Central path and Riemannian distances

Yu. Nesterov* and A. Nemirovski†

June, 2003

Abstract

In this paper we study the Riemannian length of the primal central path computed with respect to the local metric defined by a self-concordant function. We show that despite to some examples, in many important situations the length of this path is quite close to the length of geodesic curves. We show that in the case when the Riemannian structure of a bounded convex set is introduced by a ν-self-concordant barrier, the central path is sub-geodesic up to the factor $\nu^{1/4}$.

Keywords:
Riemannan geometry, convex optimization, structural optimization, interior-point methods, path-following methods, self-concordant functions, polynomial-time methods.

---

*CORE, Catholic University of Louvain, 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium; e-mail: nesterov@core.ucl.ac.be. 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.

†Technion-Israel institute of Technology, Haifa, Israel; e-mail: nemirovs@ie.technion.ac.il

1 Introduction

Motivation. The majority of the modern interior-point schemes are based on the idea of tracing some paths in the interior of convex sets. A generic case of this type is as follows. We are given a self-concordant function $f(x)$ with convex open domain $\operatorname{dom} f$ in a finite-dimensional space E (see Section 2 to recall the definitions). And let $\mathcal{A}$ be a linear subspace in E. We are interested in tracing the path

$$x(t) = \operatorname*{argmin}_{x \in \mathcal{A} + ta + b} f_t(x), \quad f_t(x) = f(x) + \langle tc + d, x \rangle. \tag{1.1}$$

Note that this framework covers seemingly all standard short-step interior-point path-following schemes (e.g. [5], [7] and [10]):

- The *primal path-following method* solving the optimization program

  $$\min_x \{ \langle c, x \rangle : \ x \in \operatorname{cl} \operatorname{dom} f \}$$

  traces, as $t \to \infty$, the *primal central path* (1.1) given by

  $$\mathcal{A} = E, \quad a = 0, \quad b = 0, \quad d = 0$$

  and a ν-self-concordant barrier f for $\operatorname{cl} \operatorname{dom} f$.

- Let $\mathcal{K}$ be a closed pointed cone with nonempty interior in a finite-dimensional linear space G, $\mathcal{L}$ be a linear subspace in G, G^* be a linear space dual to G, and $\mathcal{K}_* \subset G^*$ be a cone dual to $\mathcal{K}$. And let F be a ν-logarithmically homogeneous self-concordant barrier F for $\mathcal{K}$ with Legendre transform F_*. The *feasible start primal-dual path-following method* for solving the primal-dual pair of conic problems

  $$(P): \quad \min_u \{ \langle p, u \rangle : u \in [\mathcal{L} - q] \cap \mathcal{K} \}$$

  $$(D): \quad \max_v \left\{ \langle v, q \rangle : v \in [\mathcal{L}^\perp + p] \cap \mathcal{K}_* \right\},$$

  traces, as $t \to \infty$, the *primal-dual central path* (1.1) given by

  $$E = G \times G^*, \quad \mathcal{A} = \mathcal{L} \times \mathcal{L}^\perp, \quad f(u, v) = F(u) + F_*(v),$$

  $$a = 0, \quad b = (-q, p), \quad c = (p, q), \quad d = 0.$$

- The *infeasible-start primal-dual path-following method* for solving (P), (D) traces, as $t \to \infty$, the path (1.1) given by

  $$E = G \times G^*, \quad \mathcal{A} = \mathcal{L} \times \mathcal{L}^\perp, \quad f(u, v) = F(u) + F_*(v),$$

  $$a = (-q, p), \quad c = 0, \quad d = (p, -q)$$

  and b defined by a starting point from $\operatorname{int} \mathcal{K} \times \operatorname{int} \mathcal{K}_*$.

It can be shown that in all situation an *upper* complexity bound for tracing the corresponding trajectories by a short-step strategy is of the order $O(\sqrt{\nu}\ln\frac{\nu}{\epsilon})$ iterations, where ϵ is a relative accuracy of the final approximate solution of the problem.

Recently in [6] there was developed a technique for establishing the *lower* complexity bounds for short-step path-following schemes. This technique is based on the concepts of *Riemannian geometry* (see [1, 2, 3, 8, 9] for main definitions and their applications in optimization).

Let f be a non-degenerate self-concordant function on an open convex domain $Q \subseteq E$. We can use the Hessian of f in order to define a Riemannian structure on Q. Namely, for a continuously differentiable curve $\gamma(t) \in Q$, $t_0 \le t \le t_1$, we define the length of this curve as follows:

$$\rho[\gamma(\cdot), t_0, t_1] = \int\limits_{t_0}^{t_1} \langle f''(\gamma(t))\gamma'(t), \gamma'(t)\rangle^{1/2} dt.$$

Definition 1.1 *The infimum of the quantities $\rho[\gamma(\cdot), t_0, t_1]$ over all continuously differentiable curves $\gamma(\cdot)$ in Q linking a given pair of points x and y of the set (i.e., $\gamma(t_0) = x, \gamma(t_1) = y$) is called the* Riemannian distance *between x and y. We denote it by $\sigma(x, y)$.*

Clearly, $\sigma(x, y)$ is a distance on Q. For $\widehat{Q} \subset Q$ we define $\sigma(x, \widehat{Q}) = \inf\limits_y \{\sigma(x, y) : \ y \in \widehat{Q}\}$.

In [6], it was shown that, the value $\mathrm{O}(\sigma(x, y))$ provides us with a *lower bound* for the length of any short-step sequence, connecting the points x and y (see [6], Section 3). Moreover, in [6], Section 5, it was shown that the primal-dual feasible and infeasible path-following schemes are optimal (up to a constant factor). Thus, in the above list of interior-point methods only the primal central path trajectories still need to be checked for optimality.

Actually, it is easy to see that in the pure primal setting the central path trajectories can be very bad. Let us look at the following example.

Example 1.1 *Let $Q \equiv R^n_+$ be a positive orthant in R^n. Let us endow Q with a standard self-concordant barrier*

$$f(x) = -\sum_{i=1}^n \ln x^{(i)}, \quad \nu = n.$$

Then, the Riemannian distance in Q is defined as follows (see (6.20), [6]):

$$\sigma(x, y) = \left[\sum\limits_{i=1}^n \ln^2 \frac{x^{(i)}}{y^{(i)}} \right]^{1/2}.$$

Let us form a central path, which connects the point $x_0 = e = (1, \dots, 1)^T \in R^n$ with the simplex

$$\Delta(A) = \left\{ x \in Q : \ \sum_{i=1}^n x^{(i)} = n \cdot (1 + A) \right\}, \quad A > 0.$$

That is a solution of the following problem

$$x(t) = \arg \min_{x \in \Delta(t)} f(x), \quad 0 \le t \le A.$$

Clearly, $x(t) = (1+t) \cdot e$. *Thus, using the central path we can travel from* $x_0 = e$ *to the set* $\Delta(A)$ *in* $O(\sigma(e, x(A)))$ *iterations of a path-following scheme with*

$$\sigma(e, x(A)) = \sqrt{n} \ln(1+A).$$

However, it is easy to see that there exists a shortcut:

$$y = e + nAe_1 \in \Delta(A), \quad \sigma(e, y) = \ln(1+nA),$$

where $e_1 = (1, 0, \ldots, 0)^T \in R^n$. $\square$

Fortunately, the situation is not always so bad. The main goal of this paper is to show that in many important situations the primal central paths are sub-optimal up to a multiplicative factor $\nu^{1/4}$.

Contents. In Section 2 we recall the reader the main fact on the theory of self-concordant functions and prove several new inequalities, which are necessary for our analysis. In Section 3 we establish several lower bounds on the Riemannian distances in convex sets in terms of different local norms defined by a self-concordant function. In Section 4 we prove the main result of this paper. That is an upper bound on the Riemannian length of a segment of a central path in terms of variation of the value of corresponding self-concordant function and the logarithm of the variation of the path parameter. The last Section 5 we apply this result to different problem instances: finding a minimum of a self-concordant function (Section 5.1), feasibility problem (Section 5.2) and the standard minimization problem (Section 5.3). We show that in Example 1.1 the presence of the unpleasant factor $O(\sqrt{\nu}/\ln\nu)$ in the ratio of the length of the central path and the corresponding Riemannian distance is due to unboundedness of the basic feasible set Q. If Q is bounded, this ratio can be at most of the order $O(\nu^{1/4})$.

2 Self-concordant functions and barriers

In order to make the paper self-contained, in this section we summarize the main results on self-concordant functions, which can be found in [5], Chapter 2, and in [4], Chapter 4. We prove also some new inequalities, which are useful for working with the Riemannian distances.

Let E be a finite-dimensional real vector space and $Q \subset E$ be an open convex domain. A three-times continuously differentiable convex function

$$f(x) : \quad Q \to R$$

is called *self-concordant*, if the sets $\{x \in Q : f(x) \leq a\}$ are closed for every $a \in R$ and

$$\tfrac{d^3}{dt^3} f(x+th)|_{t=0} \leq 2 \left(\tfrac{d^2}{dt^2} f(x+th)|_{t=0} \right)^{3/2} \quad \forall x \in Q,\, h \in E.$$

Such a function is called *non-degenerate*, if its Hessian is positive definite at some (and then – at every) point of Q. This happens, for example, if Q contains no straight line.

