Local Superlinear Convergence of Polynomial-Time Interior-Point Methods for Hyperbolicity Cone Optimization Problems

Yu. Nesterov* and L. Tunçel†

November 2009, revised: August 2015

Abstract

In this paper, we establish the local superlinear convergence property of some polynomial-time interior-point methods for an important family of conic optimization problems. The main structural property used in our analysis is the logarithmic homogeneity of self-concordant barrier function, which must have *negative curvature*. We propose a new path-following predictor-corrector scheme, which works only in the dual space. It is based on an easily computable gradient proximity measure, which ensures an automatic transformation of the global linear rate of convergence to the local superlinear one under some mild assumptions. Our step-size procedure for the predictor step is related to the maximum step size maintaining feasibility. As the optimal solution set is approached, our algorithm automatically tightens 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 superlinear convergence.

*Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL), 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium; e-mail: nesterov@core.ucl.ac.be.

The research results presented in this paper have been supported by a grant “Action de recherche concertè ARC 04/09-315” from the “Direction de la recherche scientifique - Communautè française de Belgique”. The scientific responsibility rests with its author(s).

†Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada; e-mail: ltuncel@uwaterloo.ca. Research of this author was supported in part by Discovery Grants from NSERC and by ONR Research Grant N00014-15-1-2171.

1 Introduction

Motivation. Local superlinear convergence is a natural and very desirable 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, to some extent, the delay in developing an appropriate framework for analyzing the local behavior of general polynomial-time interior-point methods [13]. 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 to the particular functional inequality constraints of the initial optimization problems. Therefore, up to now, the local superlinear convergence for polynomial-time path-following methods was proved only for Linear Programming [18, 10] and for Semidefinite Programming problems [6, 16, 9, 5, 15]. 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 superlinear convergence property of interior-point path-following methods by employing some geometric properties of quite general conic optimization problem. The main structural property used in our analysis is the logarithmic homogeneity of self-concordant barrier functions, and the condition that the barrier must have negative curvature. We propose a new path-following predictor-corrector scheme, which works in the dual space. It is based on an easily computable gradient proximity measure, which ensures an automatic transformation of the global linear rate of convergence to the local superlinear rate (under a mild assumption). Our step-size procedure for the predictor step is related to the maximum step size maintaining feasibility. As the iterates approach the optimal solution set, our algorithm automatically tightens the neighborhood of the central path proportionally to the current duality gap. In the literature (as we noted above) similar conditions have been imposed on the iterates of interior-point algorithms in order to attain local superlinear convergence in the Semidefinite Programming setting. However, there exist different, weaker combinations of conditions that have been studied in the Semidefinite Programming setting as well (see for instance [15, 5, 7, 8]). As a key feature of our analysis, we avoid distinguishing individual subvectors as "large" or "small." Many of the ingredients of our approach are primal-dual symetric; however, we break the symmetry, when necessary, by our choice of assumptions and algorithms. Indeed, in general (beyond the special case of symmetric cones and self-scaled barriers), only one of the primal, dual problems may admit a logarithmically homogeneous self-concordant barrier with negative curvature.

Contents. The paper is organized as follows. In Section 2 we introduce a conic primal-dual problem and define the central path. After that, we consider a small full-dimensional dual problem and define the prediction operator. We derive some representations, which help us bound the curvature of the central path. In Section 3 we justify the choice of the fixed Euclidean metrics in the primal and dual spaces. Then, in Section 4 we introduce two main assumptions ensuring the quadratic drop in the duality gap for predicted points from a tight neighborhood of the central path. The first one is on the sharpness of the 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 4.2 which demonstrates the quadratic decrease of the distance to the optimal solution for the prediction point, measured in an appropriately chosen fixed Euclidean norm. In Section 5, 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 6 we prepare for analysis of polynomial-time predictor-corrector strategies. For that, we study an important class of barriers with *negative curvature*. This class includes at least the self-scaled barriers [14] and hyperbolic barriers [4, 1, 17], possibly many others. In Section 7 we establish some bounds on the growth of a variant of the gradient proximity measure. We show that we can achieve a local superlinear rate of convergence. It is important to relate the decrease of penalty parameter of the central path with the *distance to the boundary* of the feasible set while performing the predictor step. At the same time, we show that, for achieving the local superlinear convergence, the centering condition must be satisfied with increasing accuracy.

In Section 8 we show that the local superlinear convergence can be combined with the global polynomial-time complexity. We present a method, which works for the barriers with negative curvature, and has a desired property of cheap computation of the predictor step. Finally, in Section 9, 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 a linear transformation $A : \mathbb{E} \to \mathbb{H}^*$ we denote by A^* the corresponding *adjoint* linear transformation:

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

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

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

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, this means that $r = 1$.

In what follows, we often use the following simple result from Linear Algebra. Let self-adjoint linear operators Δ and B map $\mathbb{E}$ to $\mathbb{E}^*$, and $B \succ 0$. Then, for every tolerance parameter $\tau > 0$ we have

$$\pm \Delta \preceq \tau B \quad \Leftrightarrow \quad \Delta B^{-1} \Delta \preceq \tau^2 B. \tag{1.1}$$

For future reference, let us recall some facts from the theory of self-concordant functions. Most of these results can be found in Section 4 in [11]. 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 $Q \subset \mathbb{E}$:

$$\left| \nabla^3 \Phi(x)[h, h, h] \right| \;\; \leq \;\; 2 \langle \nabla^2 \Phi(x) h, h \rangle^{3/2}, \quad x \in \operatorname{int} Q, \; h \in \mathbb{E}, \tag{1.2}$$

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 every $u \in \operatorname{int} Q$. Self-concordant function Φ is called a ν-self-concordant barrier if

$$\langle \nabla \Phi(u), [\nabla^2 \Phi(u)]^{-1} \nabla \Phi(u) \rangle \;\; \leq \;\; \nu. \tag{1.3}$$

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 (1.3) can be written as $\| \nabla \Phi(u) \|_u^2 \leq \nu$. The following result is very useful.

Theorem 1.1 *(Theorem on Recession Direction; see Section 4 in [11] for the proof.) If h is a recession direction of the set Q and $u \in \operatorname{int} Q$, then*

$$\| h \|_u \;\; \leq \;\; \langle -\nabla \Phi(u), h \rangle. \tag{1.4}$$

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

$$\langle \nabla\Phi(v) - \nabla\Phi(u), v-u\rangle \;\geq\; \tfrac{r^2}{1+r}, \quad r \geq 0. \tag{1.5}$$

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

$$(1-r)^2\nabla^2\Phi(u) \;\preceq\; \nabla^2\Phi(v) \;\preceq\; \tfrac{1}{(1-r)^2}\nabla^2\Phi(u), \tag{1.6}$$

$$\|\nabla\Phi(v) - \nabla\Phi(u)\|_u \;\leq\; \tfrac{r}{1-r}, \tag{1.7}$$

$$\|\nabla\Phi(v) - \nabla\Phi(u) - \nabla^2\Phi(u)(v-u)\|_u \;\leq\; \tfrac{r^2}{1-r}. \tag{1.8}$$

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:

$$F(\tau x) \;=\; F(x) - \nu \ln \tau, \quad x \in \operatorname{int} K, \; \tau > 0. \tag{1.9}$$

Note that $-\nabla F(x) \in \operatorname{int} K^*$ for every $x \in \operatorname{int} K$. Equality (1.9) leads to many interesting identities:

$$\nabla F(\tau x) \;=\; \tau^{-1} \cdot \nabla F(x), \tag{1.10}$$

$$\nabla^2 F(\tau x) \;=\; \tau^{-2} \cdot \nabla^2 F(x), \tag{1.11}$$

$$\langle \nabla F(x), x\rangle \;=\; -\nu, \tag{1.12}$$

$$\nabla^2 F(x) \cdot x \;=\; -\nabla F(x), \tag{1.13}$$

$$\nabla^3 F(x)[x] \;=\; -2\nabla^2 F(x), \tag{1.14}$$

$$\|\nabla F(x)\|_x^2 \;=\; \nu, \tag{1.15}$$

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

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

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

$$\begin{array}{rclcrcl} \nabla F(-\nabla F_*(s)) & = & -s, & & \nabla^2 F(-\nabla F_*(s)) & = & [\nabla^2 F_*(s)]^{-1}, \\ \nabla F_*(-\nabla F(x)) & = & -x, & & \nabla^2 F_*(-\nabla F(x)) & = & [\nabla^2 F(x)]^{-1}, \end{array} \tag{1.16}$$

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

For normal barriers, the Theorem on Recession Direction (1.4) can be written both in primal and dual forms:

$$\|u\|_x \;\leq\; \langle -\nabla F(x), u\rangle, \quad x \in \operatorname{int} K, u \in K, \tag{1.17}$$

$$\|s\|_x \;\leq\; \langle s, x\rangle, \quad x \in \operatorname{int} K, s \in K^*. \tag{1.18}$$

The following statement is very useful.

Lemma 1.1 *Let F be a ν-normal barrier for K and $H : \mathbb{E} \to \mathbb{E}^*$, $H \succ 0$. Assume that $\mathcal{E}_H(u) \subset K$, and for some $x \in \operatorname{int} K$ we have*

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

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

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

$$\langle \nabla F(x), H^{-1}h \rangle \quad \geq \quad 0, \tag{1.19}$$

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

$$\begin{array}{rcl} \frac{\|H^{-1}h\|_x}{\|h\|_H} & \leq & \|u\|_x + \|y\|_x \stackrel{(1.17)}{\leq} \langle -\nabla F(x), u \rangle + \langle -\nabla F(x), y \rangle \\ & \stackrel{(1.19)}{\leq} & 2\langle -\nabla F(x), u \rangle \;\leq\; 2\langle -\nabla F(x), x \rangle \stackrel{(1.12)}{=} 2\nu. \end{array}$$

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

Corollary 1.1 *Let $x, u \in \operatorname{int} K$ 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 1.2 *Let $x \in \operatorname{int} K$ and $u \in K$. Then $\nabla^2 F(x+u) \preceq 4\nu^2 \nabla^2 F(x)$.*

Proof:
Denote $y = x + u \in \operatorname{int} K$. Then $\langle \nabla F(y), x - y \rangle = \langle -\nabla F(y), u \rangle \geq 0$. Hence, we can apply Corollary 1.1. □

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

$$\sigma_x(h) \quad = \quad \min_{\rho \geq 0} \{ \rho : \; \rho \cdot x - h \in K \} \;\leq\; \|h\|_x, \quad x \in \operatorname{int} K, \; h \in \mathbb{E}. \tag{1.20}$$

2 Predicting the optimal solution

Consider the standard conic optimization problem:

$$\min_{x \in K} \{ \; \langle c, x \rangle : \; Ax = b \; \}, \tag{2.1}$$

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. If A is not surjective, we either find that $Ax = b$ has no solution (implying our optimization problem is infeasible) or detect and eliminate all redundant equations in the system $Ax = b$, and redefine A and b without changing the set of feasible solutions and the set of optimal solutions. Therefore, we assume, without loss of generality, that A is surjective.

