: Theory and Applications*, SIAM, Philadelphia, 1994.

- [12] Yu. Nesterov, M. J. Todd, Self-scaled barriers and interior-point methods for convex programming, *Math. Oper. Res.* 22 (1997) 1–42.
- [13] F. A. Potra and R. Sheng, Superlinear convergence of interior-point algorithms for semidefinite programming, *J. Optim. Theory Appl.* 99 (1998) 103–119.
- [14] J. Renegar, Hyperbolic programs, and their derivative relaxations, *Found. Comput. Math.* 6 (2006) 59–79.
- [15] Y. Ye, O. Güler, R. A. Tapia and Y. Zhang, A quadratically convergent $O(\sqrt{n}L)$ -iteration algorithm for linear programming, *Math. Program.* 59 (1993) 151–162.