Augmented self-concordant barriers and nonlinear optimization problems with finite complexity

Yu. Nesterov[*] J.-Ph. Vial[†]

October 9, 2000

Abstract

In this paper we study special barrier functions for the convex cones, which are the sum of a self-concordant barrier for the cone and a positive-semidefinite quadratic form. We show that the central path of these augmented barrier functions can be traced with linear speed. We also study the complexity of finding the analytic center of the augmented barrier. This problem itself has some interesting applications. We show that for some special classes of quadratic forms and some convex cones, the computation of the analytic center requires an amount of operations independent on the particular data set. We argue that these problems form a class that is endowed with a property which we call finite polynomial complexity.

Keywords Augmented barrier, self-concordant functions, finite methods, nonlinear optimization.

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

[†]HEC, Université de Genève, 40 Bd du Pont d'Arve, CH-1211 Geneva, Switzerland. This author acknowledges the support of the grant 1214-057093-99 of the Swiss National Science Foundation.

1 Introduction

In the recent years, homotopy methods have become a focal point of interest in interior point methods for convex optimization. The basic idea of these methods is to trace approximately the *central path trajectory* $x(t)$. This path is defined by

$$x(t) = \arg\min_{x \in Q} \; t\langle c, x\rangle + F(x), \tag{1.1}$$

where $\langle \cdot, \cdot \rangle$ stands for the scalar product in R^n, c is a fixed cost vector, $Q \subset R^n$ is a bounded convex set and $F(x)$ is a convex barrier function for Q. The main issue for these methods is the following:

> *What is the class of barrier functions, for which we can follow the central path with a high speed?*

The answer to that question is given by the theory of *self-concordant functions* [3]. More specifically, the analysis relies on the concept of that *self-concordant barrier*. $F(x)$ is a self-concordant barrier if the following conditions are satisfied. Firstly, $F(x)$ must be a self-concordant function:

$$D^3F(x)[h,h,h] \le 2\langle F''(x)h, h\rangle^{3/2} \quad \forall x \in \operatorname{int} Q, \; h \in R^n.$$

Secondly, the local norm of the gradient of $F(x)$ must be uniformly bounded by some positive constant ν:

$$\langle F'(x), h\rangle^2 \le \nu \langle F''(x)h, h\rangle, \quad \forall x \in \operatorname{int} Q, \; h \in R^n.$$

The term ν is called the *parameter* of the barrier. In this framework, it can be proved that the approximations to the points $x(t)$ can be updated with a constant complexity for the change

$$t \to \left(1 \pm \frac{O(1)}{\sqrt{\nu}}\right) t. \tag{1.2}$$

The theory of self-concordant functions naturally extends to conic sets [3]. If $Q \equiv K$ is a proper cone, the theory introduces the further assumption that the self-concordant barrier $F(x)$ for this set is *logarithmically homogeneous*[1]:

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

With such a barrier one can trace the central path trajectory on the intersection of the cone K with some linear subspace [3].

Note that the problem of tracing the trajectory (1.1) for a standalone logarithmically homogeneous barrier has no sense. (In this case, the trajectory $\{x(t)\}_{t>0}$ either does not exist or it is a straight line.) However, in this paper we show, that there exists a class of barrier functions for convex cones, for which the tracing problem is meaningful and not trivial, while its complexity remains on the level of (1.2). We call these functions the *augmented self-concordant barriers*. Such a function $\Phi(x)$ is formed as follows:

$$\Phi(x) = \tfrac{1}{2}\langle Qx, x\rangle + F(x),$$

[1] The barrier possessing all above mentioned properties is called ν-normal barrier. The results on ν-normal barrier that we use in the paper can be found in [2]. A short summary of these results can also be found in [4].

where Q is a positive-semidefinite matrix and $F(x)$ is a ν-normal barrier for the cone K. Note that $\Phi(x)$ is a self-concordant function, but it is *not* a self-concordant barrier. So, the complexity of tracing its central path cannot be obtained from the standard results [3].

Augmented self-concordant barriers have already been used in the framework of cutting plane algorithms, either in the main iterative scheme [4], or in an auxiliary problem [1]. The present paper gives a more systematic treatment and discusses some additional applications. The more striking result in this paper is perhaps the fact that, for some nonlinear optimization problems, the complexity of a path-following method is independent of the particular data set. This strong property is remindful of the situation in linear algebra, for which interior point methods have *finite polynomial complexity.* For this reason, we suggest to call the methods possessing the above property "*finite methods* in nonlinear optimization".

The paper is organized as follows. In Section 2 we define the augmented self-concordant barriers and study their properties. In Section 3 we analyze the complexity of finding the analytic center of the augmented barrier. Our complexity results are given in terms of the quality of the starting point with respect to the global behavior of the quadratic form $\langle Qx, x\rangle$ on the cone K. The main motivations for such approach are some applications (Section 5), for which the complexity of finding the analytic center appears to be independent on a particular data set. We conclude the paper with a short discussion of the complexity results (Section 6).

2 Augmented self-concordant barriers

Definition 1 *Let K be a proper cone endowed with a ν-normal barrier $F(x)$ and let Q be a positive semidefinite matrix. The function*

$$\Phi(x) = \tfrac{1}{2}\langle Qx, x\rangle + F(x)$$

is called an augmented ν-self-concordant barrier *for the cone K.*

Remark 1 *An augmented self-concordant barrier with $Q \neq 0$ is* not *a self-concordant barrier itself. A self-concordant barrier has the property that the Newton decrement is uniformly bounded on* $\operatorname{int} K$*. An augmented self-concordant barrier is a self-concordant function, but the quadratic augmentation term precludes the uniform boundedness of the Newton decrement.*

Just as in the theory of self-concordant functions, we will use local norms for measuring the size of the vectors. We need three types of local norms. The first one is defined by the Hessian of the barrier $F(x)$:

$$\theta_x(u) = \langle F''(x)u, u\rangle^{1/2}, \quad \theta^*_x(g) = \langle g, [F''(x)]^{-1}g\rangle^{1/2}, \quad x \in \operatorname{int} K,\ u, g \in R^n.$$

The second one is given by the Hessian of the augmented barrier:

$$\lambda_x(u) = \langle \Phi''(x)u, u\rangle^{1/2}, \quad \lambda^*_x(g) = \langle g, [\Phi''(x)]^{-1}g\rangle^{1/2}, \quad x \in \operatorname{int} K,\ u, g \in R^n.$$

Finally, we need a mixed norm:

$$\left.\begin{aligned}\sigma_x(u) &= \langle \Phi''(x)u, [F''(x)]^{-1}\Phi''(x)u\rangle^{1/2},\\ \sigma_x^*(g) &= \langle F''(x)[\Phi''(x)]^{-1}g, [\Phi''(x)]^{-1}g\rangle^{1/2},\end{aligned}\right\} \quad x \in \operatorname{int} K,\ u, g \in R^n.$$

Note that $\Phi''(x) = F''(x) + Q \succeq F''(x)$. Therefore the above norms are related as follows:

$$\theta_x(u) \le \lambda_x(u) \le \sigma_x(u), \quad x \in \operatorname{int} K,\ u \in R^n. \tag{2.1}$$

$$\sigma_x^*(g) \le \lambda_x^*(g) \le \theta_x^*(g), \quad x \in \operatorname{int} K,\ g \in R^n. \tag{2.2}$$

In particular,

$$\sigma_x^*(\Phi'(x)) \le \lambda_x^*(\Phi'(x)) \le \theta_x^*(\Phi'(x)), \quad x \in \operatorname{int} K.$$

Note that $\lambda_x^*(\Phi'(x))$ is a local norm of the gradient of function $\Phi(x)$ with respect to its Hessian. It is uniformly bounded by $\nu^{1/2}$ only when the quadratic term is absent ($Q = 0$). However, it possible to derive a uniform bound on $\sigma_x^*(\Phi'(x))$ and a weaker bound on $\lambda_x^*(\Phi'(x))$.