The problem dual to (2.1) is then

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

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

$$\langle s_1 - s_2, x_1 - x_2 \rangle \quad = \quad 0 \tag{2.3}$$

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

Under the *strict feasibility* assumption,

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

the optimal sets of the primal and dual problems are nonempty and bounded, and there is no duality gap (see for instance [13]). Moreover, a *primal-dual central path* $z_\mu \stackrel{\mathrm{def}}{=} (x_\mu, s_\mu, y_\mu)$:

$$\left. \begin{array}{rcl} Ax_\mu & = & b \\ c + \mu \nabla F(x_\mu) & = & A^* y_\mu \\ s_\mu \; = \; -\mu \nabla F(x_\mu) & \stackrel{(1.16),(1.10)}{\Leftrightarrow} & x_\mu \; = \; -\mu \nabla F_*(s_\mu) \end{array} \right\}, \quad \mu > 0, \tag{2.5}$$

is well defined. Note that

$$\langle c, x_\mu \rangle - \langle b, y_\mu \rangle \;=\; \langle s_\mu, x_\mu \rangle \;\overset{(2.5),(1.12)}{=}\; \nu \cdot \mu. \tag{2.6}$$

The majority of modern strategies for solving the primal-dual problem pair (2.1), (2.2) suggest to follow this trajectory as $\mu \to 0$. On the one hand, it is important that μ be decreased at a linear rate to attain a polynomial-time complexity. However, on the other hand, in a small neighborhood of the solution, it is highly desirable to switch on a superlinear rate. Such a possibility was already discovered for Linear Programming problems [19, 18, 10]. There has also been significant progress in the case of Semidefinite Programming [6, 16, 9, 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

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

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:

$$s \;=\; s(y) \;\overset{\mathrm{def}}{=}\; c - A^* y.$$

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

$$\begin{array}{rcl} f^* & \overset{\mathrm{def}}{=} & \max\limits_{y \in \mathbb{H}} \{ \langle b, y \rangle : \; y \in Q \}, \\[2ex] Q & \overset{\mathrm{def}}{=} & \{ y \in \mathbb{H} : \; c - A^* y \in K^* \}. \end{array} \tag{2.7}$$

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

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

Since the optimal set of problem (2.7) is bounded, Q contains no straight line. Thus, this barrier has a nondegenerate Hessian at every strictly feasible point.

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

$$\begin{array}{rcl} b & = & \mu \nabla f(y_\mu) \;=\; -\mu A \nabla F_*(c - A^* y_\mu) \\[2ex] & = & -\mu A \nabla F_*(s_\mu) \;\overset{(2.5)}{=}\; A x_\mu, \quad \mu > 0. \end{array} \tag{2.8}$$

Let us estimate the quality of the following prediction point:

$$\begin{array}{rcl} p(y) & \overset{\mathrm{def}}{=} & y + v(y), \quad y \in \operatorname{int} Q, \\[2ex] v(y) & \overset{\mathrm{def}}{=} & [\nabla^2 f(y)]^{-1} \nabla f(y), \quad s_p(y) \overset{\mathrm{def}}{=} s(y) - A^* v(y). \end{array}$$

Definition of the displacement $v(y)$ is motivated by identity (1.13), which is valid for arbitrary convex cones. Indeed, in a neighborhood of a suitably non-degenerate solution, the barrier function should be close to the barrier of a tangent cone centered at the solution. Hence, the relation (1.13) should be satisfied with a reasonably high accuracy. For every $y \in \operatorname{int} Q$, we have

$$\begin{array}{rcl} p(y) & = & [\nabla^2 f(y)]^{-1} \cdot \left[\nabla^2 f(y) y + \nabla f(y) \right] \\[2ex] & = & [\nabla^2 f(y)]^{-1} \cdot \left[A \nabla^2 F_*(c - A^* y) A^* y \;-\; A \nabla F_*(c - A^* y) \right] \\[2ex] & \overset{(1.13)}{=} & [\nabla^2 f(y)]^{-1} \cdot \left[A \nabla^2 F_*(c - A^* y) A^* y \;+\; A \nabla^2 F_*(c - A^* y)(c - A^* y) \right] \\[2ex] & = & [\nabla^2 f(y)]^{-1} A \nabla^2 F_*(c - A^* y) \cdot c. \end{array}$$

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

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

Thus, we have proved the following representation.

Lemma 2.1 *For every $y \in \operatorname{int} Q$ and every optimal pair (s_*, y_*) of dual problem (2.2), we have*[1]

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

Remark 2.1 *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*

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

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 6) . Thus, the representation (2.9) can be rewritten as follows:

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

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.

To conclude this section, let us describe our prediction abilities from the points of the central path. For that, we need to compute derivatives of the trajectory (2.5). Differentiating the last line of this definition in μ, we obtain

$$0 \quad = \quad -A \nabla F_*(s_\mu) + \mu A \nabla^2 F_*(s_\mu) A^* y'_\mu.$$

Thus,

$$y'_\mu \quad = \quad \tfrac{1}{\mu}[A \nabla^2 F_*(s_\mu) A^*]^{-1} A \nabla F_*(s_\mu) \;=\; -\tfrac{1}{\mu}[\nabla^2 f(y_\mu)]^{-1} \nabla f(y_\mu). \tag{2.11}$$

Therefore, we have the following representation of the prediction point:

$$p(y_\mu) \quad = \quad y_\mu - \mu y'_\mu. \tag{2.12}$$

Hence, for the points of the central path, identity (2.9) can be written in the following form:

$$y_\mu - \mu y'_\mu - y^* \quad = \quad [\nabla^2 f(y_\mu)]^{-1} A \nabla^2 F_*(s_\mu) s_*. \tag{2.13}$$

For the primal trajectory, we have

$$\begin{aligned}
x'_\mu \quad &\overset{(2.5)}{=} \quad \tfrac{1}{\mu} x_\mu + \mu \nabla^2 F_*(s_\mu) A^* y'_\mu \overset{(2.11)}{=} \tfrac{1}{\mu} x_\mu + \nabla^2 F_*(s_\mu) A^* [A \nabla^2 F_*(s_\mu) A^*]^{-1} A \nabla F_*(s_\mu) \\
&\overset{(1.13)}{=} \quad \tfrac{1}{\mu} x_\mu - \nabla^2 F_*(s_\mu) A^* [A \nabla^2 F_*(s_\mu) A^*]^{-1} A \nabla^2 F_*(s_\mu)(c - A^* y_\mu) \\
&= \quad \tfrac{1}{\mu} x_\mu - \nabla^2 F_*(s_\mu) A^* [A \nabla^2 F_*(s_\mu) A^*]^{-1} A \nabla^2 F_*(s_\mu)(s_* + A^*(y^* - y_\mu)) \\
&= \quad \tfrac{1}{\mu} x_\mu - \nabla^2 F_*(s_\mu) \left(s_\mu - s_* + A^* [A \nabla^2 F_*(s_\mu) A^*]^{-1} A \nabla^2 F_*(s_\mu) s_* \right).
\end{aligned}$$

Using now identity (1.13), definition (2.5), and equation (2.13), we obtain the following representation:

$$x'_\mu \quad = \quad \nabla^2 F_*(s_\mu) s_* - \nabla^2 F_*(s_\mu) A^* \left(y_\mu - \mu y'_\mu - y^* \right). \tag{2.14}$$

Due to the primal-dual symmetry of our set-up, we have the following elegant geometric interpretation. Let $H \overset{\text{def}}{=} \nabla^2 F_*(s_\mu)$, then we have

$$\begin{aligned}
H^{-1/2}\left(x_* - x_\mu + \mu x'_\mu\right) \quad &= \quad \text{projection of } H^{-1/2} x_* \text{ onto the kernel of } A H^{1/2}, \\
H^{1/2} A^* \left(y_* - y_\mu + \mu y'_\mu\right) \quad &= \quad \text{projection of } H^{1/2} s_* \text{ onto the image of } \left(A H^{1/2}\right)^*.
\end{aligned}$$

In Section 4, we introduce two conditions, ensuring the boundedness of the derivative x'_μ and the high quality of the prediction from the central dual trajectory: $y_\mu - \mu y'_\mu - y^* \approx O(\mu^2)$. However, first we need to decide on how to measure distances in primal and dual spaces.

[1] In fact, in representation (2.9) we can replace the pair (s_*, y_*) by any pair $(\bar{s}, \bar{y})$, satisfying the condition $c = A^* \bar{y} + \bar{s}$. However, in our analysis we are interested only in predicting the optimal solutions.

3 Measuring the distances

Recall that 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 Euclidean norms in the primal and dual spaces. Recall the definitions of Euclidean norms based on a positive definite operator $B : \mathbb{E} \to \mathbb{E}^*$.

$$\|x\|_B \;=\; \langle Bx, x\rangle^{1/2},\ x \in \mathbb{E}, \quad \|s\|_B \;=\; \langle s, B^{-1}s\rangle^{1/2},\ s \in \mathbb{E}^*. \tag{3.1}$$

Using this operator, we can define another operator

$$G \;\stackrel{\mathrm{def}}{=}\; AB^{-1}A^* : \mathbb{H} \to \mathbb{H}^*. \tag{3.2}$$

By a Schur complement argument and the fact that A is surjective, we conclude that

$$A^*G^{-1}A \;\preceq\; B. \tag{3.3}$$

It is convenient to choose B related in a certain way to our cones and barriers. Sometimes it is useful to have B such that

$$BK \subseteq K^*. \tag{3.4}$$

Lemma 3.1 *Let B be as above, suppose B satisfies (3.4) and $u \succeq_K \pm v$. Then $\|v\|_B \le \|u\|_B$.*

Proof:

Let B, u and v be as above. Then, $\langle Bu, u\rangle - \langle Bv, v\rangle = \langle B(u-v), u+v\rangle \stackrel{(3.4)}{\ge} 0.$ □

In what follows, we choose B in a way that is related to the primal central path. Let us define

$$B \;=\; \nabla^2 F(x_\mu) \quad \text{with} \quad \mu = 1. \tag{3.5}$$

Remark 3.1 *As a side remark, we will see that, if F has negative curvature (see Section 6), then B chosen in (3.5) satisfies (3.4). In the general case, it is also possible to satisfy (3.4) by choosing*

$$B = \nabla^2 F(\bar{x}) + \nabla F(\bar{x}) \left[\nabla F(\bar{x})\right]^*,$$

where $\bar{x}$ is an arbitrary point from $\mathrm{int}\, K$. *This operator acts as*

$$Bh = \nabla^2 F(\bar{x})h + \langle \nabla F(\bar{x}), h\rangle \cdot \nabla F(\bar{x}), \quad \forall h \in \mathbb{E}.$$

Then, the property (3.4) easily follows from the Theorem on Recession Direction. Note that

$$\nabla^2 F(\bar{x}) \;\preceq\; B \;\stackrel{(1.15)}{\preceq}\; (\nu+1)\cdot \nabla^2 F(\bar{x}).$$

We need some bounds for the points of the primal and dual central paths. Denote by $X_* \subset \mathbb{E}$ the set of limit points of the primal central path, and by $S_* \subset \mathbb{E}^*$ the set of limit points of the dual central path.