Denote by E^* the space *dual* to E. For $h \in E$ and $\eta \in E^*$ denote by $\langle \eta, h \rangle$ the value of the linear function η on the vector h. Let f be a non-degenerate self-concordant function on Q. For every $x \in Q$ we have $f'(x) \in E^*$. Thus, the Hessian $f''(x)$ defines a non-degenerate linear operator:

$$h \mapsto f''(x)h \in E^* \quad \forall h \in E.$$

Hence, we can define a local *primal* norm:

$$\|h\|_x \quad = \quad \langle f''(x)h, h \rangle^{1/2} : \quad E \to R,$$

and, using the standard definition $\|\eta\|_x^* = \max\limits_{h:\|h\|_x \le 1} \langle \eta, h \rangle$, the corresponding *local dual* norm:

$$\|\eta\|_x^* \quad = \quad \langle \eta, [f''(x)]^{-1}\eta \rangle^{1/2} : \quad E^* \to R.$$

Denote by

$$\lambda(x) = \langle f'(x), [f''(x)]^{-1} f'(x) \rangle^{1/2} = \|f'(x)\|_x^*$$

the local norm of the gradient $f'(x)$. A non-degenerate self-concordant function f is called a *self-concordant barrier* with parameter ν, if

$$\lambda^2(x) \le \nu \quad \forall x \in Q.$$

Note that ν cannot be smaller than one.

Let us mention first the well-known facts.

Proposition 2.1 *Let f be a non-degenerate self-concordant function on an open convex domain $Q \subset E$. Then*

(i) *For every $x \in Q$, the ellipsoid $W(x) \equiv \{y : \|y - x\|_x < 1\}$ is contained in Q. For any $y \in W(x)$ and any $h \in E$ we have*

$$(1 - \|y - x\|_x)\|h\|_x \le \|h\|_y \le \tfrac{\|h\|_x}{1 - \|y-x\|_x}. \tag{2.1}$$

Moreover, for any x and y from Q

$$\|y - x\|_y \ge \tfrac{\|y-x\|_x}{1+\|y-x\|_x}, \tag{2.2}$$

and

$$\langle f'(x) - f'(y), x - y \rangle \ge \tfrac{\|x-y\|_x^2}{1+\|x-y\|_x}. \tag{2.3}$$

(ii) *The following facts are related to existence of a minimizer x_f of $f(x)$ on Q.*

- *f attains its minimum on Q if and only if it is below bounded.*
- *f attains its minimum on Q if and only if the set $\{x : \lambda(x) < 1\}$ is nonempty.*
- *If $\lambda(x) < 1$, then $f(x) - f(x_f) \le -\lambda(x) - \ln(1 - \lambda(x))$.*
- *If x_f exists, then it is unique.*
- *For every $\rho < 1$, the set $\{x \in Q : \lambda(x) \le \rho\}$ is compact.*

- *Denote* $r_f(x) = \|x - x_f\|_{x_f}$. *If* $\lambda(x) < 1$, *then*

$$r_f(x) - \ln(1 + r_f(x)) \le -\lambda(x) - \ln(1 - \lambda(x)).$$

Hence, the set $\{x : \lambda(x) \le \frac{1}{2}\}$ *is contained in the ellipsoid* $\{x : \|x - x_f\|_{x_f} \le \frac{3}{4}\}$.

(iii) *For every* $x \in Q$, *the damped Newton iterate*

$$x_+ = x - \tfrac{1}{1+\lambda(x)}[f''(x)]^{-1} f'(x)$$

belongs to Q, *and*

$$\begin{aligned} f(x_+) &\le f(x) - [\lambda(x) - \ln(1 + \lambda(x))], \\ \lambda(x_+) &\le 2\lambda^2(x). \end{aligned} \tag{2.4}$$

(iv) *The domain of the Legendre transformation*

$$f_*(\xi) = \sup_x [\langle \xi, x\rangle - f(x)] : \quad E^* \to R \cup \{+\infty\}$$

of f *is an open convex set which is exactly the image of* Q *under the one-to-one* C^2 *mapping*

$$f'(x) : \quad Q \to E^*.$$

The function f_* *is a non-degenerate self-concordant function on its domain, and the Legendre transformation of* f_* *is* f.

(v) *Let* f *be* ν*-self-concordant barrier for* $\mathrm{cl}\, Q$. *Let* x *and* y *belong to* Q. *Then*

$$\langle f'(x), y - x\rangle \le \nu.$$

If in addition, $\langle f'(x), y - x\rangle \ge 0$, *then* $\|y - x\|_x \le \nu + 2\sqrt{\nu}$.

Let us prove now some new inequalities. Let f be a nondegenerate self-concordant function with $\mathrm{dom}\, f \subseteq E$. Denote

$$\begin{aligned} \zeta(t) &= \ln(1 + t) - \tfrac{t}{1+t}, \quad t > -1, \\ \zeta_*(t) &= \ln(1 - t) + \tfrac{t}{1-t}, \quad t < 1. \end{aligned}$$

In what follows we assume that these functions are equal $+\infty$ outside their natural domains.

Lemma 2.1 *For every* $x, y \in \mathrm{dom}\, f$ *we have:*

$$f(y) \ge f(x) + \langle f'(x), y - x\rangle + \zeta(\|y - x\|_y), \tag{2.5}$$

$$f(y) \le f(x) + \langle f'(x), y - x\rangle + \zeta_*(\|y - x\|_y), \tag{2.6}$$

$$f(x) \ge f(y) + \langle f'(y), x - y\rangle + \zeta(\|f'(x) - f'(y)\|^*_y), \tag{2.7}$$

$$f(x) \le f(y) + \langle f'(y), x - y\rangle + \zeta_*(\|f'(x) - f'(y)\|^*_y). \tag{2.8}$$

Proof: For $t \in [0,1]$ consider the function

$$\phi(t) = f(y) - f(y + t(x-y)) + \langle f'(y + t(x-y)), t(x-y) \rangle.$$

Note that $\phi(0) = 0$. Then

$$\phi'(t) = t\langle f''(y + t(x-y))(x-y), x-y) \rangle.$$

Denote $r = \|x-y\|_y$. Since f is self-concordant, we have

$$\langle f''(y + t(x-y))(x-y), x-y \rangle \geq \tfrac{r^2}{(1+tr)^2}$$

(see (2.2)). Hence,

$$f(y) - f(x) + \langle f'(x), x-y \rangle = \phi(1) - \phi(0) = \int\limits_0^1 \phi'(t)dt \geq \int\limits_0^1 \tfrac{tr^2 dt}{(1+tr)^2} = \zeta(r),$$

as required in (2.5). If $r < 1$, we have also

$$\langle f''(y + t(x-y))(x-y), x-y \rangle \leq \tfrac{r^2}{(1-tr)^2}$$

(see(2.1)). Hence,

$$f(y) - f(x) + \langle f'(x), x-y \rangle = \phi(1) - \phi(0) = \int\limits_0^1 \phi'(t)dt \leq \int\limits_0^1 \tfrac{tr^2 dt}{(1-tr)^2} = \zeta_*(r),$$

and that is (2.6).

In order to prove two others inequalities, note that the Legendre transformation of f

$$f_*(s) = \max_x [\langle s, x \rangle - f(x)]$$

is nondegenerate and self-concordant along with f (Proposition 2.1.(iv)). Therefore, using (2.5) and (2.6) at the points

$$u = f'(x), \quad v = f'(y)$$

we have

$$f_*(v) - f_*(u) - \langle f_*'(u), v-u \rangle \geq \zeta(\|v-u\|_v),$$

$$f_*(v) - f_*(u) - \langle f_*'(u), v-u \rangle \leq \zeta_*(\|v-u\|_v).$$

It remains to note that

$$f_*(u) = \langle f'(x), x \rangle - f(x), \quad f_*'(u) = x,$$

$$f_*(v) = \langle f'(y), y \rangle - f(y), \quad f_*'(v) = y,$$

and that $\|h\|_v = \langle f_*''(v)h, h \rangle^{1/2} = \langle [f''(y)]^{-1}h, h \rangle^{1/2} = \|h\|_y^*$ for $h \in E^*$. □

We will need some bounds on the variation of the gradient of a self-concordant barrier in terms of Minkowski function. Let $\pi_z(x)$ be the Minkowski function of Q with the pole at $z \in Q$:

$$\pi_z(x) = \inf\{t > 0 \mid \ z + t^{-1}(x-z) \in Q\}.$$

Lemma 2.2 *Let u be an arbitrary point in Q. Then for any $v \in Q$ we have*

$$\|f'(v)\|_u^* \le \tfrac{\nu}{1-\pi_u(v)}. \tag{2.9}$$

Moreover, if $\langle f'(u), v-u\rangle \ge 0$ for some $v \in Q$, then

$$\|f'(v)\|_u^* \ge \tfrac{\pi_u(v)}{(\nu+2\sqrt{\nu})(1-\pi_u(v))}. \tag{2.10}$$

Finally, if $\langle f'(u), v-w\rangle = 0$ for some $v, w \in Q$, then

$$1-\pi_u(v) \ge \tfrac{1-\pi_u(w)}{1+\nu+2\sqrt{\nu}}. \tag{2.11}$$

Proof: The set $\mathrm{cl}\, Q$ contains a $\|\cdot\|_u$-ball of radius 1, which is centered at u. Consequently, this set contains a $\|\cdot\|_u$-ball B of radius $1-\pi_u(v)$, which is centered at v. Since f is a ν-self-concordant barrier, from Proposition 2.1.(v) we have

$$\langle f'(v), x-v\rangle \le \nu \quad \forall x \in B \subseteq \mathrm{cl}\, Q,$$

and (2.9) follows.