Lemma 1 *For any $x \in \operatorname{int} K$ we have*

$$\sigma_x^*(\Phi'(x)) \le \nu^{1/2}, \tag{2.3}$$

$$\lambda_x^*(\Phi'(x)) \le [\langle Qx, x\rangle + \nu]^{1/2}. \tag{2.4}$$

Proof:
Indeed, since the barrier $F(x)$ is logarithmically homogeneous, we have for all $x \in \operatorname{int} K$

$$\Phi'(x) = Qx + F'(x) = Qx - F''(x)x = (Q - F''(x))x.$$

Therefore, for all $x \in \operatorname{int} K$

$$\left(\sigma_x^*(\Phi'(x))\right)^2 = \langle F''(x)[Q + F''(x)]^{-1}(Q - F''(x))x, [Q + F''(x)]^{-1}(Q - F''(x))x\rangle. \tag{2.5}$$

Let

$$H = (Q - F''(x))[Q + F''(x)]^{-1}F''(x)[Q + F''(x)]^{-1}(Q - F''(x)).$$

Let V be an orthogonal matrix of eigenvectors and Λ the associated diagonal matrix of eigenvalues of $[F''(x)]^{-1/2}Q[F''(x)]^{-1/2}$. Then

$$H = [F''(x)]^{1/2}V(\Lambda - I)(\Lambda + I)^{-2}(\Lambda - I)V^T[F''(x)]^{1/2} \preceq F''(x),$$

since for positive λ_i, one has $(\lambda_i - 1)^2(\lambda_i + 1)^{-2} \le 1$. Therefore, from equation (2.5),

$$\left(\sigma_x^*(\Phi'(x))\right)^2 \le \langle F''(x)x, x\rangle = \nu.$$

On the other hand,

$$\left(\lambda_x^*(\Phi'(x))\right)^2 = \langle [Q + F''(x)]^{-1}(Q - F''(x))x, (Q - F''(x))x\rangle.$$

Using a similar argument as above, we have

$$(Q - F''(x))[Q + F''(x)]^{-1}(Q - F''(x)) \preceq Q + F''(x).$$

Thus, (2.4) holds. $\square$

The result of Lemma 2.2 is crucial for the analysis of a path-following scheme. We also need an additional result on the convergence of the Newton method on the augmented barrier.

Consider the problem

$$\min_{x} \; [\psi(x) \equiv \langle c, x\rangle + \Phi(x)]\,. \tag{2.6}$$

Let us analyze the performance of the Newton method

$$x_+ = x - [\psi''(x)]^{-1}\psi'(x) \tag{2.7}$$

for finding the minimum of ψ. Note that ψ is a self-concordant function. Therefore the standard description of the region of quadratic convergence of the Newton scheme is as follows (see [3]):

$$\lambda^*_x(\psi'(x)) \equiv \langle \psi'(x), [\psi''(x)]^{-1}\psi'(x)\rangle^{1/2} < \frac{2}{3+\sqrt{5}} = 0.3819660\ldots.$$

Let us show that, due to the quadratic augmentation term, the region of quadratic convergence of the Newton scheme for problem (2.6) is wider.

Theorem 1 *Let x_+ be defined by (2.7) and $\sigma^*_x(\psi'(x)) < 1$. Then $x_+ \in \operatorname{int} K$ and*

$$\sigma^*_{x_+}(\psi'(x_+)) \le \lambda^*_{x_+}(\psi'(x_+)) \le \theta^*_{x_+}(\psi'(x_+)) \le \left(\frac{\sigma^*_x(\psi'(x))}{1-\sigma^*_x(\psi'(x))}\right)^2. \tag{2.8}$$

The region of quadratic convergence of the Newton method is then

$$\sigma^*_x(\psi'(x)) < \frac{2}{3+\sqrt{5}}.$$

Proof:
We prove first that the process is well-defined. Denote

$$r^2 \equiv \langle F''(x)(x_+ - x), x_+ - x\rangle.$$

By definition,

$$\begin{aligned}(\sigma^*_x(\psi'(x)))^2 &= \langle F''(x)[\Phi''(x)]^{-1}\psi'(x), [\Phi''(x)]^{-1}\psi'(x)\rangle,\\ &= \langle F''(x)[\psi''(x)]^{-1}\psi'(x), [\psi''(x)]^{-1}\psi'(x)\rangle = r^2.\end{aligned}$$

The statement $x_+ \in \operatorname{int} K$ follows from $r < 1$. Thus, the scheme (2.8) is well defined.

The first two inequalities in (2.8) follow from (2.2). Let us prove the last one. From the definition of x^+, one has $\psi'(x) + \psi''(x)(x_+ - x) = 0$. Thus,

$$\begin{aligned}
\psi'(x_+) \quad &= \psi'(x_+) - \psi'(x) - \psi''(x)(x_+ - x), \\
&= \Big[\int_0^1 [\psi''(x + \tau(x_+ - x)) - \psi''(x)] d\tau \Big] (x_+ - x), \\
&= \Big[\int_0^1 [F''(x + \tau(x_+ - x)) - F''(x)] d\tau \Big] (x_+ - x) \equiv G(x_+ - x).
\end{aligned}$$

In view of Corollary 4.1.4 [2] the matrix G is comparable to $F''(x)$:

$$-\left(r - \frac{r^2}{3} \right) F''(x) \preceq G \preceq \frac{r}{1-r} F''(x).$$

On the other hand, by Theorem 2.1.1 [3], we have $F''(x_+) \succeq (1-r)^2 F''(x)$. Thus,

$$G[F''(x_+)]^{-1} G \preceq \frac{1}{(1-r)^2} G[F''(x)]^{-1} G \preceq \frac{r^2}{(1-r)^4} F''(x).$$

We conclude that

$$\begin{aligned}
\left(\theta^*_{x_+}(\psi'(x_+)) \right)^2 \quad &= \langle [F''(x_+)]^{-1} \psi'(x_+), \psi'(x_+) \rangle, \\
&= \langle [F''(x_+)]^{-1} G(x_+ - x), G(x_+ - x) \rangle, \\
&\preceq \frac{r^2}{(1-r)^4} \langle F''(x)(x_+ - x), x_+ - x \rangle = \left(\frac{\sigma^*_x(\psi'(x))}{1 - \sigma^*_x(\psi'(x))} \right)^4 .
\end{aligned}$$

□

Remark 2 *In the above proof we did not use the fact that $F(x)$ is a self-concordant barrier for a cone. Thus, the result of Theorem 1 is also valid for an arbitrary self-concordant function $F(x)$.*

Theorem 1 makes it possible to construct a path-following scheme relative to the following central path trajectory:

$$x(t) = \arg\min_t \; \psi_t(x), \quad \psi_t(x) = t\langle c, x \rangle + \Phi(x), \; t > 0.$$

Let us analyze one iteration of the path-following scheme.

Theorem 2 *Assume that for some $t > 0$ we have a point $x \in \operatorname{int} K$ such that*

$$\sigma^*_x(\psi'_t(x)) \le \beta < 3 - \sqrt{5}.$$

Let us choose $t_+ = (1 + \Delta)t$ and

$$x_+ = x - [\psi''_t(x)]^{-1} \psi'_{t_+}(x). \tag{2.9}$$

If Δ satisfies the condition

$$|\Delta|(\beta+\sqrt{\nu}) \le \frac{\sqrt{\beta}}{1+\sqrt{\beta}} - \beta,$$

then $x_+ \in \operatorname{int} K$ and $\sigma^*_{x_+}(\psi'_{t_+}(x_+)) \le \theta^*_{x_+}(\psi'_{t_+}(x_+)) \le \beta$.

Proof:
In view of the initial centering condition and Lemma 1, we have

$$\beta \ge \sigma^*_x(\psi'_t(x)) = \sigma^*_x(tc + \Phi'(x)) \ge t\sigma^*_x(c) - \sigma^*_x(\Phi'(x)) \ge t\sigma^*_x(c) - \sqrt{\nu}.$$

Thus, $t\sigma^*_x(c) \le \beta + \sqrt{\nu}$ and

$$\begin{aligned}\sigma^*_x(\psi'_{t_+}(x)) &= \sigma^*_x(\Delta tc + \psi'_t(x)) \le |\Delta| t\sigma^*_x(c) + \sigma^*_x(\psi'_t(x)),\\ &\le |\Delta|(\beta+\sqrt{\nu}) + \beta \le \tfrac{\sqrt{\beta}}{1+\sqrt{\beta}}.\end{aligned}$$

Hence, in view of Theorem 1,

$$\theta^*_{x_+}(\psi'_{t_+}(x_+)) \le \left(\frac{\sigma^*_x(\psi'_{t_+}(x))}{1-\sigma^*_x(\psi'_{t_+}(x))}\right)^2 \le \beta.$$

□

In particular, for $\beta = \frac{1}{9}$ we have the following bounds for Δ: $|\Delta| \le \frac{5}{4+36\sqrt{\nu}}$.

3 Finding the analytic center

Consider now the problem of finding the analytic center of the augmented barrier:

$$\phi^* = \min_x \ \Phi(x) \equiv \min_x \ [\tfrac{1}{2}\langle Qx, x\rangle + F(x)], \tag{3.1}$$

where Q is a positive-semidefinite matrix. For the sake of simplicity, we assume that the the solution x^* of this problem always exists.

When Q is non-degenerate, problem to (3.1) has a nice symmetric dual.

Lemma 2 *Let $F_*(s)$ be a barrier conjugate to $F(x)$. Then the problem*

$$\phi_* = \min_s [\tfrac{1}{2}\langle s, Q^{-1}s\rangle + F_*(s)] \tag{3.2}$$

is dual to (3.1) with zero duality gap: $\phi^* + \phi_* = 0$.