Lemma 3.2 *If $\mu_1 \in (0, \mu_0]$, then*

$$\|x_{\mu_1}\|_{x_{\mu_0}} \;\le\; \nu, \quad \|s_{\mu_1}\|_{s_{\mu_0}} \;\le\; \nu. \tag{3.6}$$

In particular, for every $x_* \in X_*$ and every $s_* \in S_*$ we have:

$$\|x_*\|_B \;\le\; \nu, \quad \|s_*\|_B \;\le\; \nu. \tag{3.7}$$

Moreover, if $\mu \in (0,1]$, then

$$\begin{aligned} \tfrac{1}{4\nu^2}B \;&\preceq\; \nabla^2 F(x_\mu) \;\preceq\; \tfrac{4\nu^2}{\mu^2}B, \\ \tfrac{1}{4\nu^2}B^{-1} \;&\preceq\; \nabla^2 F_*(s_\mu) \;\preceq\; \tfrac{4\nu^2}{\mu^2}B^{-1}. \end{aligned} \tag{3.8}$$

Proof:
Indeed,

$$\|x_{\mu_1}\|^2_{x_{\mu_0}} \stackrel{(1.17)}{\leq} \langle -\nabla F(x_{\mu_0}), x_{\mu_1}\rangle^2 = \tfrac{1}{\mu_0^2}\langle s_{\mu_0}, x_{\mu_1}\rangle^2 \stackrel{(2.5)}{=} \tfrac{1}{\mu_0^2}[\langle c, x_{\mu_1}\rangle - \langle b, y_{\mu_0}\rangle]^2 \stackrel{(2.6)}{\leq} \nu^2.$$

The last inequality also uses the fact that $\langle c, x_{\mu_1}\rangle \leq \langle c, x_{\mu_0}\rangle$. Applying this inequality with $\mu_0 = 1$ and taking the limit as $\mu_1 \to 0$, we obtain (3.7) in view of the choice (3.5). The reasoning for the dual central path is the same.

Further, $\langle \nabla F(x_1), x_\mu - x_1\rangle \stackrel{(2.5)}{=} \langle c, x_1 - x_\mu\rangle \geq 0$. Therefore, applying Corollary 1.1, we get the first relation in the first line of (3.8). Similarly, we justify the first relation in the second line of (3.8). Finally,

$$\nabla^2 F(x_\mu) \stackrel{(2.5)}{=} \nabla^2 F(-\mu \nabla F_*(s_\mu)) \stackrel{(1.11)}{=} \tfrac{1}{\mu^2}\nabla^2 F(-\nabla F_*(s_\mu)) \stackrel{(1.16)}{=} \tfrac{1}{\mu^2}[\nabla^2 F_*(s_\mu)]^{-1}.$$

Using this, with the first relation, in the second line of (3.8) and the Löwner order reversing property of the inverse, we conclude the second relation in the first line of (3.8). The last unproved relation can be justified in a similar way. □

Corollary 3.1 *For every $\mu \in (0,1]$ we have*

$$\|\nabla^2 F_*(s_\mu)\|_B \leq \tfrac{4\nu^2}{\mu^2}. \tag{3.9}$$

Proof:
Indeed, for every $h \in \mathbb{E}^*$, we have

$$\begin{aligned} \|\nabla^2 F_*(s_\mu)h\|^2_B &= \langle B\nabla^2 F_*(s_\mu)h, \nabla^2 F_*(s_\mu)h\rangle \\ &\stackrel{(3.8)}{\leq} \tfrac{4\nu^2}{\mu^2}\langle h, \nabla^2 F_*(s_\mu)h\rangle \stackrel{(3.8)}{\leq} \tfrac{16\nu^4}{\mu^4}\langle h, B^{-1}h\rangle. \end{aligned}$$

□

Finally, we need to estimate the norms of the initial data.

Lemma 3.3 *We have*

$$\begin{aligned} \|A\|_{G,B} &\stackrel{\text{def}}{=} \max_{h\in\mathbb{E}}\{\|Ah\|_G : \|h\|_B = 1\} \leq 1, \\ \|A^*\|_{B,G} &\stackrel{\text{def}}{=} \max_{y\in\mathbb{H}}\{\|A^*y\|_B : \|y\|_G = 1\} \leq 1, \\ \|b\|_G &\leq \nu^{1/2}. \end{aligned} \tag{3.10}$$

Proof:
Indeed, for every $h \in \mathbb{E}$, we have

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

Further,

$$\begin{aligned} \|A^*\|_{B,G} &= \max_{h\in\mathbb{E}, y\in\mathbb{H}}\{\langle A^*y, h\rangle : \|h\|_B = 1, \|y\|_G = 1\} \\ &= \max_{h\in\mathbb{E}, y\in\mathbb{H}}\{\langle Ah, y\rangle : \|h\|_B = 1, \|y\|_G = 1\} = \|A\|_{G,B} \leq 1. \end{aligned}$$

To justify the remaining inequality, note that

$$\begin{aligned} \|b\|_G^2 &= \langle b, G^{-1}b\rangle \; = \; \max_{y\in\mathbb{H}} \left[2\langle b,y\rangle - \langle A^*y, B^{-1}A^*y\rangle\right] \\ &= \max_{y\in\mathbb{H}} \left[2\langle A^*y, x_1\rangle - \langle A^*y, B^{-1}A^*y\rangle\right] \\ &\leq \max_{s\in\mathbb{E}^*} \left[2\langle s, x_1\rangle - \langle s, B^{-1}s\rangle\right] \; = \; \langle Bx_1, x_1\rangle \\ &\overset{(3.5)}{=} \langle \nabla^2 F(x_1)x_1, x_1\rangle \overset{(1.12),(1.13)}{=} \nu. \end{aligned}$$

□

4 Main assumptions

Now, we can introduce our main assumptions.

Assumption 1 *The dual problem (2.2) has a unique optimal solution y_* and there exists a constant $\gamma_d > 0$ such that*

$$f^* - \langle b, y\rangle \; = \; \langle s, x_*\rangle \;\; \geq \;\; \gamma_d\|y - y_*\|_G \overset{(3.2)}{=} \gamma_d\|s - s_*\|_B, \tag{4.1}$$

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

Lemma 4.1 *For every $y \in \operatorname{int} Q$, we have*

$$[\nabla^2 f(y)]^{-1} \;\; \preceq \;\; \tfrac{4}{\gamma_d^2}[f^* - \langle b, y\rangle]^2 \cdot G^{-1}. \tag{4.2}$$

Proof:
Let us fix some $y \in \operatorname{int} Q$. 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

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

belongs to the set Q. Therefore, in view of inequality (4.1), we have

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

Hence,

$$\begin{aligned} \tfrac{1}{\gamma_d}[f^* - \langle b, y\rangle] &\geq \tfrac{\|[\nabla^2 f(y)]^{-1}h\|_G}{\langle h, [\nabla^2 f(y)]^{-1}h\rangle^{1/2}} - \|y - y_*\|_G \\ &\overset{(4.1)}{\geq} \tfrac{\|[\nabla^2 f(y)]^{-1}h\|_G}{\langle h, [\nabla^2 f(y)]^{-1}h\rangle^{1/2}} - \tfrac{1}{\gamma_d}[f^* - \langle b, y\rangle]. \end{aligned}$$

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

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

This means that

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

and (4.2) follows. □

Now we can estimate the size of the Hessian $\nabla^2 f(y)$ with respect to the norm induced by G:

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

Corollary 4.1 *For every $y \in \mathrm{int}\, Q$, we have*

$$\|[\nabla^2 f(y)]^{-1}\|_G \leq \tfrac{4}{\gamma_d^2}[f^* - \langle b, y\rangle]^2. \tag{4.3}$$

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, (4.3) follows directly from (4.2). Further,

$$\begin{aligned} \|v(y)\|_G^2 &= \langle G[\nabla^2 f(y)]^{-1}\nabla f(y), [\nabla^2 f(y)]^{-1}\nabla f(y)\rangle \\ &\stackrel{(4.2)}{\leq} \tfrac{4}{\gamma_d^2}[f^* - \langle b, y\rangle]^2 \langle \nabla f(y), [\nabla^2 f(y)]^{-1}\nabla f(y)\rangle. \end{aligned}$$

It remains to use inequality (1.3). □

Assumption 1 and Lemma 4.1 help us bound the norm of the right-hand side of representation (2.13). However, in this expression there is one more object, which potentially can be large. This is the vector $\nabla^2 F_*(s_\mu)s_*$. Therefore, we need one more assumption.

Assumption 2 *There exists a constant σ_d such that for every $\mu \in (0, 1]$ we have*

$$\|\nabla^2 F_*(s_\mu)s_*\|_B \leq \sigma_d. \tag{4.4}$$

Note that the representation (2.13) shows that if we want to bound the error of the first-order estimate for central points y_μ in terms of the norm of the Hessian of f at y_μ, without explicitly taking into account the possible special interaction of this Hessian with the element $A\nabla^2 F_*(s_\mu)s_*$, in view of the previous results of this section, then Assumption 2 is justifiable. It is plausible (and in fact likely) that a much more involved analysis than what we are presenting here, focusing on bounding the full expression $\left[\nabla^2 f(y_\mu)\right]^{-1} A\nabla^2 F_*(s_\mu)s_*$ without bounding the three main components of this expression separately (and somewhat independently as we do below) would lead to the same convergence results as ours under a weaker set of assumptions. However, at the time of this writing, it seems wise to leave this for future research.

In what follows, we always suppose that Assumptions 1 and 2 are valid. Let us point out their immediate consequence which also provides some justification for these assumptions.

Theorem 4.1 *For every $\mu \in (0, 1]$, we have the following bounds:*

$$\|y_\mu - \mu y'_\mu - y_*\|_G \leq \tfrac{4\sigma_d \nu^2}{\gamma_d^2}\, \mu^2, \tag{4.5}$$

$$\|x'_\mu\|_B \leq \sigma_p \stackrel{\mathrm{def}}{=} \sigma_d \left(1 + \tfrac{16\nu^4}{\gamma_d^2}\right). \tag{4.6}$$

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

$$\|y_\mu - \mu y'_\mu - y_*\|_G \leq \|\nabla^2 f(y_\mu)\|_G \cdot \|A\|_{G,B} \cdot \|\nabla^2 F_*(s_\mu)s_*\|_B.$$

Thus, in view of inequalities (4.3), (3.10), and (4.4), we have

$$\| y_\mu - \mu y'_\mu - y_* \|_G \quad \leq \quad \tfrac{4\sigma_d}{\gamma_d^2} (f^* - \langle b, y_\mu \rangle)^2.$$

Applying now identity (2.6), we obtain (4.5). Further, in view of representation (2.14), we have

$$\| x'_\mu \|_B \quad \leq \quad \| \nabla^2 F_*(s_\mu) s_* \|_B + \| \nabla^2 F_*(s_\mu) \|_B \cdot \| A^* \|_{B,G} \cdot \| y_\mu - \mu y'_\mu - y_* \|_G.$$

It remains to apply inequalities (4.4), (3.9), (3.10), and (4.5). □

Corollary 4.2 *There is a unique limit point of the primal central path:* $x_* = \lim\limits_{\mu \to 0} x_\mu$. *Moreover, for every* $\mu \in (0,1]$ *we have*

$$\| x_\mu - x_* \|_B \quad \leq \quad \sigma_p \mu. \tag{4.7}$$

Note that Assumptions 1 and 2 do not guarantee the uniqueness of the primal optimal solution in the problem (2.1).