In order to prove (2.10), let us choose $v \in Q$ such that

$$\langle f'(u), v-u\rangle \ \ge 0,$$

and let $r = \|v-u\|_u$. The case of $r=0$ is trivial. Assuming $r>0$, let

$$\phi(t) = f(u + tr^{-1}(v-u)), \quad t \in \Delta = \{t \mid u + tr^{-1}(v-u) \in Q\}.$$

Note that ϕ is self-concordant barrier for Δ and the right endpoint of Δ is the point $T = r/\pi_u(v)$. By Proposition 2.1.(i), Δ contains the set $\{s : (s-t)^2\phi''(t) < 1\}$, whence $t + (\phi''(t))^{-1/2} \le T$. Thus, for $t \in \Delta$, $t \ge 0$ we have

$$\phi''(t) \ge \tfrac{1}{(T-t)^2} \tag{2.12}$$

Combining (2.12) with the relation $\phi'(0) = r^{-1}\langle f'(u), v-u\rangle \ge 0$, we get

$$\phi'(r) \ge \int\limits_0^r \tfrac{1}{(T-t)^2}dt = \tfrac{r}{T(T-r)} = \tfrac{\pi_u^2(v)}{r(1-\pi_u(v))}.$$

On the other hand, $\phi'(r) = \langle f'(v), v-u\rangle r^{-1} \le \|f'(v)\|_u^*$, and we come to

$$\|f'(v)\|_u^* \ge \tfrac{\pi_u(v)}{1-\pi_u(v)} \cdot \tfrac{\pi_u(v)}{r}. \tag{2.13}$$

Setting $x = u + \pi_u^{-1}(v)(v-u)$, we have $x \in \mathrm{cl}\, Q$ and $\langle f'(u), x-u\rangle \ge 0$, whence by Proposition 2.1.(v) we get

$$\tfrac{r}{\pi_u(v)} = \|x-u\|_u \le \nu + 2\sqrt{\nu},$$

which combined with (2.13) implies (2.10).

To prove (2.11), let $\langle f'(v), w-v\rangle = 0$, and let $Q_v = \{x \in Q \mid \langle f'(v), x-v\rangle = 0\}$. Since v is the minimizer of a ν-self-concordant barrier on Q_v, the set Q_v, regarded as a full-dimensional subset of its affine span, contains a centered at v ellipsoid and is contained in a $(\nu + 2\sqrt{\nu})$ times larger concentric ellipsoid (Proposition 2.1.(i), (v)). It follows that there exists $x \in \mathrm{cl}\, Q_v$ such that

$$v = \tfrac{1}{1+\nu+2\sqrt{\nu}} w + \tfrac{\nu+2\sqrt{\nu}}{1+\nu+2\sqrt{\nu}} x.$$

Thus,

$$\pi_u(v) \le \tfrac{1}{1+\nu+2\sqrt{\nu}} \pi_u(w) + \tfrac{\nu+2\sqrt{\nu}}{1+\nu+2\sqrt{\nu}} \pi_u(x).$$

Note that $\pi_u(x) \le 1$. Hence,

$$1 - \pi_u(v) \ge \tfrac{1-\pi_u(w)}{1+\nu+2\sqrt{\nu}},$$

as required in (2.11). □

We conclude this section with two lower bounds of the size of the gradient of self-concordant function computed with respect to the local norm defined by its minimizer.

Lemma 2.3 *Assume that there exists a minimizer x_f of a ν-self-concordant barrier $f(x)$. Then for any $\bar{x} \in \mathrm{dom}\, f$ we have*

$$f(\bar{x}) - f(x_f) \le \nu \ln\left(1 + 3(1 + 2\nu^{-1/2}) \|f'(\bar{x})\|^*_{x_f}\right) \le \nu \ln\left(1 + 9\|f'(\bar{x})\|^*_{x_f}\right).$$

Proof: Denote

$$\delta = \bar{x} - x_f, \quad r = \|\delta\|_{x_f}, \quad \phi(t) = f\left(x_f + \tfrac{t}{r}\delta\right), \quad \Delta = \mathrm{dom}\, \phi.$$

Then $\phi(\cdot)$ is a ν-self-concordant barrier for the segment $\mathrm{cl}\, \Delta$. Note that $r \in \Delta$ and ϕ attains its minimum at $t = 0$. Moreover, $\phi''(0) = 1$. By Proposition 2.1.(v), we have

$$\phi'(t)(r - t) \le \nu, \quad \forall t \in [0, r],$$

and $r \le \nu + 2\sqrt{\nu}$. Besides this,

$$0 \le \phi'(t) \le \phi'(r), \quad \forall t \in [0, r].$$

Thus,

$$\begin{array}{rcl}
f(\bar{x}) - f(x_f) & = & \phi(r) - \phi(0) = \int\limits_0^r \phi'(t)dt \\
& \le & \int\limits_0^r \min[\phi'(r), \nu(r-t)^{-1}]dt \\
& = & \left\{ \begin{array}{cl} r\phi'(r), & \text{if } \phi'(r)r \le \nu, \\ \nu(1 + \ln(r\phi'(r)/\nu)), & \text{otherwise} \end{array} \right. \\
& \le & \nu \ln\left(1 + 3\frac{r\phi'(r)}{\nu}\right) \\
\left(\phi'(r) = \frac{1}{r}\langle f'(\bar{x}), \delta\rangle\right) & \le & \nu \ln\left(1 + 3(1 + 2\nu^{-1/2})\|f'(\bar{x})\|^*_{x_f}\right).
\end{array}$$

□

Lemma 2.4 *Assume that there exists a minimizer x_f of a self-concordant barrier $f(x)$. Then for any $x \in \operatorname{dom} f$ we have*

$$\|x - x_f\|_x \le (1 + 2\|x - x_f\|_{x_f}) \cdot \|f'(x)\|^*_{x_f}. \tag{2.14}$$

If in addition f is a ν-self-concordant barrier, then

$$\|x - x_f\|_x \le (2\nu + 4\sqrt{\nu} + 1) \cdot \|f'(x)\|^*_{x_f}. \tag{2.15}$$

Proof: Denote

$$\delta = x - x_f, \quad r = \|\delta\|_{x_f}, \quad \phi(t) = f\left(x_f + \tfrac{t}{r}\delta\right).$$

Note that ϕ is self-concordant and $\phi'(0) = 0$. Therefore, in view of (2.3) we have

$$\phi'(r) \ge \tfrac{r\phi''(r)}{1 + r\sqrt{\phi''(r)}},$$

whence

$$\phi''(r) \le (\phi'(r))^2 \left(1 + \tfrac{1}{r\sqrt{\phi''(r)}}\right)^2. \tag{2.16}$$

Further, $\phi''(t) = \frac{1}{r^2}\langle f''\left(x_f + \frac{t}{r}\delta\right)\delta, \delta\rangle$. Thus, $\phi''(0) = 1$ and $\phi''(r) \ge (1 + r)^{-2}$ by (2.2). Therefore (2.16) implies that

$$\phi''(r) \le (\phi'(r))^2 (2 + r^{-1})^2.$$

Note that

$$\phi'(r) = r^{-1}\langle f'(x), \delta\rangle \le \|f'(x)\|^*_{x_f}.$$

Combining the two last inequalities, we get

$$\langle f''(x)\delta, \delta\rangle = r^2\phi''(r) \le (1 + 2r)^2(\phi'(r))^2 \le (1 + 2r)^2 \left(\|f'(x)\|^*_{x_f}\right)^2,$$

as required in (2.14). Inequality (2.15) follows from (2.14) and the second part of Proposition 2.1.(v). □

3 Lower bounds on Riemannian distances

Let us establish lower bounds for the Riemannian distance between two points in Q in terms of different local norms defined by $f(x)$.

Lemma 3.1 *Let u and v belong to Q. Then for any $h \in E\backslash\{0\}$ we have*

$$\left|\ln \tfrac{\|h\|_u}{\|h\|_v}\right| \le \sigma(u, v), \tag{3.1}$$

and for any $\eta \in E^*\backslash\{0\}$

$$\left|\ln \tfrac{\|\eta\|^*_u}{\|\eta\|^*_v}\right| \le \sigma(u, v). \tag{3.2}$$

Moreover,

$$\left|\ln \tfrac{\lambda(u)+1}{\lambda(v)+1}\right| \le \sigma(u, v), \tag{3.3}$$

and

$$\ln \tfrac{\|f'(u)\|_v^* + 1}{\|f'(v)\|_v^* + 1} \le \sigma(u, v). \tag{3.4}$$

Besides this, if f is a ν-self-concordant barrier for $\operatorname{cl} Q$*, then*

$$|f(u) - f(v)| \le \sqrt{\nu}\sigma(u, v). \tag{3.5}$$

Proof: By continuity reasons, we may assume that both u, v are distinct from x_f. Let us fix $\epsilon > 0$. Consider a C^1-curve $\gamma(t) \in Q$, $0 \le t \le 1$, which satisfies the following conditions:

$$\gamma(0) = u, \quad \gamma(1) = v,$$

$$\gamma(t) \ne x_f \quad \forall t \in [0, 1],$$

$$\rho[\gamma(\cdot), 0, 1] \le \sigma(u, v) + \epsilon.$$

To prove (3.1), let us fix $h \in E \backslash \{0\}$ and set $\psi(t) = \langle f''(\gamma(t))h, h\rangle$. Then

$$|\psi'(t)| = |D^3 f(\gamma(t))[\gamma'(t), h, h]| \le 2\langle f''(\gamma(t))\gamma'(t), \gamma'(t)\rangle^{1/2}\psi(t),$$

whence

$$\left|\ln \tfrac{\langle f''(u)h,h\rangle^{1/2}}{\langle f''(v)h,h\rangle^{1/2}}\right| = \left|\tfrac{1}{2} \ln \tfrac{\psi(1)}{\psi(0)}\right| \le \rho[\gamma(\cdot), 0, 1] \le \sigma(u, v) + \epsilon,$$

and (3.1) follows. Relation (3.2) can be derived from (3.1) using the definition of the dual norm.