Proof:
Indeed, in view of definition of the conjugate barrier we have

$$\begin{aligned}\min_x \ [\tfrac{1}{2}\langle Qx, x\rangle + F(x)] &= \min_x \ \max_s \ [\tfrac{1}{2}\langle Qx, x\rangle - \langle s, x\rangle - F_*(s)],\\ &= \max_s \ \min_x \ [\tfrac{1}{2}\langle Qx, x\rangle - \langle s, x\rangle - F_*(s)],\\ &= \max_s \ [-\tfrac{1}{2}\langle s, Q^{-1}s\rangle - F_*(s)].\end{aligned}$$

□

In the non-degenerate case, duality makes it possible to find the analytic center of the augmented barrier either from the primal or from the dual problem. Surprisingly enough, we shall see in Section 5 that the problems in this pair may have different complexity. Let us also mention that, in Section 4, we shall discuss duality theory for problem (3.1) with generically degenerate quadratic function.

Since the primal and dual problems have similar formulation, we focus without loss of generality on the primal problem (3.1). In the rest of this section, we derive complexity estimates for some basic minimization schemes.

Complexity estimates depend on characteristics of problem (3.1). Let us define the characteristics as follows. Choose some point $\hat{x} \in \operatorname{int} K$ and let the characteristics be:

$$\begin{aligned} \gamma_u(\hat{x}) &\geq \max_x \left[\langle Qx, x \rangle : \; -\langle F'(\hat{x}), x \rangle = 1, \; x \in K \right], \\ \gamma_l(\hat{x}) &\leq \min_x \left[\langle Qx, x \rangle : \; \langle F''(\hat{x})x, x \rangle = 1, \; x \in K \right]. \end{aligned}$$

Thus, for any $x \in K$ we have:

$$\gamma_u(\hat{x}) \langle F'(\hat{x}), x \rangle^2 \geq \langle Qx, x \rangle \geq \gamma_l(\hat{x}) \langle F''(\hat{x})x, x \rangle. \tag{3.3}$$

Note that $\gamma_l(\hat{x})$ can be positive even for a degenerate Q. Since $F(x)$ is a self-concordant barrier, we have

$$\langle F'(\bar{x}), x \rangle^2 \geq \langle F''(\bar{x})x, x \rangle \geq \frac{1}{\nu} \langle F'(\bar{x}), x \rangle^2$$

for any $x \in K$. So, we can use different combinations of these quadratic forms in the two-side bounds for the variation of the form $\langle Qx, x \rangle$ over the cone K. All these bounds are equivalent up to a factor that is polynomial in ν. Our choice (3.3) is motivated by some applications (see Section 5).

3.1 Damped Newton scheme

Let us choose some scaling coefficient $\rho > 0$. Consider the following Newton scheme:

$$x_0 = \rho \hat{x}, \quad x_{k+1} = x_k - \frac{[\Phi''(x_k)]^{-1}}{1 + \lambda_{x_k}(\Phi'(x_k))}, \quad k = 0, \ldots. \tag{3.4}$$

Theorem 3 *1) If we choose $\rho = \frac{1}{\sqrt{\nu}}$, then the process (3.4) enters the region of quadratic convergence of the Newton method (2.7) at most after*

$$O\left(\nu \gamma_u(\hat{x})\right) + O\left(\nu \ln \frac{1}{\gamma_l(\hat{x})}\right) + O\left(\nu \ln \nu\right).$$

iterations.

2) If we choose $\rho = \frac{1}{\sqrt{\nu \gamma_u(\hat{x})}}$, then the upper bound for the number of iterations is as follows:

$$O\left(\nu \ln \frac{\nu \gamma_u(\hat{x})}{\gamma_l(\hat{x})}\right).$$

Proof:
Indeed, (3.4) is a standard damped Newton method for minimizing the self-concordant function $\Phi(x)$. Therefore, it enters the region of the quadratic convergence after at most $O(\Phi(x_0) - \Phi(x^*))$ iterations. Note that in view of inequality (3.3) we have

$$\Phi(x^*) = \min_x \, [\tfrac{1}{2}\langle Qx, x\rangle + F(x)] \geq \min_x \, [\tfrac{1}{2}\gamma_l(\hat{x})\langle F''(\hat{x})x, x\rangle + F(x)].$$

The solution of this problem is uniquely defined by $\gamma_l(\hat{x})F''(\hat{x})x + F'(x) = 0$. Let $\bar{x} = \hat{x}/\sqrt{\gamma_l(\hat{x})}$. Since F is logarithmically homogeneous, $\gamma_l(\hat{x})F''(\hat{x})\bar{x} = -\sqrt{\gamma_l(\hat{x})}F'(\hat{x})$ and $F'(\bar{x}) = \sqrt{\gamma_l(\hat{x})}F'(\hat{x})$. Hence, $\bar{x}$ is the solution and

$$\Phi^* \geq \tfrac{1}{2}\nu(1 + \ln \gamma_l(\hat{x})) + F(\hat{x}).$$

On the other hand,

$$\begin{aligned} \Phi(x_0) &= \tfrac{1}{2}\rho^2\langle Q\hat{x}, \hat{x}\rangle + F(\rho\hat{x}), \\ &\leq \tfrac{1}{2}\rho^2\gamma_u(\hat{x})\langle F'(\hat{x}), \hat{x}\rangle^2 + F(\hat{x}) - \nu \ln \rho, \\ &= \tfrac{1}{2}\rho^2\gamma_u(\hat{x})\nu^2 + F(\hat{x}) - \nu \ln \rho. \end{aligned}$$

Thus,

$$\Phi(x_0) - \Phi(x^*) \leq \tfrac{1}{2}\nu[\rho^2\gamma_u(\hat{x})\nu - 2\ln\rho - \ln\gamma_l(\hat{x}) - 1]. \tag{3.5}$$

To conclude the proof it suffices to introduce the appropriate values for ρ into (3.5). □

3.2 Path-following scheme

Let us look now at the complexity of a path-following scheme as applied to the problem (3.1). Assume that the scaling coefficient $\rho > 0$ has been fixed to some value (to be made precise later). Set the tolerance parameter $\beta = \frac{1}{9}$ and let Δ be defined as in Theorem 2

$$\Delta = \frac{1}{\beta + \sqrt{\nu}}\left(\frac{\sqrt{\beta}}{1+\sqrt{\beta}} - \beta\right) = \frac{5}{4 + 36\sqrt{\nu}}. \tag{3.6}$$

Denote

$$\psi_t(x) = -t\langle \Phi'(\rho\hat{x}), x\rangle + \Phi(x).$$

Consider the following method:

$$\left.\begin{array}{ll} \text{Initialization:} & t_0 = 1, \quad x_0 = \rho\hat{x}. \\ \text{Iteration } \#k\text{:} & \text{If } \lambda_{x_k}(\Phi'(x_k)) > 2\beta, \text{ then set} \\ & t_{k+1} = (1-\Delta)t_k, \quad x_{k+1} = x_k - [\psi''_{t_k}(x_k)]^{-1}\psi'_{t_{k+1}}(x_k), \\ & \text{else stop.} \end{array}\right\} \tag{3.7}$$

Prior to analyzing the complexity of this scheme, we need one auxiliary result on the local norms defined by a ν-normal barrier.

Lemma 3 *Let $F(x)$ be a ν-normal barrier for the cone K. Then for any x and y from* $\mathrm{int}\, K$ *and any $s \in R^n$ we have:*

$$\theta_y^*(s) \leq -2\langle F'(x), y\rangle \cdot \theta_x^*(s), \tag{3.8}$$

and

$$[F''(y)]^{-1} \preceq 4\langle F'(x), y\rangle^2 [F''(x)]^{-1}. \tag{3.9}$$

Proof:
Since s is arbitrary, it is straightforward to see that (3.8) implies (3.9).