Next, we illustrate our assumptions on some of the most popular special cases.

Example 4.1 *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*

$$\langle e, \nabla^2 F_*(s_\mu) s_* \rangle \;=\; \sum_{i \in I_*} \tfrac{s_*^{(i)}}{(s_\mu^{(i)})^2} \;=\; \sum_{i \in I_*} \tfrac{1}{s_*^{(i)}} \left(\tfrac{s_*^{(i)}}{s_\mu^{(i)}} \right)^2 \;\leq\; \| s_* \|^2_{s_\mu} \max_{i \in I_*} \tfrac{1}{s_*^{(i)}} \;\overset{(3.6)}{\leq}\; \max_{i \in I_*} \tfrac{n^2}{s_*^{(i)}}.$$

Since vector $\nabla^2 F_*(s_\mu) s_*$ *is nonnegative, we obtain for its norm an upper bound in terms of* $\max\limits_{i \in I_*} \frac{n^2}{s_*^{(i)}}$. *Note that this bound is valid even for degenerate dual solutions (too many active facets in* Q*), or multiple dual optimal solutions (which is excluded by Assumption 1).*

It is interesting, that we can find a bound for vector $\nabla^2 F_*(s_\mu) s_*$ *based on the properties of the* primal *central path. Indeed, for all* $i \in I_*$, *we have*

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

Thus, assuming $\| x_\mu - x_* \|_B \leq O(\mu)$, *we get a bound for* $\nabla^2 F_*(s_\mu) s_*$. *In view of inequality (4.7), this confirms that for Linear Programming our assumptions are very natural.*

Example 4.2 *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,

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

It seems difficult to get an upper bound for this value in terms of $\| S_* \|^2_{S_\mu} = \langle S_\mu^{-1} S_* S_\mu^{-1}, S_* \rangle$. *However, the second approach also works here:*

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

Thus, we get an upper bound for $\| \nabla^2 F_*(S_\mu) S_* \|_B$ *assuming* $\| X_\mu - X_* \|_B \in O(\mu)$. *This last condition has been used in superlinear convergence analyses in the semidefinite programming literature (see, for instance, [9]).* □

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

$$\mathcal{N}(\mu,\beta) \quad = \quad \left\{ y \in \mathbb{H}: \; \gamma(y,\mu) \stackrel{\mathrm{def}}{=} \|\nabla f(y) - \tfrac{1}{\mu} b\|_y \leq \beta \right\}, \quad \mu \in (0,1], \; \beta \in [0, \tfrac{1}{2}]. \tag{4.8}$$

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 $s = c - A^\top y$, notice that Dikin's affine scaling direction in this case is given by $\left[AS^{-2}A^\top\right]^{-1} b$. Then, our predictor step corresponds to the search direction $\left[AS^{-2}A^\top\right]^{-1} AS^{-1}e$, and our proximity measure becomes

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

Let us prove the main result of this section.

Lemma 4.2 *Let $y \in \mathcal{N}(\mu,\beta)$ with $\mu \in (0,1]$ and $\beta \in [0, \frac{1}{9}]$. Then*

$$\|\nabla^2 F_*(s(y)) s_*\|_B \quad \leq \quad \sigma_d + \tfrac{6\nu^2}{\mu}\beta, \tag{4.9}$$

$$f^* - \langle b, y \rangle \quad \leq \quad \kappa_1 \cdot \mu, \tag{4.10}$$

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

Proof:
Indeed,

$$\|s(y) - s_\mu\|_{s(y)} \quad = \quad \langle \nabla^2 F_*(s(y)) A^*(y-y_\mu), A^*(y-y_\mu)\rangle^{1/2} \; = \; \|y - y_\mu\|_y \; \stackrel{\mathrm{def}}{=} \; r.$$

Since $y \in \mathcal{N}(\mu,\beta)$ and $\beta \in [0, \frac{1}{9}]$, we have $r < 1$; therefore, by (1.6) we have

$$(1-r)^2 \nabla^2 F_*(s_\mu) \; \preceq \; \nabla^2 F_*(s(y)) \quad \preceq \quad \tfrac{1}{(1-r)^2} \nabla^2 F_*(s_\mu).$$

Denote $H = \nabla^2 F_*(s(y)) - \nabla^2 F_*(s_\mu)$. Then,

$$\pm H \quad \preceq \quad \max\left\{ \tfrac{1}{(1-r)^2} - 1, 1 - (1-r)^2 \right\} \nabla^2 F_*(s_\mu) \quad = \quad \tfrac{r(2-r)}{(1-r)^2} \nabla^2 F_*(s_\mu).$$

Note that $\|\nabla^2 F_*(s(y)) s_*\|_B \stackrel{(4.4)}{\leq} \sigma_d + \|H s_*\|_B$. At the same time,

$$\begin{aligned} \|Hs_*\|_B^2 \quad &= \quad \langle BHs_*, Hs_* \rangle \quad \stackrel{(3.8)}{\leq} \quad \tfrac{4\nu^2}{\mu^2} \langle [\nabla^2 F_*(s_\mu)]^{-1} H s_*, H s_* \rangle \\ &\stackrel{(1.1)}{\leq} \quad \tfrac{4\nu^2 r^2 (2-r)^2}{\mu^2 (1-r)^4} \langle \nabla^2 F_*(s_\mu) s_*, s_* \rangle \quad \stackrel{(3.6)}{\leq} \quad \tfrac{4\nu^4 r^2 (2-r)^2}{\mu^2 (1-r)^4}. \end{aligned}$$

Thus, $\|\nabla^2 F_*(s(y)) s_*\|_B \leq \sigma_d + \frac{2\nu^2 r(2-r)}{\mu(1-r)^2}$. For proving (4.9), it remains to note that $r \stackrel{(1.5)}{\leq} \frac{\beta}{1-\beta}$, and

$$\tfrac{r(2-r)}{(1-r)^2} \quad = \quad \tfrac{1}{(1-r)^2} - 1 \; \leq \; \tfrac{(1-\beta)^2}{(1-2\beta)^2} - 1 \; = \; \tfrac{2\beta - 3\beta^2}{(1-2\beta)^2} \; < \; \tfrac{\beta(2-3\beta)}{1-4\beta} \; \leq \; 3\beta, \quad \beta \in [0, \tfrac{1}{9}].$$

To establish (4.10), note that

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

where the last inequality follows from the assumptions of the lemma and (1.3). □

Now, we can put all our observations together.

Theorem 4.2 *Let dual problem (2.2) satisfy Assumptions 1 and 2. If for some* $\mu \in (0,1]$ *and* $\beta \in [0, \frac{1}{9}]$ *we have* $y \in \mathcal{N}(\mu, \beta)$*, then*

$$\|p(y) - y_*\|_G \leq \frac{4}{\gamma_d^2}\left(\sigma_d + \frac{6\nu^2}{\mu}\beta\right)\langle b, y - y_*\rangle^2 \leq \frac{4\nu}{\gamma_d^2}\left(\sigma_d + \frac{6\nu^2}{\mu}\beta\right) \cdot \|y - y_*\|_G^2. \tag{4.11}$$

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

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

Now, we can use inequalities (4.3), (3.10), and (4.9). For justifying the second inequality, we apply the third bound in (3.10). □

The most important consequence of the estimate (4.11) consists in the necessity to keep the neighborhood size parameter β at the same order as the (central path) penalty parameter μ. In our reasoning below, we often use

$$\beta = \tfrac{1}{9}\mu. \tag{4.12}$$

5 Efficiency of the predictor step

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

Lemma 5.1 *If* $y \in \mathcal{N}(\mu, \beta)$*, then*

$$\|p(y) - y_*\|_y \leq \kappa_2 \cdot [f^* - \langle b, y\rangle] \leq \mu \cdot \kappa, \tag{5.1}$$

where $\kappa_2 = \frac{2}{\gamma_d}\left(\sigma_d + \frac{6\nu^2}{\mu}\beta\right)$*, and* $\kappa = \kappa_1 \cdot \kappa_2$.

Proof:
Indeed,

$$\begin{aligned}\|p(y) - y_*\|_y^2 &\overset{(2.9)}{=} \langle A\nabla^2 F_*(s(y))s_*, [\nabla^2 f(y)]^{-1} A\nabla^2 F_*(s(y))s_*\rangle \\ &\overset{(4.2)}{\leq} \frac{4}{\gamma_d^2}[f^* - \langle b, y\rangle]^2 \langle A\nabla^2 F_*(s(y))s_*, G^{-1} A\nabla^2 F_*(s(y))s_*\rangle \\ &\overset{(3.3)}{\leq} \frac{4}{\gamma_d^2}[f^* - \langle b, y\rangle]^2 \langle B\nabla^2 F_*(s(y))s_*, \nabla^2 F_*(s(y))s_*\rangle.\end{aligned}$$

It remains to use the bounds (4.9) and (4.10). □

Since $\|y_* - y\|_y \geq 1$, inequality (5.1) demonstrates a significant drop in the distance to the optimal point after a full predictor step. Recall, $v(y) = [\nabla^2 f(y)]^{-1}\nabla f(y)$ and $s_p(y) = s(y) - A^* v(y)$. The following fact is also useful.

Lemma 5.2 *For every* $y \in Q$*, we have* $A \cdot \nabla^2 F_*(s(y)) \cdot s_p(y) = 0$.

Proof:
Indeed, $A\nabla^2 F_*(s(y))s_p(y) = A\nabla^2 F_*(s(y))(s(y) - A^* v(y)) \overset{(1.13)}{=} -A\nabla F_*(s(y)) - \nabla f(y) = 0$. □

Corollary 5.1 *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 5.1 *Let $y \in \mathcal{N}(\mu, \beta)$ with $\mu \in (0,1]$ and $\beta \in [0, \frac{1}{9}]$. Then, for every $r \in (0,1)$, the point $y(\hat{\alpha})$ with*

$$\hat{\alpha} \;\stackrel{\text{def}}{=}\; \frac{r}{r + \kappa_2[f^* - \langle b, y\rangle]} \tag{5.2}$$

belongs to Q. Moreover,

$$f^* - \langle b, y(\hat{\alpha})\rangle \;\leq\; \kappa_3 \cdot [f^* - \langle b, y\rangle]^2, \tag{5.3}$$

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 $W_r(y) \subseteq Q$, 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

$$\begin{aligned}\|y(\hat{\alpha}) - (1-\hat{\alpha})y - \hat{\alpha}y_*\|_y \;&=\; \hat{\alpha}\|p(y) - y_*\|_y \\ &\stackrel{(5.1)}{\leq}\; \kappa_2\hat{\alpha}[f^* - \langle b, y\rangle] \;\stackrel{(5.2)}{=}\; r(1-\hat{\alpha}).\end{aligned}$$

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

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

Since $\|b\|_G \stackrel{(3.10)}{\leq} \sqrt{\nu}$ and

$$\|p(y) - y_*\|_G \;\stackrel{(4.11)}{\leq}\; \tfrac{2\kappa_2}{\gamma_d}[f^* - \langle b, y\rangle]^2,$$

we obtain the desired inequality (5.3). □

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 large enough. In general,

$$\bar{\alpha}(y) \;\geq\; \frac{1}{\|v(y)\|_y} \;\stackrel{(1.3)}{\geq}\; \frac{1}{\nu^{1/2}}. \tag{5.4}$$