Further, denoting $\psi(t) = \lambda(\gamma(t))$, we have

$$\begin{aligned}
|\tfrac{d}{dt}\psi^2(t)| &= |-D^3 f(\gamma(t))[\gamma'(t), [f''(\gamma(t))]^{-1} f'(\gamma(t)), [f''(\gamma(t))]^{-1} f'(\gamma(t))] \\
&\quad +2\langle [f''(\gamma(t))]^{-1} f''(\gamma(t))\gamma'(t), f'(\gamma(t))\rangle| \\
&\le 2\langle f''(\gamma(t))\gamma'(t), \gamma'(t)\rangle^{1/2} \\
&\quad \times\langle f''(\gamma(t))[f''(\gamma(t))]^{-1} f'(\gamma(t)), [f''(\gamma(t))]^{-1} f'(\gamma(t))\rangle \\
&\quad +2\langle [f''(\gamma(t))]^{-1} f'(\gamma(t)), f'(\gamma(t))\rangle^{1/2} \langle f''(\gamma(t))\gamma'(t), \gamma'(t)\rangle^{1/2} \\
&= 2\psi(t)(\psi(t) + 1)\langle f''(\gamma(t))\gamma'(t), \gamma'(t)\rangle^{1/2}.
\end{aligned}$$

Since $\psi(\cdot) > 0$ on $[0, 1]$, we get

$$\left|\tfrac{d}{dt} \ln(1 + \psi(t))\right| \le \langle f''(\gamma(t))\gamma'(t), \gamma'(t)\rangle^{1/2},$$

whence

$$\left|\ln \tfrac{\lambda(u)+1}{\lambda(v)+1}\right| \le \rho[\gamma(\cdot), 0, 1] \le \sigma(u, v) + \epsilon,$$

and (3.3) follows.

To prove (3.4), let $\psi(t) = \|f'(\gamma(t))\|_v^*$ and $r(t) = \rho[\gamma(\cdot), 0, t]$. Then

$$\begin{array}{rcl}
\frac{d}{dt}\psi^2(t) & = & \frac{d}{dt}\langle f'(\gamma(t)), [f''(v)]^{-1} f'(\gamma(t))\rangle = 2\langle f''(\gamma(t))\gamma'(t), [f''(v)]^{-1} f'(\gamma(t))\rangle \\
& \leq & 2\langle f''(\gamma(t))\gamma'(t), \gamma'(t)\rangle^{1/2} \langle f''(\gamma(t))[f''(v)]^{-1} f'(\gamma(t)), [f''(v)]^{-1} f'(\gamma(t))\rangle^{1/2} \\
\text{(by (3.1))} & \leq & 2r'(t)e^{r(t)} \langle f''(v)[f''(v)]^{-1} f'(\gamma(t)), [f''(v)]^{-1} f'(\gamma(t))\rangle^{1/2} \\
& = & 2r'(t)e^{r(t)}\psi(t),
\end{array}$$

whence

$$\psi'(t) \leq r'(t)e^{r(t)},$$

so that

$$\begin{array}{rcl}
\|f'(u)\|_v^* + 1 & = & \psi(1) + 1 \leq \psi(0) + 1 + [e^{r(1)} - e^{r(0)}] \\
& \leq & \|f'(v)\|_v^* + \exp\{\sigma(u, v) + \epsilon\} \\
& \leq & (\|f'(v)\|_v^* + 1) \exp\{\sigma(u, v) + \epsilon\},
\end{array}$$

and (3.4) follows.

Finally, to prove (3.5) it suffices to note that if f is a ν-self-concordant barrier, then

$$\begin{array}{rcl}
|\frac{d}{dt} f(\gamma(t))| = |\langle f'(\gamma(t)), \gamma'(t)\rangle| & \leq & \lambda(\gamma(t)) \cdot \langle f''(\gamma(t))\gamma'(t), \gamma'(t)\rangle^{1/2} \\
& \leq & \sqrt{\nu} \cdot \langle f''(\gamma(t))\gamma'(t), \gamma'(t)\rangle^{1/2},
\end{array}$$

whence

$$|f(u) - f(v)| \leq \sqrt{\nu}\rho[\gamma(\cdot), 0, 1] \leq \sqrt{\nu}[\sigma(u, v) + \epsilon],$$

and (3.5) follows. □

4 Riemannian length of central path

Let f be a nondegenerate self-concordant function with domain $Q \subseteq E$. Given a nonzero vector $e \in E^*$, consider the associated *central path*

$$x(t) = \underset{x}{\operatorname{argmin}}[-t\langle e, x\rangle + f(x)]. \tag{4.1}$$

By Proposition 2.1.(iv) the domain of this curve is an open interval Δ on the axis. From now on we assume that this interval contains a given segment $[t_0, t_1]$ with $0 \leq t_0 < t_1 < \infty$. Note that by the Implicit Function Theorem the path $x(t)$ is continuously differentiable on its domain and satisfies the relations

$$f'(x(t)) = te; \quad x'(t) = [f''(x(t))]^{-1} e = t^{-1}[f''(x(t))]^{-1} f'(x(t)). \tag{4.2}$$

Our goal now is to obtain a useful upper bound on the Riemannian length $\rho[x(\cdot), t_0, t_1]$ of the central path.

Lemma 4.1 (i) *We always have*

$$\rho[x(\cdot),t_0,t_1] \leq \sqrt{[f(x(t_1))-f(x(t_0))]\ln\tfrac{t_1}{t_0}}. \tag{4.3}$$

If in addition $\lambda(x(t_0)) \geq \frac{1}{2}$, *then*

$$\ln\tfrac{t_1}{t_0} = \ln\tfrac{\|f'(x(t_1))\|^*_{x(t_0)}}{\|f'(x(t_0))\|^*_{x(t_0)}} \leq \sigma(x(t_0),x(t_1)) + \ln 3. \tag{4.4}$$

(ii) *If* $\lambda(x(t_0)) < \frac{1}{2}$, *then*

$$\rho[x(\cdot),t_0,t_1] \leq \ln 2 + \sqrt{\left[f(x(t_1)) - f(x(\widehat{t}))\right]\ln\tfrac{t_1}{\widehat{t}}}, \tag{4.5}$$

where $\widehat{t}$ *is the largest* $t \in [t_0,t_1]$ *such that* $\lambda(x(t)) \leq \frac{1}{2}$, *and*

$$\ln\tfrac{t_1}{\widehat{t}} \leq \ln\left(\max\left\{1, 4\|f'(x(t_1))\|^*_{x_f}\right\}\right) \leq \sigma(x(t_0),x(t_1)) + \ln 12. \tag{4.6}$$

Proof: Let

$$\phi(t) = f(x(t)), \quad r(t) = \rho[x(\cdot),t_0,t], \quad t_0 \leq t \leq t_1.$$

Since from (4.2) $x'(t) = t^{-1}[f''(x(t))]^{-1}f'(x(t))$, we have

$$\begin{aligned}
\phi'(t) &= \langle f'(x(t)), x'(t)\rangle = t^{-1}\langle [f''(x(t))]^{-1}f'(x(t)), f'(x(t))\rangle = t^{-1}\lambda^2(x(t)),\\
r'(t) &= \langle [f''(x(t))]x'(t), x'(t)\rangle^{1/2} = t^{-1}\langle [f''(x(t))]^{-1}f'(x(t)), f'(x(t))\rangle^{1/2}\\
&= t^{-1}\lambda(x(t)),
\end{aligned}$$

whence by the Cauchy inequality

$$\begin{aligned}
\rho^2[x(\cdot),t_0,t_1] &= \left(\int_{t_0}^{t_1} t^{-1}\lambda(x(t))dt\right)^2 \leq \left(\int_{t_0}^{t_1} t^{-1}\lambda^2(x(t))dt\right)\left(\int_{t_0}^{t_1} t^{-1}dt\right)\\
&= (\phi(t_1)-\phi(t_0))\ln\tfrac{t_1}{t_0},
\end{aligned}$$

as required in (4.3).