Let us fix an arbitrary x and y in $\mathrm{int}\, K$. Note that $x \in \mathrm{int}\, K$, implies that $F'(x) \in \mathrm{int}\, K_*$, and thus $\langle F'(x), y\rangle < 0$. Let us scale y to

$$\bar{y} = \frac{\nu y}{-\langle F'(x), y\rangle},$$

so that $-\langle F'(x), \bar{y}\rangle = \nu$. Let us fix an arbitrary direction $s \in R^n$. Denote

$$u = [F''(\bar{y})]^{-1} s / \theta_{\bar{y}}^*(s)$$

and consider two points $y_\pm = \bar{y} \pm u$. Both points belong to K, since $\theta_{\bar{y}}(u) = 1$, and at least one of them belongs to the set

$$S = \{v \in K : \; -\langle F'(x), v\rangle \leq \nu\}.$$

Since F is a ν-normal barrier, $\langle F''(x)x, x\rangle = \nu$ and $F''(x)x = F'(x)$; besides, for any $v \in K$, $\langle F''(x)v, v\rangle = \langle F'(x), v\rangle^2$. Therefore, if we take any $v \in S \subset K$, we have

$$\begin{aligned}
(\theta_x(v - x))^2 &= \langle F''(x)(v - x), v - x\rangle, \\
&= \langle F''(x)v, v\rangle + 2\langle F'(x), v\rangle + \nu, \\
&\leq \langle F'(x), v\rangle^2 + 2\langle F'(x), v\rangle + \nu.
\end{aligned}$$

Recall that for a ν-normal barrier $-\langle F'(x), v\rangle \geq 0$, while, by definition of S, $-\langle F'(x), v\rangle \leq -\nu$. Hence, the quadratic form $t^2 + 2t + \nu$, with $-\nu \leq t \leq 0$ achieves its maximum at the end point of the interval. Thus, we may conclude

$$(\theta_x(v - x))^2 \leq \langle F'(x), v\rangle^2 + 2\langle F'(x), v\rangle + \nu \leq \nu^2. \tag{3.10}$$

Now, using (3.10) and Cauchy-Schwarz inequality, we have for any $v \in S$ and $g \in R^n$

$$\langle g, v\rangle = \langle g, x\rangle + \langle g, v - x\rangle \leq \langle g, x\rangle + \nu\theta_x^*(g). \tag{3.11}$$

If $y_+ \in S$, let us take $v = y_+$ and $g = s$ in (3.11). Then,

$$\langle s, \bar{y}\rangle + \theta_{\bar{y}}^*(s) \leq \langle s, x\rangle + \nu\theta_x^*(s). \tag{3.12}$$

Now, if $y_- \in S$, let us take $v = y_-$ and $g = -s$ in (3.11). Then,

$$-\langle s, \bar{y}\rangle + \theta_{\bar{y}}^*(s) \leq -\langle s, x\rangle + \nu\theta_x^*(s). \tag{3.13}$$

Together, the inequalities (3.12) & (3.13) are equivalent to

$$\theta^*_{\bar{y}}(s) \leq |\langle s, x - \bar{y}\rangle| + \nu\theta^*_x(s) \leq 2\nu\theta^*_x(s).$$

It remains to note that in view of homogeneity of the barrier $F(x)$ we have

$$(\theta^*_{\bar{y}}(s))^2 = \langle [F''(\bar{y})]^{-1}s, s\rangle = \frac{\nu^2}{\langle F'(x), y\rangle^2}\langle [F''(y)]^{-1}s, s\rangle = \frac{\nu^2}{\langle F'(x), y\rangle^2}(\theta^*_y(s))^2.$$

□

Now we can give a complexity estimate for (3.7).

Theorem 4 *The scheme (3.7) terminates no more than after N iterations, where N satisfies the following inequality:*

$$(1-\Delta)^N\nu^2\left[\rho^2\gamma_u(\hat{x}) + 8\frac{\gamma_u(\hat{x})\nu^2}{\gamma_l(\hat{x})} + \frac{16}{\rho^4\gamma_l^2(\hat{x})}\left(4\nu + \frac{\rho^2}{16}\gamma_l(\hat{x})\right)\right] \leq \beta.$$

Proof:
Note that $\psi'_{t_0}(x_0) = 0$, and thus

$$\theta^*_{x_0}(\psi'_{t_0}(x_0)) = \sigma^*_{x_0}(\psi'_{t_0}(x_0)) = 0 \leq \beta.$$

Hence, we can apply the path-following scheme (3.7), starting from x_0. In view of Theorem 2 and relation (2.2), for all $k \geq 0$, we have

$$\beta \geq \theta^*_{x_k}(\psi'_{t_k}(x_k)) \geq \lambda^*_{x_k}(\psi'_{t_k}(x_k)) = \lambda^*_{x_k}(-t_k\Phi'(x_0) + \Phi'(x_k)). \quad (3.14)$$

Thus,

$$\lambda^*_{x_k}(\Phi'(x_k)) \leq \beta + t_k\lambda^*_{x_k}(\Phi'(x_0)). \quad (3.15)$$

We shall try to bound the right-hand side of (3.15). To this end, we introduce the quantity $\omega_k = -\langle F'(x_0), x_k\rangle$. In view of Lemma 3,

$$[F''(x_k)]^{-1} \preceq 4\omega_k^2[F''(x_0)]^{-1}.$$

Equivalently, $F''(x_k) \succeq \frac{1}{4\omega_k^2}F''(x_0)$. Therefore,

$$\begin{aligned}(\lambda^*_{x_k}(\Phi'(x_0)))^2 &= \langle [Q + F''(x_k)]^{-1}\Phi'(x_0), \Phi'(x_0)\rangle,\\ &\leq \langle [Q + \tfrac{1}{4\omega_k^2}F''(x_0)]^{-1}\Phi'(x_0), \Phi'(x_0)\rangle,\\ &= \langle [Q + \tfrac{1}{4\omega_k^2}F''(x_0)]^{-1}(Q - F''(x_0))x_0, (Q - F''(x_0))x_0\rangle.\end{aligned}$$

Using the same argument as in Lemma 1, we check that

$$(Q - F''(x_0))[Q + \frac{1}{4\omega_k^2}F''(x_0)]^{-1}(Q - F''(x_0)) \preceq (Q + 4\omega_k^2F''(x_0)).$$

Thus,

$$(\lambda^*_{x_k}(\Phi'(x_0)))^2 \le \langle (Q + 4\omega_k^2 F''(x_0))x_0, x_0\rangle \le \langle Qx_0, x_0\rangle + 4\nu\omega_k^2. \tag{3.16}$$

In view of (3.15) this implies that the algorithm terminates as soon as

$$t_k[\langle Qx_0, x_0\rangle + 4\nu\omega_k^2] \le \beta. \tag{3.17}$$

Note that $t_k = (1-\Delta)^k$ and

$$\langle Qx_0, x_0\rangle = \rho^2\langle Q\hat{x}, \hat{x}\rangle \le \rho^2\gamma_u(\hat{x})\langle F'(\hat{x}), \hat{x}\rangle^2 = \rho^2\gamma_u(\hat{x})\nu^2. \tag{3.18}$$

Therefore, in order to justify our complexity estimate, we need only to find a uniform upper bound for the special quantity ω_k. From (3.17) it is clear that the bound on the number of iterations will involve the quantity ω_k only under a logarithm. Consequently, it is not necessary to look for a sharp bound on ω_k.

Since x_k is close to the central path, we shall derive the bound on ω_k from a closer look at that path. Let $x(t)$ denote the central path trajectory for problem (3.1). By definition, $\psi_t'(x(t)) = 0$; thus,

$$\Phi'(x(t)) = Qx(t) + F'(x(t)) = t\Phi'(x_0), \quad 0 \le t \le 1.$$

Since $\Phi(x)$ is a convex function, we have

$$\begin{aligned}
\Phi(x_0) &\ge \Phi(x(t)) + \langle \Phi'(x(t)), x_0 - x(t)\rangle,\\
&= \Phi(x(t)) + t\langle \Phi'(x_0), x_0 - x(t)\rangle,\\
&\ge \Phi(x(t)) + t(\Phi(x_0) - \Phi(x(t))).
\end{aligned}$$

Thus, $\Phi(x(t)) \le \Phi(x_0)$, for all $0 \le t \le 1$. Recall that by (3.14)

$$\lambda^*_{x_k}(\psi'_{t_k}(x_k)) \le \beta.$$

Since the function $\psi_t(x)$ is self-concordant, we have by inequality 4.1.15 in [2]

$$\psi_{t_k}(x_k) \le \psi_{t_k}(x(t_k)) + \omega_*(\beta),$$

where $\omega_*(t) = -t - \ln(1-t)$. Note that

$$\begin{aligned}
\Phi(x_k) &= \psi_{t_k}(x_k) + t_k\langle \Phi'(x_0), x_k\rangle,\\
&\le \psi_{t_k}(x(t_k)) + t_k\langle \Phi'(x_0), x_k\rangle + \omega_*(\beta),\\
&= \Phi(x(t_k)) + t_k\langle \Phi'(x_0), x_k - x(t_k)\rangle + \omega_*(\beta),\\
&= \Phi(x(t_k)) + \langle \Phi'(x(t_k)), x_k - x(t_k)\rangle + \omega_*(\beta), && \text{(from the central path)}\\
&\le \Phi(x(t_k)) + \langle \Phi'(x_k), x_k - x(t_k)\rangle + \omega_*(\beta), && \text{(by monotony of } \Phi'\text{)}\\
&\le \Phi(x_0) + \lambda^*_{x_k}(\Phi'(x_k)) \cdot \lambda_{x_k}(x_k - x(t_k)) + \omega_*(\beta). && \text{(by Cauchy-Schwarz)}
\end{aligned}$$