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

Theorem 5.2 *Let $y \in \mathcal{N}(\mu, \beta)$ with $\mu \in (0,1]$ and $\beta \in [0, \frac{1}{9}]$. Then*

$$1 - \bar{\alpha}(y) \;\leq\; \frac{\kappa\mu}{1+\kappa\mu}. \tag{5.5}$$

Moreover, if

$$\mu \;<\; \frac{1-2\beta}{\kappa}, \tag{5.6}$$

then

$$\bar{\alpha}(y) - 1 \;\leq\; \frac{\kappa\mu}{1-\kappa\mu-2\beta}, \tag{5.7}$$

and

$$\|y(\bar{\alpha}) - y^*\|_y \;\leq\; \kappa\mu\left(1 + \frac{\sqrt{\nu}}{1-\kappa\mu-2\beta}\right). \tag{5.8}$$

Proof:
Since for every $r \in (0,1)$

$$\bar{\alpha} \stackrel{(5.2)}{\geq} \hat{\alpha} = \tfrac{r}{r+\kappa_2[f^*-\langle b,y\rangle]},$$

we have $1-\bar{\alpha} \leq \frac{\kappa_2[f^*-\langle b,y\rangle]}{1+\kappa_2[f^*-\langle b,y\rangle]} \stackrel{(4.10)}{\leq} \frac{\kappa\mu}{1+\kappa\mu}$, which is (5.5). On the other hand,

$$\begin{aligned}\langle b, v(y)\rangle &= \langle b, [f''(y)]^{-1}(f'(y) - \tfrac{1}{\mu}b + \tfrac{1}{\mu}b\rangle \;\geq\; \tfrac{1}{\mu}\|b\|_y^2 - \beta\|b\|_y \\ &= \|b\|_y(\|f'(y) + \tfrac{1}{\mu}b - f'(y)\|_y - \beta) \;\geq\; \|b\|_y(\|f'(y)\|_y - 2\beta) \\ &= \|b\|_y(\|p(y) - y\|_y - 2\beta) \;\geq\; \|b\|_y(\|y-y_*\|_y - \|p(y) - y_*\|_y - 2\beta).\end{aligned}$$

Thus, using the estimate (5.1) and the bound $\|y-y_*\|_y \geq 1$ (since y_* is on the boundary of Q), we get

$$\langle b, v(y)\rangle \;\geq\; \|b\|_y(1-\kappa\mu-2\beta). \tag{5.9}$$

Therefore, condition (5.6) guarantees that $\langle b, v(y)\rangle > 0$. Define $\tilde{\alpha} = \frac{f^*-\langle b,y\rangle}{\langle b,v(y)\rangle}$. Then

$$\langle b, y + \tilde{\alpha}v(y)\rangle \;=\; f^*.$$

Therefore, $y + \tilde{\alpha}v(y) \notin Q$. Hence,

$$\bar{\alpha} \;\leq\; \tilde{\alpha} \;=\; 1 + \tfrac{\langle b, y_*-y-v(y)\rangle}{\langle b,v(y)\rangle} \stackrel{(5.1)}{\leq} 1 + \tfrac{\|b\|_y\cdot\kappa\mu}{\langle b,v(y)\rangle} \stackrel{(5.9)}{\leq} 1 + \tfrac{\kappa\mu}{1-\kappa\mu-2\beta}.$$

Further,

$$\begin{aligned}\|y(\bar{\alpha}) - y^*\|_y &\leq \|y(\bar{\alpha}) - p(y)\|_y + \|p(y) - y^*\|_y \stackrel{(5.1)}{\leq} |1-\bar{\alpha}|\cdot\|v(y)\|_y + \kappa\cdot\mu \\ &\stackrel{(1.3)}{\leq} |1-\bar{\alpha}|\sqrt{\nu} + \kappa\cdot\mu.\end{aligned}$$

Taking into account that in view of (5.5) and (5.7) $|1-\bar{\alpha}| \leq \frac{\kappa\mu}{1-\kappa\mu-2\beta}$, we get inequality (5.8). □

Despite the extremely good progress in function value, we have to worry about the distance to the central path. 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.

6 Barriers with negative curvature

Definition 6.1 *Let F be a normal barrier for the regular cone K. We say that F has* negative curvature *if for every $x \in \operatorname{int} K$ and $h \in K$ we have*

$$\nabla^3 F(x)[h] \;\preceq\; 0. \tag{6.1}$$

It is clear that self-scaled barriers have negative curvature (see [14]). Some other important barriers, like the negation of the logarithms of *hyperbolic polynomials* (see [4]) also share this property.

Theorem 6.1 *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 $x \in \operatorname{int} K$ and $h \in \mathbb{E}$ we have*

$$-\nabla^3 F(x)[h,h] \quad \in \quad K^*; \tag{6.2}$$

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

$$\nabla F(x+h) - \nabla F(x) \;\preceq_{K^*}\; \nabla^2 F(x) h. \tag{6.3}$$

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

$$0 \quad \geq \quad \nabla^3 F(x)[h,h,u] \;=\; \langle \nabla^3 F(x)[h,h], u\rangle. \tag{6.4}$$

Clearly, this condition is equivalent to (6.2). Therefore, statement 1. implies statement 2. Next, suppose statement 2. holds. Then, using the equation in (6.4) and the fact that statement 2. holds, we see that statement 1. holds. Now, suppose statement 2. holds. Then, for every $x \in \operatorname{int} K$ and for every $h \in \mathbb{E}$ such that $x + h \in \operatorname{int} K$, we have

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

Therefore, statement 3. holds. Finally, suppose statement 3. holds. Then, for every $x \in \operatorname{int} K$ and for every $h \in \mathbb{E}$ such that $x + h \in \operatorname{int} K$, and for every $t \in (0,1)$, we have

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

This implies, upon dividing both sides by t^2, for every $t \in (0,1)$,

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

Now, taking the limit as $t \to 0^+$, we obtain statement 1. □

Theorem 6.2 *Let the curvature of F be negative. Then for every $x \in K$, we have*

$$\nabla^2 F(x) h \quad \succeq_{K^*} \quad 0, \quad \forall h \in K, \tag{6.5}$$

and, consequently,

$$\nabla F(x+h) - \nabla F(x) \quad \succeq_{K^*} \quad 0. \tag{6.6}$$

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

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

Note that $s'(t) = \nabla^3 F(x+th)[h,h] \overset{(6.2)}{\preceq_{K^*}} 0$. This means that

$$\nabla^2 F(x) h \quad \succeq_{K^*} \quad \nabla^2 F(x+th) h \overset{(1.11)}{=} \tfrac{1}{t^2} \nabla^2 F(h + \tfrac{1}{t} x) h.$$

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

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

□

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

Lemma 6.1 *Let both F and F^* have negative curvature. Then K is a symmetric cone.*

Proof:
Indeed, for every $x \in \mathrm{int}\, K$ 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 $\nabla^2 F_*(s) \stackrel{(1.16)}{=} [\nabla^2 F(x)]^{-1}$, this means $K^* \subseteq \nabla^2 F(x)K$. Thus $K^* = \nabla^2 F(x)K$. Now, using the same arguments as in [14] it is easy to prove that for every pair $x \in \mathrm{int}\, K$ and $s \in \mathrm{int}\, K^*$ there exists a scaling point $w \in \mathrm{int}\, K$ 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. □

Recall,

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

Theorem 6.3 *Let K be a regular cone and F be a normal barrier for K, which has negative curvature. Further let $x, x+h \in \mathrm{int}\, K$. Then for every $\alpha \in [0,1)$ we have*

$$\tfrac{1}{(1+\alpha\sigma_x(h))^2}\nabla^2 F(x) \;\preceq\; \nabla^2 F(x+\alpha h) \;\preceq\; \tfrac{1}{(1-\alpha)^2}\nabla^2 F(x). \tag{6.7}$$

Proof:
Indeed,

$$\begin{aligned}\nabla^2 F(x+\alpha h) \;&=\; \nabla^2 F((1-\alpha)x + \alpha(x+h))\\ &\stackrel{(6.1)}{\preceq}\; \nabla^2 F((1-\alpha)x) \stackrel{(1.11)}{=} \tfrac{1}{(1-\alpha)^2}\nabla^2 F(x).\end{aligned}$$

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

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

Therefore, by the second inequality in (6.7), we have

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

□

7 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 $\bar{s} = s(y(\bar{\alpha})) \in K^*$. Recall that our proximity measure is:

$$\gamma(y,\mu) = \left\|\nabla f(y) - \frac{1}{\mu}b\right\|_y.$$

Therefore, first we try to estimate how $\nabla f(y(\alpha))$ varies with the step size α.

Lemma 7.1 *Let F_* have negative curvature. Then, for every $\alpha \in [0, \bar\alpha)$, we have*

$$\delta_y(\alpha) \stackrel{\rm def}{=} \|\nabla f(y(\alpha)) - \tfrac{\bar\alpha}{\bar\alpha-\alpha}\nabla f(y)\|_G \le \tfrac{\alpha\bar\alpha}{(\bar\alpha-\alpha)^2}\|\nabla^2 F_*(s(y))\bar s\|_B. \tag{7.1}$$

Proof:
Indeed,

$$\begin{aligned} \delta_y^2(\alpha) &= \langle G^{-1}(\nabla f(y(\alpha)) - \tfrac{\bar\alpha}{\bar\alpha-\alpha}\nabla f(y)), \nabla f(y(\alpha)) - \tfrac{\bar\alpha}{\bar\alpha-\alpha}\nabla f(y)\rangle \\ &= \langle G^{-1}A(\nabla F_*(s(y(\alpha))) - \tfrac{\bar\alpha}{\bar\alpha-\alpha}\nabla F_*(s(y))), A(\nabla F_*(s(y(\alpha))) - \tfrac{\bar\alpha}{\bar\alpha-\alpha}\nabla F_*(s(y)))\rangle \\ &\stackrel{(3.3)}{\le} \langle B(\nabla F_*(s(y(\alpha))) - \nabla F_*((1-\tfrac{\alpha}{\bar\alpha})s(y))), \nabla F_*(s(y(\alpha))) - \nabla F_*((1-\tfrac{\alpha}{\bar\alpha})s(y))\rangle. \end{aligned}$$

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

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

Since F_* has negative curvature, we have

$$x' \stackrel{\rm def}{=} \nabla F_*(s(y(\alpha))) - \nabla F_*\left(\left(1-\frac{\alpha}{\bar\alpha}\right)s(y)\right) \succeq_K 0.$$

Using (6.3) in Theorem 6.1, we also have

$$x'' \stackrel{\rm def}{=} \nabla^2 F_*\left(\left(1-\frac{\alpha}{\bar\alpha}\right)s(y)\right)\cdot\left(\frac{\alpha}{\bar\alpha}\bar s\right) \succeq_K x'.$$

Thus, $x'' \succeq_K \pm x'$. By (6.5) in Theorem 6.2 and Lemma 3.1, we obtain $\langle Bx'', x''\rangle \ge \langle Bx', x'\rangle$ which gives the desired conclusion. □

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 (7.1) suggests the following dependence:

$$\mu(\alpha) \approx \left(1-\tfrac{\alpha}{\bar\alpha}\right)\cdot\mu. \tag{7.2}$$