Let us prove now inequality (4.4). Assume that $\lambda(x(t_0)) \geq \frac{1}{2}$. Since

$$f'(x(t_1)) = \tfrac{t_1}{t_0}f'(x(t_0)),$$

using (3.4) and the fact that $\lambda(x(t_0)) = \|f'(x(t_0))\|^*_{x(t_0)} \geq \frac{1}{2}$, we get

$$\ln\tfrac{t_1}{t_0} = \ln\tfrac{\|f'(x(t_1))\|^*_{x(t_0)}}{\|f'(x(t_0))\|^*_{x(t_0)}} \leq \ln\left(3\tfrac{\|f'(x(t_1))\|^*_{x(t_0)}+1}{\|f'(x(t_0))\|^*_{x(t_0)}+1}\right) \leq \sigma(x(t_0),x(t_1)) + \ln 3,$$

as required in (4.4).

Now assume that $\lambda(x(t_0)) < \frac{1}{2}$. By Proposition 2.1(ii), under this assumption f attains its minimum on Q at a unique point x_f, the quantity

$$T = \max\{t \geq 0 : t \in \Delta, \lambda(x(t)) \leq \tfrac{1}{2}\}$$

is well-defined and $\lambda(x(T)) = \frac{1}{2}$. Multiplying the vector e in (4.1) by an appropriate constant, we may assume that $T = 1$. Note that since $\lambda(x(t_0)) < \frac{1}{2}$, we have

$$t_0 < \widehat{t} \le T = 1.$$

Let us first prove that

$$\rho[x(\cdot), 0, 1] \le \ln 2. \tag{4.7}$$

Note that $e = f'(x(1))$. Denote by f_* the Legendre transformation of f. Since f attains it minimum on Q, we have $0 \in \operatorname{dom} f_*$ and since $1 = T \in \Delta$, we have $e \in \operatorname{dom} f_*$. Besides this, f_* is nondegenerate and self-concordant on its domain in view of Proposition 2.1.(iv). Thus,

$$\langle e, f_*''(e)e\rangle = \langle f'(x(1)), [f''(x(1))]^{-1} f'(x(1))\rangle = \lambda^2(x(1)) = \tfrac{1}{4},$$

whence, by Proposition 2.1.(i)

$$\langle e, f_*''(te)e\rangle \le \tfrac{\lambda^2(x(1))}{(1-(1-t)\lambda(x(1)))^2}, \quad 0 \le t \le 1.$$

Hence,

$$\int_0^1 \langle e, f_*''(te)e\rangle^{1/2} dt \le \int_0^1 \tfrac{\lambda(x(1))}{1-(1-t)\lambda(x(1))} dt = -\ln(1 - \lambda(x(1))) = \ln 2.$$

On the other hand, we have $f'(x(t)) = tf'(x(1))$, whence

$$f''(x(t))x'(t) = f'(x(1)) = t^{-1} f'(x(t)),$$

and consequently

$$\begin{aligned} \rho[x(\cdot), 0, 1] &= \int_0^1 \langle f''(x(t))x'(t), x'(t)\rangle^{1/2} dt = \int_0^1 \langle f'(x(1)), [f''(x(t))]^{-1} f'(x(1))\rangle^{1/2} dt \\ &= \int_0^1 \langle e, f_*''(f'(x(t)))e\rangle^{1/2} dt = \int_0^1 \langle e, f_*''(te)e\rangle^{1/2} dt \le \ln 2, \end{aligned}$$

as required in (4.7).

Now let us prove (4.5). In the case of $t_1 \le 1\, [\equiv T]$ inequality (4.5) immediately follows from (4.7). In the case of $1 < t_1$ we have $\widehat{t} = 1$, and

$$\rho[x(\cdot), t_0, t_1] \le \rho[x(\cdot), 0, t_1] = \rho[x(\cdot), 0, 1] + \rho[x(\cdot), 1, t_1] \le \ln 2 + \rho[x(\cdot), 1, t_1].$$

Bounding $\rho[x(\cdot), 1, t_1] = \rho[x(\cdot), \widehat{t}, t_1]$ from above by (4.3), we get (4.5).

It remains to prove (4.6). There is nothing to prove if $\widehat{t} = t_1$. Thus, we may assume that $\widehat{t} < t_1$, whence, in particular, $\lambda(x(\widehat{t})) = \frac{1}{2}$. In view of the latter observation we can apply (4.4) and get

$$\ln \tfrac{t_1}{\widehat{t}} = \ln \frac{\|f'(x(t_1))\|^*_{x(\widehat{t})}}{\|f'(x(\widehat{t})\|^*_{x(\widehat{t})}} \le \sigma(x(\widehat{t}), x(t_1)) + \ln 3,$$

or, which is the same,

$$\ln \tfrac{t_1}{\widehat{t}} = \ln \left(2\|f'(x(t_1))\|^*_{x(\widehat{t})}\right) \le \sigma(x(\widehat{t}), x(t_1)) + \ln 3. \tag{4.8}$$

By (4.7) we have

$$\sigma(x(t_0), x(\hat{t})) \le \rho[x(\cdot), t_0, \hat{t}] \le \rho[x(\cdot), 0, \hat{t}] \le \rho[x(\cdot), 0, 1] \le \ln 2, \tag{4.9}$$

whence by triangle inequality

$$\sigma(x(\hat{t}), x(t_1)) \le \ln 2 + \sigma(x(t_0), x(t_1)),$$

and, using (3.2) and (4.9) we get

$$\tfrac{1}{2}\|f'(x(t_1))\|^*_{x_f} \le \|f'(x(t_1))\|^*_{x(\hat{t})} \le 2\|f'(x(t_1))\|^*_{x_f}.$$

Combining these relations and (4.8), we come to (4.6). □

Remark 4.1 *Note that in the proof of (4.3) we did not use the fact that f is self-concordant.*

We can establish now the *sub-geodesic* property of the central path associated with a self-concordant barrier (see [6] for terminology).

Theorem 4.1 *Let f be a nondegenerate ν-self-concordant barrier for $\mathrm{cl}\, Q$, and let $x(t)$ be the central path given by*

$$f'(x(t)) = te, \quad 0 \le t_0 \le t \le t_1.$$

Then

$$\rho[x(\cdot), t_0, t_1] \le \ln 2 + \nu^{1/4}\sqrt{\sigma(x(t_0), x(t_1))\,[\sigma(x(t_0), x(t_1)) + \ln 12]}. \tag{4.10}$$

If $\lambda(x(t_0)) \ge \frac{1}{2}$, then

$$\rho[x(\cdot), t_0, t_1] \le \nu^{1/4}\sqrt{\sigma(x(t_0), x(t_1))\,[\sigma(x(t_0), x(t_1)) + \ln 3]}. \tag{4.11}$$

Proof: It suffices to combine (4.3), (4.5), (3.5), (4.4) and (4.6). □

5 Applications

In this section we apply the results of Section 4 to the analysis of several short-step interior-point methods. We consider the following three problems:

1. Finding an approximation to the minimizer of a self-concordant function f.
2. Finding a point in a nonempty intersection of a bounded convex open domain Q and an affine plane. We assume that Q is represented by a ν-self-concordant barrier f with $\mathrm{dom}\, f = Q$, and that the minimizer x_f is known.
3. Finding an ϵ-solution to the optimization problem $\min\limits_{\mathrm{cl}\, Q}\langle c, x\rangle$, with Q represented in the same way as in Item 2.

Our main results state that in problems 2 and 3 (same as in problem 1, provided that f is a ν-self-concordant barrier), appropriate well-known short-step path-following methods are "suboptimal" within the factor $\nu^{1/4}$. Namely, the number of Newton steps, which is required by the methods, coincides, up to a factor $O(1)\nu^{1/4}$, with the Riemannian distance, between the starting point and the set of solutions. Note that this distance, as shown in [6], is a natural *lower* bound on the number of iterations of the short-step interior-point methods.

5.1 Minimization of a self-concordant function

Consider the following problem:

$$\min_{x \in Q} f(x), \tag{5.1}$$

where f is a nondegenerate below bounded self-concordant function with $\operatorname{dom} f = Q$. Assume that we have a starting point $\bar{x} \in Q$. Let us analyze the efficiency of a short-step path-following method $\mathcal{M}(\bar{x})$, which traces the central path $x(t)$:

$$f'(x(t)) = t f'(\bar{x}), \quad t \in [0, 1]. \tag{5.2}$$

as t decreases. By Proposition 2.1.(ii), $0 \in \operatorname{dom} f_*$, and of course $f'(\bar{x}) \in \operatorname{dom} f_*$. Consequently, the path is well-defined (Proposition 2.1.(iv)), and $x(0)$ is the minimizer x_f of $f(x)$ over Q.

To avoid trivialities, we assume that $\lambda(\bar{x}) > \frac{1}{2}$, i.e., that $\bar{x}$ does not belong to the domain of quadratic convergence of the Newton method as applied to f. Note that in this case

$$f(\bar{x}) - f(x_f) \geq O(1), \quad \|f'(\bar{x})\|^*_{x_f} \geq O(1), \quad \sigma(x_f, \bar{x}) \geq O(1) \tag{5.3}$$

(from now on, $O(1)$'s are positive absolute constants). Indeed, the first inequality is readily given by Proposition 2.1.(iii); the second inequality follows from the first one in view of (2.8) applied with $x = \bar{x}, y = x_f$, and the third inequality follows from the second one in view of (3.4).