Moreover, in view of inequality (4.1.16) in [2], we have

$$\lambda_{x_k}(x_k - x(t_k)) \le \frac{\beta}{1-\beta}.$$

Using the last two inequalities and inequality (2.4) we get

$$\tfrac{1}{2}\langle Qx_k, x_k\rangle + F(x_k) \le \tfrac{1}{2}\langle Qx_0, x_0\rangle + F(x_0) + \frac{\beta}{1-\beta}\left[\langle Qx_k, x_k\rangle + \nu\right]^{1/2} + \omega_*(\beta). \quad (3.19)$$

Consider the function $f(u) = \frac{1}{2}u - \beta/(1-\beta)\sqrt{u+\nu}$, with $u = \langle Qx_k, x_k\rangle$. With our choice of $\beta(=1/9)$, the function $f(u)$ is increasing. We can therefore replace $\langle Qx_k, x_k\rangle$ in (3.19) by any appropriate lower bound. Precisely, by definition of $\gamma_l(\hat{x})$ and the basic property of self concordant barriers, one may write

$$\langle Qx_k, x_k\rangle \ge \gamma_l(\hat{x})\langle F''(\hat{x})x_k, x_k\rangle \ge \tfrac{\gamma_l(\hat{x})}{\nu}\langle F'(\hat{x}), x_k\rangle^2 = \tfrac{1}{\nu}\rho^2\gamma_l(\hat{x})\omega_k^2.$$

Also, by definition of ω_k,

$$F(x_k) \ge F(x_0) + \langle F'(x_0), x_k - x_0\rangle = F(x_0) + \nu - \omega_k.$$

Using these bounds in (3.19), we obtain:

$$\begin{aligned}\tfrac{1}{2\nu}\rho^2\gamma_l(\hat{x})\omega_k^2 - \omega_k &\le \tfrac{1}{2}\langle Qx_0, x_0\rangle + \tfrac{\beta}{1-\beta}\left[\tfrac{1}{\nu}\rho^2\gamma_l(\hat{x})\omega_k^2 + \nu\right]^{1/2} + \omega_*(\beta) - \nu, \\ &\le \tfrac{1}{2}\langle Qx_0, x_0\rangle + \tfrac{\beta}{1-\beta}\sqrt{\tfrac{\gamma_l(\hat{x})}{\nu}}\rho\omega_k + \tfrac{\beta}{1-\beta}\sqrt{\nu} + \omega_*(\beta) - \nu.\end{aligned}$$

Since $\beta = 1/9$, one easily checks that $\frac{\beta}{1-\beta}\sqrt{\nu} + \omega_*(\beta) - \nu < 0$. Thus, using the bound (3.18) on $\langle Qx_0, x_0\rangle$, and replacing $\frac{\beta}{1-\beta} = \frac{1}{8}$ by its value, we get the following quadratic inequality:

$$a\omega_k^2 \le b + c\omega_k,$$

with $a = \frac{1}{2\nu}\rho^2\gamma_l(\hat{x})$, $b = \frac{1}{2}\rho^2\gamma_u(\hat{x})\nu^2$ and $c = \left(1 + \frac{\rho}{8}\sqrt{\frac{\gamma_l(\hat{x})}{\nu}}\right)$.

From the above inequality, we derive the trivial bound

$$\omega_k \le \sqrt{\frac{b}{a}} + \frac{c}{a},$$

i.e.,

$$\omega_k \le \sqrt{\frac{\gamma_u(\hat{x})\nu^3}{\gamma_l(\hat{x})}} + \frac{1}{\rho^2\gamma_l(\hat{x})}\left(2\nu + \frac{\rho}{4}\sqrt{\gamma_l(\hat{x})\nu}\right).$$

Finally, squaring the above inequality and using the obvious inequality $(u+v)^2 \le 2(u^2+v^2)$ twice, we get

$$\omega_k^2 \le 2\frac{\gamma_u(\hat{x})\nu^3}{\gamma_l(\hat{x})} + \frac{4\nu}{\rho^4\gamma_l^2(\hat{x})}\left(4\nu + \frac{\rho^2}{16}\gamma_l(\hat{x})\right).$$

Inserting the bound on ω_k^2 into (3.17), we conclude that the algorithm terminates as soon as

$$(1-\Delta)^k \nu^2 \left[\rho^2 \gamma_u(\hat{x}) + 8\frac{\gamma_u(\hat{x})\nu^2}{\gamma_l(\hat{x})} + \frac{4}{\rho^4 \gamma_l^2(\hat{x})} \left(16\nu + \frac{\rho^2}{4}\gamma_l(\hat{x}) \right) \right] \leq \beta.$$

□

Theorem 4 implies the following complexity estimates on the path-following scheme. If we can choose $\rho = 1/\sqrt{\gamma_u(\hat{x})}$, then the number of iterations of that scheme is of the order of

$$O(\sqrt{\nu}\ln \nu) + O(\ln \sqrt{\nu}\kappa(\hat{x})),$$

where $\kappa(\hat{x}) = \gamma_u(\hat{x})/\gamma_l(\hat{x})$ is a kind of condition number for problem (3.1) with starting point $\hat{x}$.

If $\gamma_u(\hat{x})$ cannot be estimated beforehand, we can still choose ρ as an absolute constant. The number of iterations of the path-following scheme remains comparable and still quite moderate:

$$O(\sqrt{\nu}\ln \nu) + O(\sqrt{\nu}\ln \gamma_u(\hat{x})) + O(\sqrt{\nu}\ln \frac{1}{\gamma_l(\hat{x})}).$$

This result basically improves Theorem 3 by a factor $\sqrt{\nu}$.

4 Constrained analytic center

Let us look now at the problem of finding a constrained analytic center of the augmented barrier:

$$\phi^* = \min_x \left[\tfrac{1}{2}\langle Qx, x\rangle + F(x) : \ Ax = 0 \right], \tag{4.1}$$

where Q is a symmetric positive definite matrix and the matrix $A \in R^{m\times n}$ is non-degenerate ($m < n$).

Of course, this problem can be solved in two stages. The first stage aims to find a point $\bar{x}$ in the interior of the cone

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

The second stage computes the analytic center of the augmented barrier restricted to the subspace $\{x : \ Ax = 0\}$, by an interior point scheme starting form $\bar{x}$. However, it turns out that problem (4.1) has a dual that can be solved in a single stage and yields a solution to the primal (4.1). (Problem (4.1) is studied in [1] in the framework of the positive orthant.)

Denote by Q_A^* the following matrix:

$$Q_A^* = Q^{-1} - Q^{-1}A^T \left[AQ^{-1}A^T \right]^{-1} AQ^{-1}.$$

Since $m < n$, then Q_A^* is a degenerate matrix.

Theorem 5 *Assume that problem (4.1) is strictly feasible. Consider the associated problem*

$$\phi_* = \min_s \left[\tfrac{1}{2}\langle s, Q_A^* s\rangle + F_*(s) \right]. \tag{4.2}$$

Then,
1) Problems (4.1) and (4.2) are both solvable. They are dual to one another with a zero duality gap: $\phi^* + \phi_* = 0$*.*
2) The solutions x^* *and* s^* *are unique and satisfy the following relations:*

$$x^* = -F_*'(s^*), \quad s^* = -F'(x^*), \tag{4.3}$$

$$Q_A^* s^* = x^*, \tag{4.4}$$

$$\langle Qx^*, x^*\rangle = \langle s^*, Q_A^* s^*\rangle = \nu. \tag{4.5}$$

Proof:
Let us show first that both problems achieve their minimum value and that $\phi^* + \phi_* \geq 0$. Since $\operatorname{int}\mathcal{F} \neq \emptyset$ and $Q \succ 0$, the solution x^* of the primal problem exists. Using the inequality $F(x) + F_*(s) \geq -\langle s, x\rangle$, we have

$$\begin{aligned}
\phi^* &= \min[\tfrac{1}{2}\langle Qx, x\rangle + F(x) : \ Ax = 0], \\
&\geq -F_*(s) + \min_{x:Ax=0} [\tfrac{1}{2}\langle Qx, x\rangle - \langle s, x\rangle], \\
&= [-\tfrac{1}{2}\langle s, Q_A^* s\rangle - F_*(s)].
\end{aligned}$$

This shows that the function

$$\Phi_*(s) = \tfrac{1}{2}\langle s, Q_A^* s\rangle + F_*(s)$$

is bounded on K from below. Since $\Phi_*(s)$ is a self-concordant function, its minimum s^* exists. Finally, since K^* is a proper cone, the Hessian $F_*(s)$ is non-degenerate at any $s \in \operatorname{int} K^*$. Thus, s^* is a unique solution to (4.2). By taking the maximum over $s \in \operatorname{int} K^*$, we obtain $\phi^* + \phi_* \geq 0$.