However, if $\bar\alpha$ is close to its lower limit (5.4), this strategy may be too aggressive. Indeed, in a small neighborhood of the point y we can guarantee only

$$\begin{aligned} \|\nabla f(y(\alpha)) - (1+\alpha)\nabla f(y)\|_y &= \|\nabla f(y(\alpha)) - \nabla f(y) - \alpha\nabla^2 f(y)v(y)\|_y \\ &\stackrel{(1.8)}{\le} \tfrac{\alpha^2\|v(y)\|_y^2}{1-\alpha\|v(y)\|_y}. \end{aligned} \tag{7.3}$$

In this situation, a more reasonable strategy for decreasing μ seems to be:

$$\mu(\alpha) \approx \tfrac{\mu}{1+\alpha}. \tag{7.4}$$

It appears that it is possible to combine both strategies (7.2) and (7.4) in a single expression. Denote

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

Note that

$$\xi_{\bar\alpha}(\alpha) = 1 + \alpha + \tfrac{\alpha^2}{\bar\alpha-\alpha} = \tfrac{\bar\alpha}{\bar\alpha-\alpha} - \tfrac{\alpha(1-\bar\alpha)}{\bar\alpha-\alpha}. \tag{7.5}$$

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

Theorem 7.1 *Suppose F_* has negative curvature. Let $y \in \mathcal{N}(\mu,\beta)$ with $\mu \in (0,1]$ and $\beta \in [0, \frac{1}{9}]$, satisfying condition (5.6). Then, for $y(\alpha) = y + \alpha v(y)$ with $\alpha \in (0,\bar\alpha)$, we have*

$$\begin{array}{rcl} \gamma\left(y(\alpha), \frac{\mu}{\xi_{\bar\alpha}(\alpha)}\right) & \leq & \Gamma_\mu(y,\alpha) \stackrel{\mathrm{def}}{=} \left(1 + \alpha \cdot \sigma_{s(y)}\left(-A^* v(y)\right)\right) \|\nabla f(y(\alpha)) - \frac{\xi_{\bar\alpha}(\alpha)}{\mu} \cdot b\|_y \\ & \leq & \left(1 + \alpha \cdot \sigma_{s(y)}(-A^* v(y))\right) \left[\gamma_1(\alpha) + \beta \cdot \left(1 + \frac{\alpha\bar\alpha}{\bar\alpha - \alpha}\right)\right], \\ \gamma_1(\alpha) & \stackrel{\mathrm{def}}{=} & \|\nabla f(y(\alpha)) - \xi_{\bar\alpha}(\alpha) \nabla f(y)\|_y \\ & \leq & \frac{\alpha\bar\alpha\mu}{(\bar\alpha-\alpha)^2}\left(\left(1 - \frac{\alpha}{\bar\alpha}\right)\kappa\sqrt{\nu} + \frac{2\kappa_1}{\gamma_d}\left[\sigma_d + \frac{6\nu^2\beta}{\mu} + 2\kappa\nu\left(1 + \frac{\sqrt{\nu}}{1-\kappa\mu-2\beta}\right)\frac{1-\beta}{1-2\beta}\right]\right). \end{array} \tag{7.6}$$

Proof:
Indeed,

$$\begin{array}{rcl} \gamma\left(y(\alpha), \frac{\mu}{\xi_{\bar\alpha}(\alpha)}\right) & = & \|\nabla f(y(\alpha)) - \frac{\xi_{\bar\alpha}(\alpha)}{\mu} \cdot b\|_{y(\alpha)} \\ & \stackrel{(6.7)}{\leq} & (1 + \alpha\sigma_{s(y)}(-A^* v(y))) \cdot \|\nabla f(y(\alpha)) - \frac{\xi_{\bar\alpha}(\alpha)}{\mu} \cdot b\|_y. \end{array}$$

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 $\gamma_1(\alpha)$.

$$\begin{array}{rcl} \gamma_1(\alpha) & \stackrel{(7.5)}{\leq} & \|\nabla f(y(\alpha)) - \frac{\bar\alpha}{\bar\alpha-\alpha}\nabla f(y)\|_y + \frac{\alpha(1-\bar\alpha)}{\bar\alpha-\alpha}\|\nabla f(y)\|_y \\ & \stackrel{(5.5)}{\leq} & \|\nabla f(y(\alpha)) - \frac{\bar\alpha}{\bar\alpha-\alpha}\nabla f(y)\|_y + \frac{\alpha}{\bar\alpha-\alpha} \cdot \frac{\kappa\mu\sqrt{\nu}}{1+\kappa\mu}. \end{array}$$

For the second inequality above, we also used (1.3). Note that

$$\begin{array}{rl} & \|\nabla f(y(\alpha)) - \frac{\bar\alpha}{\bar\alpha-\alpha}\nabla f(y)\|_y^2 \\ = & \langle [\nabla^2 f(y)]^{-1}(\nabla f(y(\alpha)) - \frac{\bar\alpha}{\bar\alpha-\alpha}\nabla f(y)), \nabla f(y(\alpha)) - \frac{\bar\alpha}{\bar\alpha-\alpha}\nabla f(y)\rangle \\ \stackrel{(4.2)}{\leq} & \frac{4}{\gamma_d^2}[f^* - \langle b, y\rangle]^2 \cdot \|\nabla f(y(\alpha)) - \frac{\bar\alpha}{\bar\alpha-\alpha}\nabla f(y)\|_G^2 \stackrel{(4.10)}{\leq} \frac{4\kappa_1^2\mu^2}{\gamma_d^2} \cdot \delta_y^2(\alpha). \end{array}$$

Moreover,

$$\begin{array}{rcl} \delta_y(\alpha) & \stackrel{(7.1)}{\leq} & \frac{\alpha\bar\alpha}{(\bar\alpha-\alpha)^2}\|\nabla^2 F_*(s(y))\bar s\|_B \\ & \leq & \frac{\alpha\bar\alpha}{(\bar\alpha-\alpha)^2}\left[\|\nabla^2 F_*(s(y))s_*\|_B + \|\nabla^2 F_*(s(y))(\bar s - s_*)\|_B\right] \\ & \stackrel{(4.9)}{\leq} & \frac{\alpha\bar\alpha}{(\bar\alpha-\alpha)^2}\left[\sigma_d + \frac{6\nu^2}{\mu}\beta + \|\nabla^2 F_*(s(y))(\bar s - s_*)\|_B\right]. \end{array}$$

It remains to estimate the last term.

Denote $r = \|s(y) - s_\mu\|_{s(y)} \stackrel{(1.5)}{\leq} \frac{\beta}{1-\beta}$. Since $\beta \in [0, 1/9]$, we have $r < 1$. Then, using $\mu \in (0,1]$,

$$B \stackrel{(3.8)}{\preceq} \frac{4\nu^2}{\mu^2}[\nabla^2 F_*(s_\mu)]^{-1} \stackrel{(1.6)}{\preceq} \frac{4\nu^2}{\mu^2(1-r)^2}[\nabla^2 F_*(s(y))]^{-1}.$$

Therefore,

$$\|\nabla^2 F_*(s(y))(\bar s - s_*)\|_B \leq \frac{2\nu}{\mu(1-r)}\|y(\bar\alpha) - y^*\|_y \stackrel{(5.8)}{\leq} 2\kappa\nu\left(1 + \frac{\sqrt{\nu}}{1-\kappa\mu-2\beta}\right) \cdot \frac{1-\beta}{1-2\beta}.$$

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

□

Taking into account the definition of $\xi_{\bar\alpha}(\alpha)$, we can now show that our predictor-corrector scheme with neighborhood size parameter $\beta = O(\mu)$ has local superlinear convergence.

8 Polynomial-time path-following method

Let us describe now our path-following predictor-corrector strategy. It employs the following univariate function:

$$\eta_{\bar\alpha}(\alpha) \quad = \quad \begin{cases} 2\alpha, & \alpha \in [0, \frac{1}{3}\bar\alpha), \\ \frac{\alpha+\bar\alpha}{2}, & \alpha \in [\frac{1}{3}\bar\alpha, \bar\alpha]. \end{cases} \tag{8.1}$$

This function will be used for updating the length of the current predictor step α by the rule $\alpha_+ = \eta_{\bar\alpha}(\alpha)$. If the current step is small, then it will be doubled. On the other hand, if α is close enough to the maximal step size $\bar\alpha$, then this distance for the new value α_+ will be halved.

Let us fix the tolerance parameter $\beta' = \frac{1}{6}$ for the proximity measure Γ. Consider the following method.

Dual path-following method
for barriers with negative curvature

1. Set $\mu_0 = 1$ and find point $y_0 \in \mathcal{N}\left(\mu_0, \frac{1}{25}\right)$.

2. **For** $k \geq 0$ **iterate:** a) Compute $\bar\alpha_k = \bar\alpha(y_k)$.

b) Set $\alpha_{k,0} = \frac{1}{6}\min\left\{1, \frac{1}{\|v(y_k)\|_{y_k}}\right\}$. Using recurrence $\alpha_{k,i+1} = \eta_{\bar\alpha_k}(\alpha_{k,i})$, find the maximal $i \equiv i_k$, such that $\Gamma_{\mu_k}(y_k, \alpha_{k,i}) \leq \beta'$. (8.2)

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 $y_{k+1} \in \mathcal{N}(\mu_{k+1}, \beta_{k+1})$ with $\beta_{k+1} = \frac{\mu_{k+1}}{25}$.

First, we analyze the predictor step. Recall that the bound

$$\Gamma_\mu(y, \alpha) \quad = \quad (1 + \alpha\sigma_{s(y)}(-A^* v(y)))\|\nabla f(y(\alpha)) - \tfrac{1}{\mu}\xi_{\bar\alpha}(\alpha) b\|_y$$

is explicitly computable for different values of α (we need to compute only the new vectors of the gradients $\nabla f(y(\alpha))$).

Let us show now that the predictor-corrector scheme (8.2) has polynomial-time complexity.

Lemma 8.1 *Suppose F_* has negative curvature and let $y \in \mathcal{N}(\mu, \beta)$ with $\beta \leq \frac{1}{25}$. Then for all*

$$\alpha \in \left[0, \frac{1}{6}\min\left\{1, \frac{1}{\|v(y)\|_y}\right\}\right] \tag{8.3}$$

we have $\Gamma_\mu(y, \alpha) \leq \beta'$.

Proof:

Denote $r = \|v(y)\|_y$, and $\hat{r} = \max\{1, \|v(y)\|_y\}$. For every $\alpha \in [0, \frac{1}{6\hat{r}}]$ we have

$$\Gamma_\mu(y,\alpha) \stackrel{(1.20)}{\leq} (1+\alpha r) \cdot \|\nabla f(y(\alpha)) - \tfrac{\xi_{\bar{\alpha}}(\alpha)}{\mu} \cdot b\|_y$$

$$\stackrel{(7.5)}{\leq} (1+\alpha r) \cdot \left(\|\nabla f(y(\alpha)) - (1+\alpha)\nabla f(y)\|_y + \tfrac{\alpha^2 r}{\bar{\alpha}-\alpha} + \xi_{\bar{\alpha}}(\alpha) \|\nabla f(y) - \tfrac{1}{\mu} \cdot b\|_y \right)$$

$$\stackrel{(7.3),(5.4)}{\leq} (1+\alpha r) \cdot \left(\tfrac{2\alpha^2 r^2}{1-\alpha r} + \beta \cdot \left[1 + \tfrac{\alpha}{1-\alpha r} \right] \right) \leq (1+\alpha\hat{r}) \cdot \tfrac{2\alpha^2\hat{r}^2 + \beta(1+\alpha)}{1-\alpha\hat{r}}.$$

To derive the second inequality above, we add and subtract $(1+\alpha)\nabla f(y(\alpha))$ inside the norm, then collect the necessary terms and finally use (7.5) and the definition of r.