Applying Lemma 4.1.(ii) and taking into account (5.3), we get the following

Theorem 5.1 *Let $\lambda(\bar{x}) > \frac{1}{2}$ (i.e., the starting point does not belong to the domain of quadratic convergence of the Newton method as applied to f). Then*

$$\begin{aligned} N(\bar{x}) &\leq O(1)\sqrt{[f(\bar{x}) - f(x_f)] \ln\left(1 + \|f'(\bar{x})\|^*_{x_f}\right)} \\ &\leq O(1)\sqrt{[f(\bar{x}) - f(x_f)]\, \sigma(x_f, \bar{x})}. \end{aligned} \tag{5.4}$$

Let us discuss the complexity bound (5.4).

1. The only previously known efficiency estimate for the problem (5.1) is

$$O(1)(f(\bar{x}) - f(x_f)) \tag{5.5}$$

Newton steps (recall that we do not assume f to be a self-concordant barrier). The simplest method providing us with this estimate is the usual damped Newton method.

However, using a standard argumentation, it is not difficult to see that the short-step path-following scheme as applied to (5.1) also share the same estimate (5.5). Let us demonstrate that the bound (5.4) is sharper. Indeed, using inequality (2.7), we have

$$f(\bar{x}) - f(x_f) \geq \ln(1 + \|f'(\bar{x})\|^*_{x_f}) - 1,$$

so that the bound (5.5) on $N(\bar{x})$ follows from the first inequality in (5.4) combined with initial conditions (5.3).

Note that the bound (5.4) can be much smaller than (5.5).

Example 5.1 *Let B_n be the unit n-dimensional box:*

$$B_n = \{x \in R^n : \mid x^{(i)} \mid \leq 1,\ i = 1, \dots, n\},$$

and $f(x) = -\sum\limits_{i=1}^{n} \ln(1 - (x^{(i)})^2)$. Without loss of generality we may assume that $\bar{x}^{(i)} \geq 0$, so that $\bar{x}^{(i)} = [1 - \epsilon_i]^{1/2}$, for some $\epsilon_i \in (0, 1]$, $i = 1, \dots, n$. Then

$$f(\bar{x}) = \sum_{i=1}^{n} \ln \tfrac{1}{\epsilon_i}.$$

At the same time,

$$\left(\|f'(\bar{x})\|^*_{x_f}\right)^2 = \tfrac{1}{2} \sum_{i=1}^{n} \tfrac{4(\bar{x}^{(i)})^2}{(1-(x^{(i)})^2)^2} = 2 \sum_{i=1}^{n} \tfrac{1-\epsilon_i}{\epsilon_i^2} \leq 2 \sum_{i=1}^{n} \tfrac{1}{\epsilon_i^2} \leq \tfrac{2n}{\min\limits_{1 \leq i \leq n} \epsilon_i^2}.$$

It follows that the ratio of the complexity bound (5.4) to the bound (5.5) does not exceed

$$O(1) \left(\left[1 + \max_{1 \leq i \leq n} \ln \tfrac{n}{\epsilon_i}\right] / \left[1 + \sum_{i=1}^{n} \ln \tfrac{1}{\epsilon_i}\right] \right)^{1/2},$$

and the latter quantity can be arbitrary close to $n^{-1/2}$. □

2. Consider a particular case of problem (5.1), with f being a ν-self-concordant barrier. In this case the known complexity estimate for a short-step path-following scheme is

$$O(1)\sqrt{\nu} \ln \left(1 + \nu \|f'(\bar{x})\|^*_{x_f}\right), \tag{5.6}$$

(see, for example, [4]). However, from Lemma 2.3 it is clear that the estimate (5.4) is sharper.

3. As it is shown in [6], a natural lower bound on the number of Newton steps in every short-step interior-point method for solving (5.1) is the Riemannian distance from $\bar{x}$ to x_f. In the case when f is a ν-self-concordant barrier, the performance of the path-following scheme, in view of Theorem 4.1 (or in view of (5.4) combined with (3.5)), is at most $O(1)\nu^{1/4}$ times worse than this lower bound.

5.2 Finding a feasible point

Consider the following problem:

$$\text{Find a point } \bar{x} \in \mathcal{F} = \{x \in Q,\ Ax = b\}, \tag{5.7}$$

where Q is an open and bounded convex set endowed with a ν-self-concordant barrier $f(x)$, and $A : x \mapsto Ax$ is a linear mapping from E onto a linear space F. Since the mapping A is an onto one, the conjugate mapping $A^* : F^* \to E^*$ is an embedding. From now on, we assume that problem (5.7) is feasible, and that we know the minimizer x_f of f on Q. Without loss of generality, we assume that $x_f = 0$.

Let f_* be the Legendre transformation of f. Since $x_f = 0$, we have

$$f_*'(0) = 0. \tag{5.8}$$

Note that $f_*(s)$ is a self-concordant function with $\operatorname{dom} f_* = E^*$ such that

$$\langle s, f_*''(s)s\rangle \le \nu \quad \forall s \in E^*, \tag{5.9}$$

(see [5], Theorem 2.4.2). In order to avoid trivial cases, let us assume that

$$\{x :\ Ax = b\} \bigcap \{x :\ \|x\|_0 < 1\} = \emptyset \tag{5.10}$$

(recall that $\|\cdot\|_0 \equiv \|\cdot\|_{x_f}$). Indeed, otherwise a solution to (5.7) can be found by projecting the origin onto the plane $Ax = b$ in the Euclidean metric $\|\cdot\|_0$ (see Proposition 2.1(i)).

In order to solve (5.7), we can trace the path of minimizers of f on the sets E_t,

$$x(t) = \operatorname*{argmin}_{E_t} f(x), \quad E_t = \{x \in Q \mid Ax = tb\},$$

as t varies from 0 to 1. Note that this path is well-defined: E_0 and E_1 are nonempty and bounded by assumption. Therefore all E_t, $t \in (0,1)$ are also nonempty and bounded.

As it is shown in [6], the path $x(\cdot)$ can be traced by an appropriate short-step path-following sequence, which length is proportional to $\rho[x(\cdot), 0, 1]$. Thus, in order to establish the complexity of our problem, we need to find some bounds on the Riemannian length of the path $x(t)$.

Observe that tracing $x(t)$ as t varies from 1 to 0 is equivalent to tracing a dual central path

$$y(t) : \quad h'(y(t)) = th'(0), \tag{5.11}$$

but with t varying from 0 to 1. This path is associated with a non-degenerate (since A^* is an embedding) self-concordant function

$$h(y) = f_*(A^*y) - \langle y, b\rangle : F^* \to R.$$

The relation between the paths $x(t)$ and $y(t)$ is given by the following Lemma.

Lemma 5.1 *For any $t \in [0,1]$ we have*

$$x(1-t) = f_*'(A^*y(t)). \tag{5.12}$$

Proof: Indeed, from the origin of $x(1-t)$ it follows that $f'(x(1-t)) \in (\operatorname{Ker} A)^{\perp}$. This means that $f'(x(1-t)) = A^* y(t)$ for certain uniquely defined $y(t)$, and this is equivalent to (5.12). Besides this, $Ax(1-t) = (1-t)b$, whence $Af'_*(A^*y(t)) = (1-t)b$, or

$$h'(y(t)) = Af'_*(A^*y(t)) - b = -tb = th'(0),$$

where the concluding equality is readily given by (5.8). We have arrived at (5.11). $\square$

It turns out that the Riemannian length $\rho[y(\cdot), 0, 1]$ of the path $y(\cdot)$ in the Riemannian structrure given by $h(\cdot)$ is equal to $\rho[x(\cdot), 0, 1]$:

$$\begin{array}{rcl} \rho[x(\cdot),0,1] & = & \int\limits_0^1 \langle f''(x(t))x'(t), x'(t)\rangle^{1/2} dt \\ & = & \int\limits_0^1 \langle f''(f'_*(A^*y(1-t)))f''_*(A^*y(1-t))A^*y'(1-t), \\ & & f''_*(A^*y(1-t))A^*y'(1-t)\rangle^{1/2} dt \\ & = & \int\limits_0^1 \langle A^*y'(1-t), f''_*(A^*y(1-t))A^*y'(1-t)\rangle^{1/2} dt \\ & = & \int\limits_0^1 \langle h''(y(1-t))y'(1-t), y'(1-t)\rangle^{1/2} dt = \rho[y(\cdot),0,1] \end{array}$$

Observe that by (5.11), $y(1) = 0$, while $y(0)$ is the minimizer of $h(\cdot)$. In view of the latter fact and Lemma 4.1 we have

$$\rho[x(\cdot),0,1] = \rho[y(\cdot),0,1] \le O(1)\left[\ln 2 + \sqrt{[h(y(1)) - h(y(0))]\ln \tfrac{1}{\widehat{t}}}\right], \tag{5.13}$$

where $\widehat{t}$ is the largest $t \in [0,1]$ such that $\lambda_h(y(t)) \le \frac{1}{2}$, $\lambda_h(y)$ being the local noram of the gradient of $h(\cdot)$ at $y \in \operatorname{dom} h(\cdot) \equiv F^*$.