Let us show now that the duality gap is zero. In view of the first order optimality conditions for problem (4.2), we have

$$-F_*'(s^*) = Q_A^* s^*. \tag{4.6}$$

Since $F_*(s)$ is logarithmically homogeneous,

$$\langle s^*, Q_A^* s^*\rangle = -\langle s^*, F_*'(s^*)\rangle = \nu.$$

Denote $\bar{x} = -F_*'(s^*)$. Then $F'(\bar{x}) = -s^*$. Since $AQ_A^* = 0$, then (4.6) implies $A\bar{x} = 0$. Besides, we have

$$\begin{aligned}
\Phi'(\bar{x}) &= Q\bar{x} + F'(\bar{x}) = -QF_*'(s^*) - s^*, \\
&= QQ_A^* s^* - s^* = -A^T \left[AQ^{-1}A^T\right]^{-1} AQ^{-1} s^*.
\end{aligned} \tag{4.7}$$

In other words, $\bar{x}$ solves the first order optimality conditions of problem (4.1), and thus $x^* \equiv \bar{x}$.

Using (4.7), we also have

$$\langle Qx^*, x^* \rangle + \langle F'(x^*), x^* \rangle = 0.$$

Since F is logarithmically homogeneous, we conclude

$$\langle Qx^*, x^* \rangle = -\langle F'(x^*), x^* \rangle = \nu.$$

Therefore

$$\begin{aligned} \phi^* &= \tfrac{1}{2}\langle Qx^*, x^* \rangle + F(x^*) = \tfrac{1}{2}\nu + F(-F_*'(s^*)), \\ &= -\tfrac{1}{2}\nu - F_*(s^*) = -\tfrac{1}{2}\langle Q_A^* s^*, s^* \rangle - F_*(s^*) = -\phi_*. \end{aligned}$$

□

Theorem 5 shows that the solution of the linearly constrained minimization problem (4.1) can be obtained by solving problem (4.2), which is of the type studied in the previous section. Clearly, one can apply a damped Newton Method, or a path-following scheme, starting from any arbitrary point in $\operatorname{int} K^*$.

Note that the problem (4.2) has *no linear constraints*. All necessary information on the matrix A is incorporated in the matrix Q_A^*. Thus, this approach can be seen as *an infeasible start dual technique* for solving the constrained minimization problem (4.1).

5 Applications

In this section we consider some applications of the results of Sections 3 and 4. Theorem 4 tells us that the complexity of the path-following scheme depends on the choice of the $\hat{x}$ and on the characteristics $\gamma_u(\hat{x})$ and $\gamma_l(\hat{x})$. In particular, if one can estimate $\gamma_u(\hat{x})$, it is possible to set the scaling factor at the value $\rho = 1/\sqrt{\gamma_u(\hat{x}))}$ in the path-following scheme (3.7) and get the best complexity estimate. In the applications we review, we shall discuss these issues and propose ways to compute those critical entities. In particular, we shall see that $\gamma_u(\hat{x})$ is easily computable in all cases.

5.1 Conic feasibility problem

Assume we need to solve the following *conic feasibility problem*:

$$\text{Find } x^* \in \mathcal{F} = \{x \in K : \ Ax = 0\}. \tag{5.1}$$

To apply the techniques previously devised, we choose a point $\hat{x} \in \operatorname{int} K$, and we transform the feasibility problem (5.1) into the auxiliary optimization problem

$$\text{Find } x^* = \arg\min_x \, [F(x) : \ Ax = 0, \ \langle F''(\hat{x})x, x \rangle \le \nu \,]. \tag{5.2}$$

Using the first order optimality conditions and the fact that F is logarithmically homogeneous, we can readily verify that (5.2) is equivalent to

$$\min_x \, [\, \tfrac{1}{2}\langle F''(\hat{x})x, x \rangle + F(x) : \ Ax = 0 \,]. \tag{5.3}$$

Let us write down this problem in the dual form. Denote $\hat{s} = -F'(\hat{x})$. Then

$$F''_*(\hat{s}) = [F''(\hat{x})]^{-1}.$$

Denote

$$H_*(\hat{x}) = F''_*(\hat{s}) - F''_*(\hat{s})A^T \left[AF''_*(\hat{s})A^T\right]^{-1} AF''_*(\hat{s}).$$

Thus, in accordance with Theorem 5, the problem dual to (5.3) is as follows:

$$\min_s \; [\; \tfrac{1}{2}\langle s, H_*(\hat{s})s\rangle + F_*(s) \;]. \tag{5.4}$$

Let us estimate the complexity of this problem. To this end, we introduce the following characteristics:

$$\sigma(\hat{x}) = \max_\alpha \; [\; \alpha : \; x^* - \alpha\hat{x} \in K \;].$$

In view of this definition, $x^* - \sigma(\hat{x})\hat{x} \in K$.

Theorem 6 *The solution s^* of the problem (5.4) can be found by the path-following scheme (3.7) in*

$$O\left(\sqrt{\nu} \ln \frac{\nu}{\sigma(\hat{x})}\right)$$

iterations.

Proof:
In order to use the result of Theorem 4, we need to find bounds $\gamma_u(\hat{s})$ and $\gamma_l(\hat{s})$. Note that for any $s \in K^*$ we have:

$$\langle s, H_*(\hat{s})s\rangle \leq \langle s, F''_*(\hat{s})s\rangle \leq \langle s, F'_*(\hat{s})\rangle^2.$$

Therefore, we can choose $\gamma_u(\hat{s}) = 1$. On the other hand, in view of Theorem 5, for $s^* = -F'(x^*)$ we have:

$$H_*(\hat{s})s^* - \sigma(\hat{x})\hat{x} = x^* - \sigma(\hat{x})\hat{x} \in K.$$

Therefore, for any $s \in K^*$ we get the following relation:

$$\sigma(\hat{x})\langle s, \hat{x}\rangle \leq \langle s, H_*(\hat{s})s^*\rangle.$$

Hence, since $\hat{x} = -F'_*(\hat{s})$, in view of (4.5) we obtain:

$$\begin{aligned}\sigma^2(\hat{x})\langle s, F''_*(\hat{s})s\rangle \; &\leq \sigma^2(\hat{x})\langle s, F'_*(\hat{s})\rangle^2 = \sigma^2(\hat{x})\langle s, \hat{x}\rangle^2 \\ &\leq \langle s, H_*(\hat{s})s^*\rangle^2 \leq \langle s^*, H_*(\hat{s})s^*\rangle \cdot \langle s, H_*(\hat{s})s\rangle = \nu \cdot \langle s, H_*(\hat{s})s\rangle.\end{aligned}$$

Thus, we can take $\gamma_l(\hat{s}) = \frac{1}{\nu}\sigma^2(\hat{x})$. □

Note that the path-following scheme for the problem (5.4) can be seen as an infeasible start methods employing the dual technique for solving the initial feasibility problem (5.1).

5.2 Affine feasibility problem

Consider the feasibility problem

$$\text{Find } x^* \in \mathcal{F} = \{x \in K : \ Ax = b\}. \tag{5.5}$$

This problem can be homogenized in a standard way:

$$\text{Find } (x^*, \tau^*) \in \widehat{\mathcal{F}} = \{(x, \tau) \in K \times R_+ : \ Ax = \tau b\}. \tag{5.6}$$

Note that if the problem (5.5) has an interior solution x^*, then the problem (5.6) has an interior solution $(x^*, 1)$. Thus, we can treat the problem (5.6) using the dual infeasible start scheme. The complexity of finding the solution is

$$O\left(\sqrt{\nu} \ln \frac{\nu}{\sigma(\hat{x}, 1)}\right)$$

iterations. Note that our approach is well defined even for unbounded $\mathcal{F}$.

5.3 Primal-dual conic problem

Consider the following primal-dual conic problem:

$$\left\{ \begin{array}{c} \min\limits_{x} \langle c, x \rangle \\ \\ \text{s.t. } \ Ax = b \\ \\ x \in K \end{array} \right\} = \left\{ \begin{array}{c} \max\limits_{s,y} \langle b, y \rangle \\ \\ \text{s.t. } \ s + A^T y = c \\ \\ s \in K^*, \ y \in R^m \end{array} \right\}. \tag{5.7}$$

Let us choose some accuracy level $\epsilon > 0$ for our approximate solution of the problem (5.7). Then we can reformulate this problem as the following conic feasibility problem:

$$\begin{array}{rl} & \text{Find } (x^*, s^*, \tau^*) \in \widetilde{\mathcal{F}}, \text{ where} \\ \widetilde{\mathcal{F}} = & \{(x, s, \tau) \in K \times K^* \times R_+ : \ \ Ax = \tau b, \ Bs = \tau Bc, \\ & \qquad\qquad\qquad \langle c, x \rangle + \langle s - c, x_0 \rangle = \tau \epsilon \ \}, \end{array} \tag{5.8}$$

where B is a matrix complementary to A, $BA^T = 0$, and x_0 is an arbitrary point such that $Ax_0 = b$. If the initial primal-dual problem is strictly feasible, then the problem (5.8) also has an interior solution.