Hence, $\Gamma_\mu(y,\alpha) \stackrel{(8.3)}{\leq} \left(1 + \frac{1}{6}\right) \cdot \frac{\frac{2}{36} + \frac{1}{25}(1+\frac{1}{6})}{1-\frac{1}{6}} < \frac{1}{6}$. □

Corollary 8.1 *Let sequence $\{\mu_k\}$ be generated by method (8.2). Then for every $k \geq 0$ we have*

$$\mu_{k+1} \leq \left(1 + \tfrac{1}{6\nu^{1/2}}\right)^{-1} \mu_k. \tag{8.4}$$

Proof:
Note that by Lemma 8.1, we get $\Gamma_{\mu_k}(y_k, \alpha_{k,0}) \leq \beta'$. Therefore, in method (8.2) we always have $\alpha_k \geq \alpha_{k,0} \stackrel{(1.3)}{\geq} \frac{1}{6\nu^{1/2}}$. It remains to note that $\xi_{\bar{\alpha}}(\alpha_k) \stackrel{(7.5)}{\geq} 1 + \alpha_k$. □

Secondly, we rigorously define and analyze the corrector step of the method (8.2). After Step c) of the method, we have $y_k \in \mathcal{N}(\mu_{k+1}, \beta')$, by Lemma 8.1. Our goal is to quickly find $y_{k+1} \in \mathcal{N}(\mu_{k+1}, \beta)$. So, we minimize the function

$$g(y) := f(y) - \frac{1}{\mu_{k+1}} b^\top y,$$

utilizing the Newton method. Indeed, the Newton direction to minimize g is the unique solution d of the linear system

$$\nabla^2 f(y_k) d = -\left[\nabla f(y_k) - \frac{1}{\mu_{k+1}} b \right].$$

Moreover, g is a self-concordant barrier and the Newton decrement for g at y_k is precisely

$$\|\nabla g(y_k)\|_{y_k} = \gamma(y_k, \mu_{k+1}) \leq \beta'.$$

By the proof of Theorem 2.2.2 in [13] (page 20), defining

$$y_{k+1} = y_k + d,$$

we have

$$\gamma(y_{k+1}, \mu_{k+1}) \leq \frac{[\gamma(y_k, \mu_{k+1})]^2}{[1 - \gamma(y_k, \mu_{k+1})]^2} \leq \frac{(\beta')^2}{(1-\beta')^2},$$

achieving quadratic convergence, as desired. Thus, we have proved the following lemma.

Lemma 8.2 *Let $\beta' \leq 1/6$ and for every $y \in \mathcal{N}(\mu, \beta')$, define d as above (the Newton direction at y for minimizing g). Then, $(y+d) \in \mathcal{N}(\mu, \beta)$, where $\beta \leq \frac{(\beta')^2}{(1-\beta')^2}$.*

At the same time, it follows from Theorem 7.1 that method (8.2) can be accelerated. Taking into account the choice $\beta_k = \frac{\mu_k}{25}$ in (8.2), we see that

$$\kappa_1 \leq \nu + \tfrac{\frac{1}{25}(\frac{1}{25} + \nu^{1/2})}{1-\frac{1}{25}},$$

$$\kappa_2 \leq \tfrac{2}{\gamma_d}(\sigma_d + \tfrac{6}{25}\nu^2),$$

and $\kappa = \kappa_1\kappa_2$. Let us assume that μ_k is small enough. Namely, we assume that

$$\frac{1}{1-\kappa\mu_k-2\beta_k} \leq 2 \quad \Leftrightarrow \quad \mu_k \leq \frac{1}{2(\kappa+\frac{2}{25})}. \tag{8.5}$$

Then

$$\begin{aligned} \bar{\alpha}_k &\overset{(5.7)}{\leq} 1 + \frac{\kappa\mu_k}{1-\kappa\mu_k-\frac{2}{25}\mu_k} \leq 1+2\kappa\mu_k \overset{(8.5)}{\leq} 2, \\ \bar{\alpha}_k &\overset{(5.5)}{\geq} \frac{1}{1+\kappa\mu_k} \overset{(8.5)}{\geq} \frac{2}{3}. \end{aligned} \tag{8.6}$$

Therefore, taking into account that

$$\sigma_{s(y)}(-A^*v(y)) \overset{(1.20)}{\leq} \|A^*v(y)\|_{s(y)} = \|v(y)\|_y \overset{(1.3)}{\leq} \nu^{1/2},$$

we have the following bound on the growth of the proximity measure:

$$\begin{aligned} \Gamma_{\mu_k}(y_k,\alpha) &\overset{(7.6)}{\leq} (1+\bar{\alpha}_k\nu^{1/2})\left[\frac{\alpha\bar{\alpha}_k}{(\bar{\alpha}_k-\alpha)^2}\mu_k c_0 + \frac{\mu_k}{25}\left(1+\frac{\alpha\bar{\alpha}_k}{\bar{\alpha}_k-\alpha}\right)\right] \\ &\overset{(8.6)}{\leq} (1+2\nu^{1/2})\mu_k\left[\frac{\bar{\alpha}_k^2}{(\bar{\alpha}_k-\alpha)^2}c_0 + \frac{1}{25}\left(1+\frac{2\bar{\alpha}_k}{\bar{\alpha}_k-\alpha}\right)\right], \end{aligned} \tag{8.7}$$

where $c_0 \overset{(8.5)}{=} \kappa\nu^{1/2} + \frac{2\kappa_1}{\gamma_d}\left[\sigma_d + \frac{6}{25}\nu^2 + 2\kappa\nu(1+2\nu^{1/2})\frac{1-\frac{1}{25}}{1-\frac{2}{25}}\right]$.

To establish the asymptotic superlinear convergence of our method, we next assume μ_k is even smaller, namely, that

$$\mu_k \leq \frac{\beta'}{(1+2\nu^{1/2})(9c_0+\frac{7}{25})}. \tag{8.8}$$

Theorem 8.1 *Suppose that μ_k satisfies condition (8.8). Then method (8.2) has a local superlinear rate of convergence:*

$$\mu_{k+1} \leq \frac{9}{c_1^{1/2}}\mu_k^{3/2}.$$

Proof:
Assume μ_k satisfies (8.8). Denote by $\xi(\mu)$ the unique positive solution of the equation

$$c_0\xi^2 + \frac{1}{25}(1+2\xi) = \frac{\beta'}{(1+2\nu^{1/2})\mu}. \tag{8.9}$$

In view of assumption (8.8), we have $1 \leq \frac{1}{3}\xi(\mu_k)$. Therefore,

$$\frac{\beta'}{(1+2\nu^{1/2})\mu_k} \leq (c_0 + \frac{7}{9\cdot 25})\xi^2(\mu_k). \tag{8.10}$$

Note that for $\alpha(\mu_k)$ defined by the equation

$$\frac{\bar{\alpha}_k}{\bar{\alpha}_k-\alpha(\mu_k)} = \xi(\mu_k) \overset{(8.8)}{\geq} 3,$$

we have $\Gamma_{\mu_k}(y_k,\alpha(\mu_k)) \leq \beta'$. Therefore, $y_k + \alpha(\mu_k)v(y_k) \in Q$. At the same time,

$$\alpha(\mu_k) \geq \frac{2}{3}\bar{\alpha}_k. \tag{8.11}$$

Note that in view of the termination criterion of Step b) in method (8.2) we either have $2\alpha_k \geq \alpha(\mu_k)$, or $\frac{1}{2}(\bar{\alpha}_k+\alpha_k) \geq \alpha(\mu_k)$. The first case is possible only if $\alpha_k < \frac{1}{3}\bar{\alpha}_k$. This implies $\alpha(\mu_k) < \frac{2}{3}\bar{\alpha}_k$, and this contradicts the lower bound (8.11). Thus, we have

$$\alpha_k \geq 2\alpha(\mu_k) - \bar{\alpha}_k.$$

Therefore,

$$\begin{aligned} \xi_{\bar{\alpha}_k}(\alpha_k) &= 1 + \frac{\alpha_k \bar{\alpha}_k}{\bar{\alpha}_k - \alpha_k} \geq 1 + \frac{\bar{\alpha}_k(2\alpha(\mu_k) - \bar{\alpha}_k)}{2(\bar{\alpha}_k - \alpha(\mu_k))} \stackrel{(8.11)}{\geq} 1 + \frac{\bar{\alpha}_k^2}{6(\bar{\alpha}_k - \alpha(\mu_k))} \\ &= 1 + \tfrac{1}{6}\bar{\alpha}_k \xi(\mu_k) \stackrel{(8.6)}{\geq} 1 + \tfrac{1}{9}\xi(\mu_k) \stackrel{(8.10)}{\geq} 1 + \tfrac{1}{9}\sqrt{\tfrac{c_1}{\mu_k}}, \end{aligned}$$

where $c_1 = \frac{\beta'}{(1+2\nu^{1/2})(c_0 + \frac{7}{9 \cdot 25})}$. Whence, $\mu_{k+1} \leq \frac{9}{c_1^{1/2}} \mu_k^{3/2}$, as desired. □

It remains to estimate the full complexity of one iteration of method (8.2). It has two auxiliary search procedures. The first one (that is Step b)) consists in finding an appropriate value of the predictor step. We need an auxiliary statement on performance of its recursive rule $\alpha_+ = \eta_{\bar{\alpha}}(\alpha)$.

Lemma 8.3 *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, $\xi_{\bar{\alpha}}(\alpha_i) \geq 1 + (\xi_{\bar{\alpha}}(\alpha_0) - 1) \cdot 2^i \stackrel{(7.5)}{\geq} 1 + \alpha_0 \cdot 2^i$. □

Note that the number of evaluations of the proximity measure $\Gamma_{\mu_k}(y_k, \cdot)$ at Step b) of method (8.2) is equal to $i_k + 2$. Therefore, for the first N iterations of this method we have

$$\sum_{k=0}^{N-1} (i_k + 2) \leq 2N + \sum_{k=0}^{N-1} \log_2 \frac{\mu_k}{\mu_{k+1}\alpha_{k,0}} \leq N(2 + \tfrac{1}{2}\log_2 \nu) - \log_2 \mu_N.$$

Taking into account that for solving the problem (2.2) with absolute accuracy ϵ we need to ensure $\mu_N \leq \frac{\nu}{\epsilon}$, and using the rate of convergence (8.4), we conclude that the total number of evaluations of the proximity measure $\Gamma_{\mu_k}(y_k, \cdot)$ does not exceed $O(\nu^{1/2} \ln \nu \ln \frac{\nu}{\epsilon})$.