Since $y(1) = 0$, we have $h'(y(1)) = -b$. Moreover, the point

$$z = [f''(0)]^{-1}A^*[A[f''(0)]^{-1}A^*]^{-1}b$$

clearly belongs to E_1, so that by (5.10) we have

$$\begin{array}{rcl} 1 & \le & \|z\|_0^2 = \langle b, [A[f''(0)]^{-1}A^*]^{-1}b\rangle = \langle b, [Af''_*(0)A^*]^{-1}b\rangle \\ & = & \langle h'(0), [h''(0)]^{-1}h'(0)\rangle = \lambda_h^2(0) = \lambda_h^2(y(1)). \end{array}$$

From $\lambda_h(y(1)) \ge 1$ and $\lambda_h(y(0)) = 0$ it follows that

$$\begin{array}{rcl} h(y(1)) - h(y(0)) & = & h(y(1)) - \min h \ge 1 - \ln 2, \\ \lambda_h(\widehat{t}) & = & \frac{1}{2}, \\ \ln \frac{1}{\widehat{t}} & \ge & O(1) \end{array} \tag{5.14}$$

(see Proposition 2.1). Consequently, (5.13) implies that

$$\rho[y(\cdot),0,1] \le \sqrt{[h(y(1))-h(y(0))]\ln\tfrac{1}{\hat{t}}}. \tag{5.15}$$

The next statement expresses the complexity bound (5.15) in terms of f.

Theorem 5.2 *For every $x\in\mathcal{F}\equiv E_1\cap Q$ we have*

$$\begin{array}{rcl}\rho[x(\cdot),0,1] & \le & O(1)\sqrt{[f(x)-f(x_f)]\,[O(1)+\ln\nu+\ln\|f'(x)\|^*_{x_f}]}\\ & \le & O(1)\sqrt{\nu}\ln\Big(O(1)\nu\|f'(x)\|^*_{x_f}\Big).\end{array} \tag{5.16}$$

Proof: Recall that we have assumed $x_f=0$. Observe, first, that

$$h(y(1)) = h(0) = f_*(0) = -\min_Q f = -f(0).$$

Therefore, using (5.12), we get

$$\begin{array}{rcl}h(y(0)) & = & f_*(A^*y(0))-\langle y(0),b\rangle\\ & = & \langle A^*y(0),f_*'(A^*y(0))\rangle - f(f_*'(A^*y(0)))-\langle y(0),b\rangle\\ & = & \langle y(0),Ax(1)-b\rangle - f(x(1)) = -f(x(1)).\end{array}$$

Thus,

$$h(y(1))-h(y(0)) = f(x(1))-\min_Q f = f(x(1))-f(0). \tag{5.17}$$

Let us prove now that

$$\ln\tfrac{1}{\hat{t}} \le O(1)\ln\langle f''(x(1))x(1),x(1)\rangle. \tag{5.18}$$

Denote $d=\|h'(y(1))\|^*_{y(0)}$. Then, either $d\ge 1$, or $d<1$. In the latter case applying (2.8) with h playing the role of f and $x=y(1)$, $y=y(0)$, we get

$$h(y(1))-h(y(0)) \le \ln(1-d)+\tfrac{d}{1-d},$$

which combined with the first relation in (5.14) results in

$$\ln(1-d)+\tfrac{d}{1-d} \ge O(1),$$

whence in any case $d\ge O(1)$.

The last conclusion combined with Lemma 4.1.(ii) (see (4.6)) implies that

$$\ln\tfrac{1}{\hat{t}} \le O(1)\ln d. \tag{5.19}$$

Recalling that $y(1)=0$, $h'(0)=-b$, we get

$$\begin{array}{rcl}d^2 & = & \langle h'(y(1)),[h''(y(0))]^{-1}h'(y(1))\rangle = \langle b,[Af_*''(A^*y(0))A^*]^{-1}b\rangle\\ & = & \langle Ax(1),[A[f''(x(1))]^{-1}A^*]^{-1}Ax(1)\rangle \le \langle f''(x(1))x(1),x(1)\rangle,\end{array}$$

and (5.18) follows from (5.19).

Combining (5.15), (5.17), (5.18) and (2.15), we come to the following inequality

$$\rho[x(\cdot),0,1] \leq O(1)\sqrt{[f(x(1)) - f(0)]\ln\left(O(1)\nu\|f'(x(1))\|_0^*\right)}. \tag{5.20}$$

Let us derive inequality (5.16) from (5.20). Since $x(1)$ is the minimizer of f on $\mathcal{F}$ and $x(0) = 0$, for every $x \in \mathcal{F}$ we have

$$f(x(1)) - f(x(0)) \leq f(x) - f(0) \leq \langle f'(x), x\rangle \leq \|x\|_0\|f'(x)\|_0^*.$$

We already have mentioned that $\|x\|_0 \leq \nu + 2\sqrt{\nu}$ for every $x \in Q$. Thus, the first part of (5.16) is proved. Further, let $x \in \mathcal{F}$. Applying (2.9) with $u = 0$ and $v = x(1)$ we have

$$\begin{array}{rcl} \|f'(x(1))\|_0^* & \leq & \frac{\nu}{1-\pi_0(x(1))} \\ [(2.11.)\ \text{with}\ u=0, v=x(1), w=x] & \leq & \frac{\nu(1+\nu+2\sqrt{\nu})}{1-\pi_0(x)} \\ [(2.10.)\ \text{with}\ u=0, v=x] & \leq & \frac{\nu(\nu+2\sqrt{\nu})(1+\nu+2\sqrt{\nu})}{\pi_0(x)}\|f'(x)\|_0^* \\ & \leq & O(1)\nu^4\|f'(x)\|_0^*, \end{array}$$

the concluding inequality being given by (5.10) combined with the fact that Q is contained in $\|\cdot\|_0$-ball of the radius $\nu + 2\sqrt{\nu}$ centered at 0 (Proposition 2.1.(v)). Thus, we get the second part of inequality (5.16). □

Corollary 5.1 *Under assumption* (5.10)*, the Riemannian length of the central path can be bounded from above as*

$$\rho[x(\cdot),0,1] \leq O(1)\nu^{1/4}(\sigma(0,\mathcal{F}) + \ln\nu), \tag{5.21}$$

where $\sigma(x, X)$ *is the infimum of Riemannian lengths of curves starting at a point* x *and ending at a point from the set* X*.*

Proof: Let $x \in \mathcal{F}$. By (3.5), we have

$$f(x) - f(0) \leq \sqrt{\nu}\sigma(0,x),$$

while by (3.4) we have

$$\ln(1 + \|f'(x)\|_0^*) \leq \sigma(0,x).$$

Combining these inequalities and (5.16), we get

$$\rho[x(\cdot),0,1] \leq O(1)\nu^{1/4}(\sigma(0,x) + \ln\nu);$$

since the resulting inequality is valid for every $x \in \mathcal{F}$ and $\sigma(0,x) \geq O(1)$ by (5.10), (5.21) follows. □

Note that in view of Example 1.1 this statement is not valid for an unbounded Q.

5.3 Central path in standard minimization problem

Now let us look at the optimization problem

$$\min\{\langle c, x\rangle : x \in \operatorname{cl} Q\}, \tag{5.22}$$

where Q is a bounded open convex set endowed with a ν-self-concordant barrier f. Assuming that the minimizer x_f of f on Q is available, consider the standard f-generated short-step path-following method for solving (5.22), where one traces the central path

$$x(t) : \quad f'(x(t)) = -tc, \quad t \geq 0,$$

as $t \to \infty$. The standard complexity bound for the number of Newton steps required to trace the segment $0 < t_0 \leq t \leq t_1$ of the path is

$$N_f(t_0, t_1) \leq O(1)\left(1 + \sqrt{\nu} \ln \tfrac{t_1}{t_0}\right). \tag{5.23}$$

The result of Lemma 4.1 combined with the technique proposed in [6], yields another bound:

$$\begin{array}{rcl} N_f(t_0, t_1) & \leq & O(1)\left(1 + \rho[x(\cdot), t_0, t_1]\right), \\ \rho[x(\cdot), t_0, t_1] & \leq & \sqrt{[f(x(t_1)) - f(x(t_0))] \ln \tfrac{t_1}{t_0}}. \end{array} \tag{5.24}$$

Observe that the new bound is sharper than the old one. Indeed, by (3.5) we have

$$f(x(t_1)) - f(x(t_0)) \leq \sqrt{\nu}\sigma(x(t_0), x(t_1)) \leq \sqrt{\nu}\rho[x(\cdot), t_0, t_1],$$

so that the second inequality in (5.24) says that $\rho[x(\cdot), t_0, t_1] \leq \sqrt{\nu} \ln \frac{t_1}{t_0}$. With this upper bound for $\rho[x(\cdot), t_0, t_1]$, the first inequality in (5.24) implies (5.23). As it is demonstrated by Example 5.1, the ratio of the right-hand side in (5.24) to that one of (5.23) can be as small as $O(1)\nu^{-1/2}$.