Thus, the solution to (5.8) can be found by the dual infeasible start scheme in

$$O\left(\sqrt{\nu} \ln \frac{\nu}{\sigma}\right)$$

iterations, where the characteristics σ depends on ϵ and the deepness of the interior of the feasible set in (5.7). Note that this scheme has only one stage. Recall, that the standard infeasible start primal dual schemes provide the same efficiency estimate. But their justification is crucially based on the skew-symmetry of the primal-dual system of linear equations. In our case we derive this efficiency bound from a completely general complexity result.

5.4 Scaling a positive semi-definite matrix

In this problem K is a positive orthant. Thus,

$$K = \{x \in R^n : \ x \geq 0\}, \quad F(x) = -\sum_{i=1}^{n} \ln x^{(i)}, \quad \nu = n.$$

The problem we need to solve is as follows:

$$\min_{x} \ \tfrac{1}{2}\langle Qx, x\rangle - \sum_{i=1}^{n} \ln x^{(i)}, \tag{5.9}$$

where Q is a positive semi-definite matrix. We assume that $\langle Qx, x\rangle > 0$ for all $x \in K$, $x \neq 0$. Hence, problem (5.9) is bounded and achieves its minimum value.

Let us choose

$$\hat{x} = e = (1, \dots, 1)^T \in R^n.$$

Then $F''(\hat{x}) = I$ and we can take in (3.3)

$$\gamma_l(\hat{u}) = \lambda_{\min}(Q).$$

On the other hand, since Q is positive semidefinite, for any $x \geq 0$ we have

$$\langle Qx, x\rangle \leq \max_{1\leq i\leq n} Q_{ii} \ \cdot \left(\sum_{i=1}^{n} x^{(i)}\right)^2.$$

Since $-F'(\hat{x}) = e$, we have $-\langle F'(\hat{x}), x\rangle = \sum\limits_{i=1}^{n} x^{(i)}$. So, we can take

$$\gamma_u(\hat{x}) = \max_{1\leq i\leq n} Q_{ii}.$$

Note that this value is useful for getting a good starting point for the path-following scheme (3.7). Note also that

$$\kappa(\hat{x}) = \frac{\gamma_u(\hat{x})}{\gamma_l(\hat{x})} \leq \mu(Q) \equiv \frac{\lambda_{\max}(Q)}{\lambda_{\min}(Q)}.$$

Thus, the complexity of the problem (5.9) does not exceed

$$O\left(\sqrt{n}\,(\ln n + \ln \mu(Q))\right)$$

iterations of the path-following scheme (3.7).

If $\lambda_{\min}(Q)$ is small or equal to zero, we can regularize the matrix Q. Indeed, we can form the matrix

$$Q_+ = Q + \mu \cdot \left(\sum_{i=1}^{n} Q_{ii}^{1/2} x^{(i)}\right)^2.$$

Note that the assumption that Q is positive definite on the nonnegative orthant implies that $Q_{ii} > 0$. Let us choose $\hat{x}^{(i)} = 1/Q_{ii}^{1/2}$, $i = 1, \dots, n$. Then the gradient $F'(\hat{x})$ has the following components:

$$\left(F'(\hat{x})\right)^{(i)} = -Q_{ii}^{1/2}, \quad i = 1, \dots, n.$$

Hence, since Q is a positive-semidefinite matrix, we have

$$(1+\mu)\langle F'(\hat{x}), x\rangle^2 \geq \langle Q_+ x, x\rangle \geq \mu\langle F'(\hat{x}), x\rangle^2 \geq \frac{\mu}{n}\langle F''(\hat{x})x, x\rangle.$$

Thus, we can take $\gamma_u(\hat{x}) = 1+\mu$ and $\gamma_l(\hat{x}) = \frac{\mu}{n}$. Therefore the complexity of the regularized problem does not exceed

$$O\left(\sqrt{n}\ln\frac{n}{\mu}\right) \tag{5.10}$$

iterations of (3.7). Note that this complexity is *constant*; it does not depends anyhow on a particular matrix Q.

5.5 Scaling a positive semi-definite matrix with nonnegative coefficients

In this problem K is again a positive orthant and the problem we solve is as follows:

$$\min_x \ \tfrac{1}{2}\langle Qx, x\rangle - \sum_{i=1}^n \ln x^{(i)}, \tag{5.11}$$

where Q is a positive semi-definite matrix with *non-negative elements*. Note that in this case for any $x \geq 0$ we have

$$\sum_{i=1}^n Q_{ii}\left(x^{(i)}\right)^2 \leq \langle Qx, x\rangle \leq \left(\sum_{i=1}^n Q_{ii}^{1/2} x^{(i)}\right)^2.$$

This means that we can take in (3.3)

$$\hat{x}^{(i)} = Q_{ii}^{-1/2}, \quad i = 1,\ldots,n,$$
$$\gamma_l(\hat{x}) = \gamma_u(\hat{x}) = 1.$$

Thus, the complexity of the problem (5.11) is *constant*:

$$O(\sqrt{n}\ln n) \tag{5.12}$$

iterations of path-following scheme (3.7).

Note that sometimes it is possible to get the problem (5.11) as a dual to our initial problem. Indeed, assume that in the matrix scaling problem (5.9) the matrix Q has the following form:

$$Q = D - B,$$

where D is a diagonal matrix and B is a matrix with non-negative elements. Then, if Q is non-degenerate, one can show using the Perron-Frobenius theorem that the inverse matrix Q^{-1} has non-negative elements. Therefore, in view of Lemma 2, we can solve our problem with the complexity (5.12).

5.6 Auxiliary problem in cutting plane schemes

In some cutting plane schemes for nonsmooth optimization at each iteration we need to solve the following problem (see [1]):

$$\min_x \{-\sum_{i=1}^{p} \ln\langle g_k, x\rangle : \langle Bx, x\rangle \le 1\}, \tag{5.13}$$

where B is a symmetric positive definite matrix and $g_k \in R^n$ are some cutting planes. The natural assumption on this problem is that the matrix

$$G = (g_1, \ldots, g_p)$$

is non-degenerate. Note that the problem (5.13) can be written also in the following form:

$$\min_x \frac{p}{2}\langle Bx, x\rangle - \sum_{i=1}^{p} \ln\langle g_k, x\rangle.$$

Now, introducing the slack variables $y \in R^p$ we can rewrite this problem as follows:

$$\begin{aligned} &\min_{x,y} \left[\frac{p}{2}\langle Bx, x\rangle - \sum_{i=1}^{p} \ln y^{(i)} : G^T x = y, \ y \ge 0 \right] \\ &\qquad = \min_y \left[\frac{p}{2}\langle Qy, y\rangle - \sum_{i=1}^{p} \ln y^{(i)} : y \ge 0 \right], \end{aligned}$$

where $Q = p[G^T B^{-1} G]^{-1}$.

Thus, the problem (5.13) is equivalent to a matrix scaling problem. Note that the elements of the matrix Q^{-1} are related to the angles between the cutting planes g_k, computed with respect to the metric $\langle B(\cdot), (\cdot)\rangle^{1/2}$. Thus, in practical implementations of these schemes we have different possibilities. We can keep all cuts, but then we need to solve at each iteration the general problem (5.9). Or, we keep only cuts with all positive (obtuse) angles. Or, we keep the cuts with only negative (acute) angles. With the last two strategies, the complexity of the auxiliary problem will be constant (see (5.12)).

5.7 Semidefinite scaling problem

Let K be a cone of symmetric positive semidefinite matrices:

$$K = \{X \in R^{n(n+1)/2} : X \succeq 0\}.$$

Then we can take

$$F(X) = -\ln\det X, \quad \nu = n.$$

Note that the gradient and the Hessian of the barrier $F(X)$ are defined as follows:

$$\begin{aligned} \langle F'(X), H\rangle &= -\langle X^{-1}, H\rangle, \\ \langle F''(X)H, H\rangle &= \langle X^{-1}HX^{-1}, H\rangle, \end{aligned} \tag{5.14}$$

where $X \succ 0$, $H \in R^{n(n+1)/2}$ and notation $\langle \cdot, \cdot \rangle$ stands for the scalar product of two matrices:

$$\langle X, Y \rangle = \mathrm{Trace}\, XY^T.$$

Similarly to the problem (5.11) we can look at the matrix scaling problem over the cone of positive semidefinite matrices. However, for this cone we can have two types of problems.

The problem of the first type is as follows:

$$\min_X \sum_{k=1}^{p} \langle A_k X A_k, X \rangle - \ln \det X, \tag{5.15}$$

where A_k are symmetric positive semidefinite matrices. We need the following non-degeneracy assumption.

Assumption 1 *The matrix*

$$\hat{Y} = \sum_{k=1}^{p} A_k \tag{5.16}$$

is positive definite.