It remains to estimate the complexity of the correction process (this is Step d)). This process cannot be too long either. Note that the penalty parameters μ_k are bounded from below by $\frac{\epsilon}{\nu}$, 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 by Lemma 8.2. Therefore, the number of iterations at Step d) is bounded by $O(\ln \ln \frac{\kappa_1}{\epsilon})$. In Section 9, we will demonstrate on simple examples that the high accuracy in approximating the trajectory of central path is crucial for local superlinear convergence of the proposed algorithm.

There exists another possibility for organizing the correction process at Step d). We can apply a gradient method in the metric defined by the Hessian of the objective function at point p_k. Then the rate of convergence will be linear, and we might not be able to prove an upper bound better than $\Omega(\ln \frac{\nu}{\epsilon})$ correction iterations at each Step d) of method (8.2). However, each such corrector direction will be cheap to compute since this type of corrector steps do not require us to reevaluate the Hessian. (We would evaluate the Hessian at p_k only, then compute a suitable decomposition of it, e.g., Cholesky, and then only perform backsolves with this decomposition until a point in the desired neighborhood of the central path is obtained.)

9 Discussion

9.1 2D-examples

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

$$\max_{y\in\mathbb{R}^2}\{\langle b,y\rangle :\ y_2\geq 0,\ y_1\geq y_2^2\}. \tag{9.1}$$

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

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

We will take $b=[-1,0]^\top$ and check our conditions for the optimal point $y_*=0$.

Problem (9.1) can be seen as a restriction of the following conic problem:

$$\max_{s,y}\{\langle b,y\rangle :\quad s_1=y_1,\ s_2=y_2,\ s_3=1,\ s_4=y_2,\quad s_1s_3\geq s_2^2,\quad s_4\geq 0\}, \tag{9.2}$$

endowed with the barrier $F_*(s)=-\ln(s_1s_3-s_2^2)-\ln s_4$. Note that

$$\nabla F_*(s)=\left(\frac{-s_3}{s_1s_3-s_2^2},\frac{2s_2}{s_1s_3-s_2^2},\frac{-s_1}{s_1s_3-s_2^2},\frac{-1}{s_4}\right)^\top.$$

Denote $\omega=s_1s_3-s_2^2$. Since in problem (9.2) $y_*=0$ corresponds to $s_*=e_3$, we have the following representation:

$$\nabla^2F_*(s)s_*=\nabla^2F_*(s)e_3=\frac{1}{\omega^2}\cdot\left(s_2^2,-2s_1s_2,s_1^2,0\right)^\top.$$

Let us choose in the primal space the norm

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

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

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

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

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

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

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

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 $\mu\to 0$. It is interesting that in our example, Assumption 2 is valid if and only if the problem (9.1) with $y^*=0$ satisfies Assumption 1.

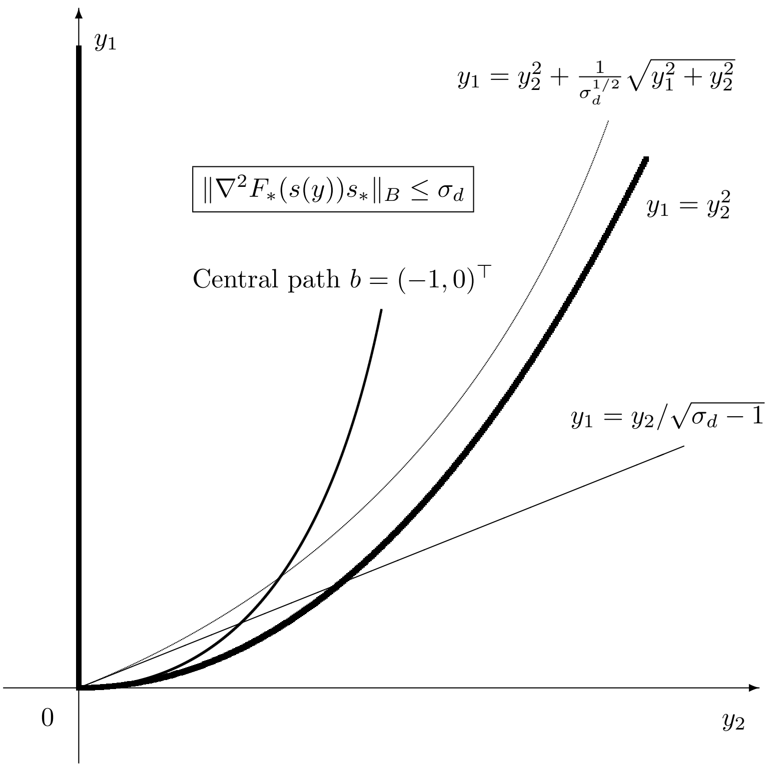

Figure 1. Behavior of $\|\nabla^2 F_*(s(y))s_*\|_B$.

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

$$\begin{aligned} \mathcal{M}(\beta) &= \mathrm{Cl}\left(\bigcup_{\mu\in\mathbb{R}} \mathcal{N}(\mu,\beta)\right) \\ &= \left\{ y : \theta^2(y) \overset{\text{def}}{=} \|\nabla f(y)\|_y^2 - \tfrac{1}{\|b\|_y^2}\langle \nabla f(y), [\nabla^2 f(y)]^{-1} b\rangle^2 \le \beta^2 \right\}. \end{aligned} \tag{9.3}$$

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

Consider the following problem:

$$\max_{y\in\mathbb{R}^2}\{y_1 : \|y\| \le 1\}. \tag{9.4}$$

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

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

Therefore,

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

and for $b = [1,0]^\top$ we have

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

Thus,

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

We conclude that for problem (9.4) the maximal neighborhood of the central path has the following representation:

$$\mathcal{M}(\beta) \quad = \quad \left\{ y \in \mathbb{R}^2 : \; y_1^2 + \tfrac{2-\beta^2}{\beta^2} \cdot y_2^2 \; \leq \; 1 \right\} \tag{9.5}$$

(see Figure 2).

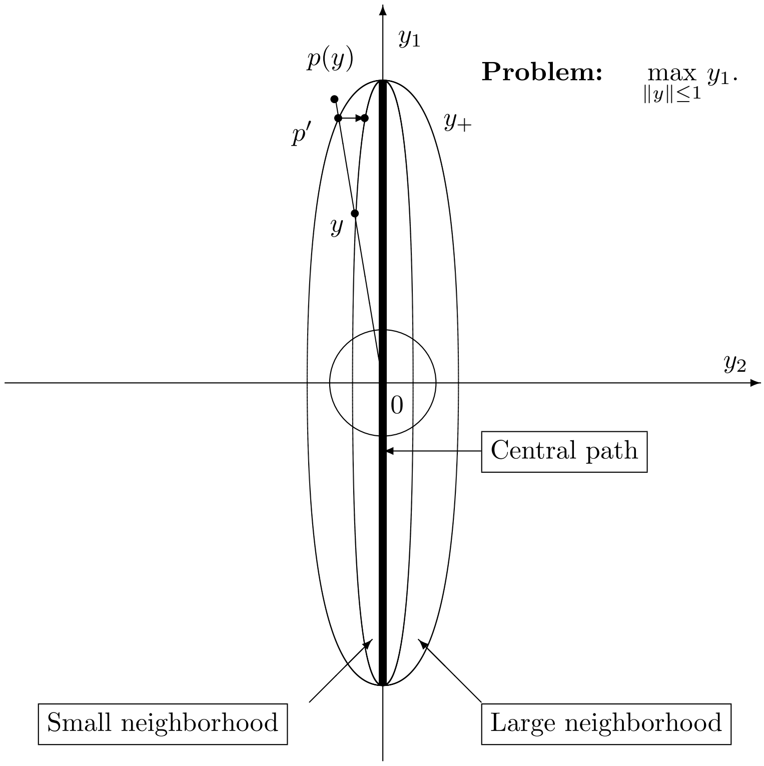

Figure 2. Prediction in the absence of sharp maximum.

Note that $p(y) = \frac{2y}{1+\|y\|^2} \in \operatorname{int} Q$. 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 $y \to y_+$ shown in 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 $\mu \to 0$.

9.2 Examples of cones with negative curvature

In accordance with the definition (6.1), negative curvature of barrier functions is preserved by the following operations.

- If barriers F_i for cones $K_i \subset \mathbb{E}_i$, $i \in \{1,2\}$, 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 $K_i \subset \mathbb{E}$, $i \in \{1,2\}$, have negative curvature, then the curvature of the barrier $F_1 + F_2$ for the cone $K_1 \bigcap K_2$ 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 [14]).
- 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 9.1 *Consider the cone of nonnegative polynomials:*

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

The dual to this cone is the cone of positive semidefinite Hankel matrices. For $k \in \{0, 1, \ldots, 2n\}$, denote

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

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

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

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

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

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 [12]):

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

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 superlinearly convergent path-following scheme (8.2).

To conclude the paper, let us mention that the negative curvature seems to be a natural property of some self-concordant barriers. Indeed, let us move from some point $x \in \operatorname{int} K$ 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:

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

We should expect that, in this situation, for a good self-concordant barrier, 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: $\nabla^2 F(x) \succeq \nabla^2 F(u)$. Thus, we are led to the following unsolved problem: *What is the class of regular convex cones that admit a logarithmically homogeneous self-concordant barrier with negative curvature?*

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 and P. Van Dooren, Optimization problems over positive pseudo-polynomial matrices, *SIAM J. Matrix Anal. Appl.* 25 (2003) 57–79.

[3] A. V. Fiacco and 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. Lu and R. D. C. Monteiro, Limiting behavior of the Alizadeh-Haeberly-Overton weighted paths in semidefinite programming, *Optim. Methods Softw.* 22 (2007) 849–870.

[8] Z. Lu and R. D. C. Monteiro, Error bounds and limiting behavior of weighted paths associated with the SDP map $X^{1/2}SX^{1/2}$, *SIAM J. Optim.* 15 (2004/05) 348–374.

[9] Z.-Q. Luo, J. F. Sturm and S. Zhang, Superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming, *SIAM J. Optim.* 8 (1998) 59–81.

[10] S. Mehrotra, Quadratic convergence in a primal-dual method, *Math. Oper. Res.* 18 (1993) 741–751.

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

[12] 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.

[13] Yu. Nesterov and A. Nemirovskii, *Interior Point Polynomial Methods in Convex Programming: Theory and Applications,* SIAM, Philadelphia, 1994.

[14] Yu. Nesterov and M. J. Todd, Self-scaled barriers and interior-point methods for convex programming, *Math. Oper. Res.* 22 (1997) 1–42.

[15] F. A. Potra and R. Sheng, Superlinear convergence of a predictor-corrector method for semidefinite programming without shrinking central path neighborhood, *Bull. Math. Soc. Sci. Math. Roumanie* 43 (2000) 107–124.

[16] F. A. Potra and R. Sheng, Superlinear convergence of interior-point algorithms for semidefinite programming, *J. Optim. Theory Appl.* 99 (1998) 103–119.

[17] J. Renegar, Hyperbolic programs, and their derivative relaxations, *Found. Comput. Math.* 6 (2006) 59–79.

[18] 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.

[19] Y. Zhang, R. A. Tapia and J. E. Dennis, On the superlinear and quadratic convergence of primal-dual interior point linear programming algorithms, *SIAM J. Optim.* 2 (1992) 304–324.