Bound (5.24) allows to obtain certain result on suboptimality of the path-following method in the family of all short-step interior point methods associated with the same self-concordant barrier. Namely, consider a segment $t \in [t_0, t_1]$, $t_0 > 0$, $t_1 < \infty$, of the central path and assume that at certain moment we stand at the point $x(t_0)$. Starting from this moment, the path-following method reaches the point $x(t_1)$ with the value of the objective $\langle c, x(t_1)\rangle < \langle c, x(t_0)\rangle$ in $N_f(t_0, t_1)$ Newton steps. Let us ask ourselves whether it is possible to reach a point with the value of the objective at least $\langle c, x(t_1)\rangle$ by a short-step interior point method, associated with f, for which the number of iterations is essentially smaller than $N_f(t_0, t_1)$. As it is shown in [6], the number of steps of any competing method is bounded below by $O(1)$ times the Riemannian distance $\sigma(x(t_0), Q_{t_1})$ from $x(t_0)$ to Q_{t_1}, where

$$Q_{t_1} = \{x \in Q \mid \langle c, x\rangle \leq \langle c, x(t_1)\rangle\}.$$

Since by [6] $N_f(t_0, t_1) = O(\rho[x(\cdot), x(t_0), x(t_1)])$, the aforementioned question can be posed as follows:

(Q) *How large could be the ratio* $\rho[x(\cdot), x(t_0), x(t_1)]/\sigma(x(t_0), Q_{t_1})$ *?*

The answer is given by the following theorem.

Theorem 5.3 *Assume that the ellipsoid*

$$W(t_0) = \{x \mid \|x - x(t_0)\|_{x(t_0)} \le 0.1\}$$

does not intersect Q_{t_1}[1]. *Then*

$$\rho[x(\cdot), t_0, t_1] \le O(1)\nu^{1/4}\sqrt{\sigma(x(t_0), Q_{t_1})\,[\sigma(x(t_0), Q_{t_1}) + \ln \nu]}. \tag{5.25}$$

Proof: 1^0. Let $H = \{x \in Q \mid \langle c, x\rangle = \langle c, x(t_1)\rangle\}$. Since $x(t_0) \in Q\backslash Q_{t_1}$, we clearly have

$$\sigma(x(t_0), H) \le \sigma(x(t_0), Q_{t_1}). \tag{5.26}$$

Note that f attains its minimum on H at the point $x(t_1)$, so that

$$\langle f'(x(t_1)), x - x(t_1)\rangle = 0 \quad \forall x \in H. \tag{5.27}$$

Since $f'(x(t_1)) = -t_1 c$, $f'(x(t_0)) = -t_0 c$ and $\langle c, x(t_1)\rangle < \langle c, x(t_0)\rangle$, we have also

$$\langle f'(x(t_0)), x - x(t_0)\rangle = \langle f'(x(t_0)), x(t_1) - x(t_0)\rangle \ge 0 \quad \forall x \in H. \tag{5.28}$$

2^0. Let $x \in H$. Since f attains its minimum on H at the point $x(t_1)$ and in view of (3.5), we have

$$f(x(t_1)) - f(x(t_0)) \le f(x) - f(x(t_0)) \le \sqrt{\nu}\sigma(x(t_0), x). \tag{5.29}$$

Furthermore, by (3.4)

$$\begin{aligned} \|f'(x)\|^*_{x(t_0)} + 1 &\le \exp\{\sigma(x(t_0), x)\}\left(\|f'(x(t_0))\|^*_{x(t_0)} + 1\right) \\ &\le \exp\{\sigma(x(t_0), x)\}(\sqrt{\nu} + 1). \end{aligned} \tag{5.30}$$

Using (2.9) with $u = x(t_0)$ and $v = x(t_1)$ we get

$$\begin{aligned} \|f'(x(t_1))\|^*_{x(t_0)} &\le \tfrac{\nu}{1 - \pi_{x(t_0)}(x(t_1))} \\ [(2.11) \text{ with } u = x(t_0), v = x(t_1), w = x \text{ and } (5.27)] &\le \tfrac{\nu(1+\nu+2\sqrt{\nu})}{1 - \pi_{x(t_0)}(x)} \\ [(2.10) \text{ with } u = x(t_0),\ v = x \text{ and } (5.28)] &\le \tfrac{\nu(\nu+2\sqrt{\nu})(1+\nu+2\sqrt{\nu})}{\pi_{x(t_0)}(x)}\|f'(x)\|^*_{x(t_0)} \\ &\le O(1)\nu^4\|f'(x)\|^*_{x(t_0)}, \end{aligned}$$

where the concluding inequality follows from the fact that, on one hand, $W(t_0)$ does not intersect H, whence $\|x - x(t_0)\|_{x(t_0)} \ge 0.1$, and, on the other hand, the $\|\cdot\|_{x(t_0)}$-distance from $x(t_0)$ to the boundary of Q in the direction $x - x(t_0)$ does not exceed $\nu + 2\sqrt{\nu}$ in view of (5.28) and Proposition 2.1.(v).

---

[1] It is easily seen that the case when $W(t_0)$ intersects Q_{t_1} is trivial: $\rho[x(\cdot), t_0, t_1] \le O(1)$.

Thus,

$$\|f'(x(t_1))\|^*_{x(t_0)} \leq O(1)\nu^4 \|f'(x)\|^*_{x(t_0)} \leq O(1)\nu^5 \exp\{\sigma(x(t_0),x)\}, \tag{5.31}$$

the concluding inequality being given by (5.30).

Observe that since the ellipsoid $W(t_0)$ does not intersect Q_{t_1} and $x \in Q_{t_1}$, we have

$$\sigma(x(t_0),x) \geq O(1). \tag{5.32}$$

3^0. Assume first that $\lambda(x(t_0)) \geq \frac{1}{2}$. Then by Lemma 4.1 we have

$$\begin{array}{rcl} \rho[x(\cdot),t_0,t_1] & \leq & \sqrt{[f(x(t_1))-f(x(t_0))]\ln\frac{\|f'(x(t_1))\|^*_{x(t_0)}}{\|f'(x(t_0))\|^*_{x(t_0)}}} \\ \text{[by (5.29) and } \lambda(x(t_0)) \geq \frac{1}{2}] & \leq & \sqrt{\nu^{1/2}\sigma(x(t_0),x)\ln(2\|f'(x(t_1))\|^*_{x(t_0)})} \\ \text{[by (5.31) and (5.32)]} & \leq & O(1)\sqrt{\nu^{1/2}\sigma(x(t_0),x)(\sigma(x(t_0),x)+\ln\nu)} \end{array} \tag{5.33}$$

Now let $\lambda(x(t_0)) \leq \frac{1}{2}$. Then $\|x(t_0)-x_f\|_{x_f} \leq \frac{3}{4}$ by Proposition 2.1.(ii), whence, in view of Item (i) of the same proposition,

$$O(1) \leq \|h\|^*_{x(t_0)}/\|h\|^*_{x_f} \leq O(1) \quad \forall h \in E^*\backslash\{0\}. \tag{5.34}$$

Now by Lemma 4.1.(ii) we have

$$\begin{array}{rcl} \rho[x(\cdot),t_0,t_1] & \leq & \ln 2 + \sqrt{[f(x(t_1))-f(x(t_0))]\ln\left(\max\left\{1,4\|f'(x(t_1))\|^*_{x_f}\right\}\right)} \\ \text{[by (5.29), (5.34)]} & \leq & \ln 2 + \sqrt{\nu^{1/2}\sigma(x(t_0),x)\ln\left(\max\left\{1,O(1)\|f'(x(t_1))\|^*_{x(t_0)}\right\}\right)} \\ {[W(t_0)\cap H=\emptyset]} & \leq & \ln 2 + O(1)\sqrt{\nu^{1/2}\sigma(x(t_0),x)\ln(O(1)\|f'(x(t_1))\|^*_{x(t_0)})} \\ \text{[by (5.31), (5.32)]} & \leq & \ln 2 + O(1)\sqrt{\nu^{1/2}\sigma(x(t_0),x)(\sigma(x(t_0),x)+\ln\nu)}. \end{array}$$

Recalling (5.32), we conclude that

$$\rho[x(\cdot),t_0,t_1] \leq O(1)\nu^{1/4}\sqrt{\sigma(x(t_0),x)(\sigma(x(t_0),x)+\ln\nu)}.$$

Since $x \in H$ is arbitrary, we get

$$\rho[x(\cdot),t_0,t_1] \leq O(1)\nu^{1/4}\sqrt{\sigma(x(t_0),H)(\sigma(x(t_0),H)+\ln\nu)}.$$

□

We conclude that the ratio in (Q) is, up to logarithmic in ν terms, at most $\nu^{1/4}$.

References

[1] A. Edelman, T. Arias, and S. T. Smith, The geometry of algorithms with orthogonality constraints, *SIAM Journal on Matrix Analysis and Applications* **20** (1998), 303–353.

[2] S. Helgason, *Differential Geometry and Symmetric Spaces*, American Mathematical Society, Providence, RI, 2000.

[3] M. Koecher, Positivitätsbereiche im R^n, *American Journal of Mathematics* **79** (1957), 575–596.

[4] Yu. Nesterov. *Introductory Lectures on Convex Optimization. Basic course.* Kluwer, 2003.

[5] Yu. Nesterov, A. Nemirovskii. *Interior point polynomial methods in convex programming: Theory and Applications,*, SIAM, Philadelphia, 1994.

[6] Yu. Nesterov, M. J. Todd. On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Foundations of Computational Mathematics.*, No 2, 2002, pp.333-361.

[7] Yu.Nesterov, M.J.Todd and Y.Ye. Infeasible-start Primal-Dual Mathods and Infeasibility Detectors. *Mathematical Programming*, 84(1999), 227-267.

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

[9] K. Tanabe, Geometric method in nonlinear programming, *Journal of Optimization Theory and Applications* **30** (1980), 181–210.

[10] S. J. Wright, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, PA, 1997.