Consider the following function of symmetric matrices X and Y:

$$\phi(X, Y) = \langle YXY, X \rangle.$$

Note that $\phi(X, Y) = \phi(Y, X)$. Moreover, if $\bar{Y} \succeq 0$ then the function $\bar{\phi}(X)$ is quadratic and non-negative on $R^{n(n+1)/2}$. Thus, $\bar{\phi}(X)$ is a convex quadratic form. This implies, in particular, that the problem (5.15) is convex.

In order to find a reasonable $\hat{X}$ for the problem (5.15) we need the following result.

Lemma 4 *Let A_k, $k = 1, \ldots, p$, are symmetric positive semidefinite matrices and let the matrix $\hat{Y}$ is given by (5.16). Then for any $X \succeq 0$ we have*

$$\frac{1}{p} \langle \hat{Y} X \hat{Y}, X \rangle \leq \sum_{k=1}^{p} \langle A_k X A_k, X \rangle \leq \langle \hat{Y}, X \rangle^2. \tag{5.17}$$

Proof:
Let us prove the first inequality. Let us fix an arbitrary $X \succeq 0$ and consider the function $\bar{\phi}(Y) = \phi(Y, X)$. This function is convex in Y. Therefore

$$\frac{1}{p^2} \langle \hat{Y} X \hat{Y}, X \rangle = \bar{\phi}\left(\frac{1}{p} \hat{Y} \right) \leq \frac{1}{p} \sum_{k=1}^{p} \bar{\phi}(A_k) = \frac{1}{p} \sum_{k=1}^{p} \langle A_k X A_k, X \rangle.$$

In order to prove the second inequality, it is enough to show that

$$\langle AXA, X \rangle \leq \langle A, X \rangle^2 \tag{5.18}$$

for any positive semidefinite X and A. Indeed, since X is positive semidefinite, it can be represented as follows:

$$X = \sum_{i=1}^{n} x_i x_i^T, \quad x_i \in R^n, \ i = 1, \ldots, n.$$

Therefore, since A is positive semidefinite, we have

$$\begin{aligned} \langle AXA, X\rangle &= \langle A \sum_{i=1}^n x_i x_i^T A, \sum_{j=1}^n x_j x_j^T\rangle = \sum_{i=1}^n \sum_{j=1}^n \langle A x_i x_i^T A, x_j x_j^T\rangle = \sum_{i=1}^n \sum_{j=1}^n \langle A x_i, x_j\rangle^2 \\ &\leq \sum_{i=1}^n \sum_{j=1}^n \langle A x_i, x_i\rangle \cdot \langle A x_j, x_j\rangle = \left(\sum_{i=1}^n \langle A x_i, x_i\rangle\right)^2 = \langle A, X\rangle^2. \end{aligned}$$

□

In view of Lemma 4 and expressions (5.14) we can choose

$$\hat{X} = \hat{Y}^{-1}, \quad \gamma_u(\hat{X}) = 1, \quad \gamma_l(\hat{X}) = \frac{1}{p}.$$

Then, $\kappa(\hat{X}) = p$. Since X belongs to the space of dimension $n(n+1)/2$, we can always ensure that p in the problem (5.15) does not exceed that dimension. Thus, we need at most

$$O(\sqrt{n} \ln n) \tag{5.19}$$

iterations of the path-following scheme (3.7). So again, the complexity of the problem (5.15) is *polynomial* and *constant*.

The semidefinite matrix scaling problem of the second type can be written as follows:

$$\min_X \sum_{k=1}^p \langle A_k, X\rangle^2 - \ln\det X. \tag{5.20}$$

We assume that this problem satisfies our assumptions on problem (5.15). Then, since all A_k are positive semidefinite, for any $X \succeq 0$ we have

$$\sum_{k=1}^p \langle A_k, X\rangle^2 \leq \left(\sum_{k=1}^p \langle A_k, X\rangle\right)^2 = \langle \hat{Y}, X\rangle^2.$$

On the other hand, in view of convexity of the function $\langle A, X\rangle^2$ and inequality (5.17) we get

$$\sum_{k=1}^p \langle A_k, X\rangle^2 \geq \frac{1}{p}\langle \hat{Y}, X\rangle^2 \geq \frac{1}{p^2}\langle \hat{Y} X \hat{Y}, X\rangle.$$

Thus, again we can take $\hat{X} = \hat{Y}^{-1}$. Then

$$\gamma_u(\hat{X}) = 1, \quad \gamma_l(\hat{X}) = \frac{1}{p^2}, \quad \kappa(\hat{X}) = p^2.$$

So, the scaling problem (5.20) has also the constant complexity (5.19).

6 Discussion

In this paper we introduced a new object: the augmented self-concordant barrier. We saw that it makes it possible to obtain new complexity results and to solve new applications. Those points are worth further discussion.

6.1 Finite methods in mathematical programming

One of the possibly surprising result in this paper is that, in some applications, the complexity of computational scheme is independent of the data of the problem instance. This strong property usually occurs in very specific contexts. Thus, we need to position our result with respect to the standard theory.

Let us look at the following groups of algorithms, traditionally recorded polynomial-time methods:

- Finite schemes in Linear Algebra.
- Finite schemes in Linear Programming.
- Polynomial-time methods in Nonlinear Programming.

The interpretation of complexity estimates for the first group is transparent. Take, for example, matrix inversion. It is possible to count the exact number of elementary operations that are needed to compute the result. However, the elementary operations are not all of the same nature. A division or a square root operation do not have the same level of complexity as an addition or a multiplication. Moreover, with the existing computer arithmetic, it impossible to compute the exact result in finite time. Nevertheless we treat these operations the same. Why so? The reason is that the computation of such an operation takes the *same computer time, independently* of the input data. Take for instance, the square root operation. Roughly speaking, it is implemented as follows. First, we scale the input to fit an appropriate interval. After that, we apply some iterative scheme, which converges quadratically (or even faster), almost at the start of the process.

Finite algorithms in Linear Programming have another justification. It is still possible to bound the number of elementary operations to reach a solution, but this bound depends on the input data, namely the dimensions and the bit length of the input. A good algorithm must have a complexity that is polynomial in the dimensions and in the logarithm of the bit length.

Finally, the complexity of a computational scheme for nonlinear programming is usually measured by an estimate for the number of iterations to achieve a certain level of accuracy. The scheme is a polynomial time one if this number is bounded by the product of a polynomial in the dimension (or another structural characteristics of the problem), with a sum of logarithms. One of these logarithms involves the solution accuracy; the other one, the "distance to infeasibility" [5]. In particular, if the distance is small, the algorithm can require very many iterations to converge.

The above review shows that the complexity results (5.10), (5.12) and (5.19) pertaining to problems (5.11), (5.15) and (5.20) do not fit well any of these groups of methods. In those examples, (5.11), (5.15) and (5.20), the proposed methods have two stages. The first stage consists of preliminary iterations to attain the domain of quadratic convergence. The length of this stage is *uniformly bounded* for the three problems. The second stage, on which we apply the Newton method, can be seen as finite, since its length is proportional to the double logarithm of the length of the computer word, not even the length of the input. This property makes this particular group of methods closer to the finite methods in linear algebra discussed above. Hence, we suggest to call the methods possessing the above properties "*finite methods* in nonlinear optimization". Hopefully, the problems we have seen in this paper are not the only ones which allow such complexity results.

To summarize, we have investigated several easy nonlinear optimization problems, which have *finite polynomial complexity.* These problems have some practical applications. Some of them already appeared as auxiliary problems in more general iterative schemes. We hope that further applications of that type can be found. Indeed, general iterative schemes are usually composed of basic operations, or set of operations, whose complexity is considered to be finite.

6.2 Impact of augmented barrier on interior point schemes.

The complexity results of this paper are obtained using the new objects, the augmented self-concordant barriers. In fact, the only non-trivial problem we can solve with such a barrier is the problem of finding its analytic center. Surprisingly enough, this problem appears to be almost universal. We have seen that the general feasibility problem and primal-dual conic minimization problem can be posed in that form. Moreover, in order to get the complexity results for the primal-dual feasibility problem we do not use its skew-symmetry and the corresponding properties of the central path. This gives us a kind of confirmation that the quadratic term in the augmented barrier improves the properties of self-concordant barrier. We hope also that this modification can also improve the numerical stability of other minimization schemes.

References

[1] J.-L. Goffin and J.-Ph. Vial. "Multiple cuts in the analytic center cutting plane method". *SIAM Journal on Optimization* 11, 266 - 288, 2000

[2] Yu. Nesterov. *Introductory lectures on Convex Optimization.* Unpublished manuscript, Louvain, Belgium, 1996.

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

[4] Yu. Nesterov and J.-Ph. Vial. "Homogeneous Analytic Center Cutting Plane Methods for Convex Problems and Variational Inequalities". *SIAM Journal on Optimization*, 9:707-728, 1999.

[5] J.Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization.* CORE Lecture Series, 1999.