Self-concordant functions in structured convex optimization

Image Technical Report 0007

François Glineur*
Service de Mathématique et de Recherche Opérationnelle,
Faculté Polytechnique de Mons,
Rue de Houdain, 9, B-7000 Mons, Belgium,
`Francois.Glineur@fpms.ac.be`.

October 2000

*The author is supported by a grant from the F.N.R.S. (Belgian National Fund for Scientific Research).

Abstract

This paper provides a self-contained introduction to the theory of self-concordant functions [8] and applies it to several classes of structured convex optimization problems. We describe the classical short-step interior-point method and optimize its parameters to provide its best possible iteration bound. We also discuss the necessity of introducing two parameters in the definition of self-concordancy, how they react to addition and scaling and which one is the best to fix. A lemma from [2] is improved and allows us to review several classes of structured convex optimization problems and evaluate their algorithmic complexity, using the self-concordancy of the associated logarithmic barriers.

Keywords. convex optimization, self-concordant functions, interior-point methods, entropy optimization, geometric optimization, l_p-norm optimization.

Contents

1 Introduction

1.1 Convex optimization

Convex optimization deals with the following problem

$$\inf_{x \in \mathbb{R}^n} f_0(x) \quad \text{s.t.} \quad x \in C \;, \tag{C}$$

where $C \subseteq \mathbb{R}^n$ is a closed convex set and $f_0 : C \mapsto \mathbb{R}$ is a convex function defined on C. Convexity of f_0 and C plays a very important role in this problem, since it is responsible for the following two important properties [11, 13]:

- Any local optimum for (C) is also a global optimum, which implies that the objective value is equal for all local optima. Moreover, all these optima can be shown to form a convex set.
- It is possible to use Lagrange duality to derive a dual problem strongly related to (C). Namely, this pair of problems satisfies a weak duality property (the objective value of any feasible solution for one of these problems provides a bound on the optimum objective value for the dual problem) and, under a Slater-type condition, a strong duality property (equality and attainment of the optimum objective values for the two problems).

We first note that it can be assumed with any loss of generality that the objective function f_0 is linear, so that we can define it as $f_0(x) = c^T x$ using a vector $c \in \mathbb{R}^n$. Indeed, it is readily seen that problem (C) is equivalent to the following problem with a linear objective:

$$\min_{x \in \mathbb{R}^{n+1}} x_{n+1} \quad \text{s.t.} \quad x \in \bar{C} \;,$$

where $\bar{C} \subseteq \mathbb{R}^{n+1}$ is suitably defined as

$$\bar{C} = \left\{(x,t) \in \mathbb{R}^{n+1} \mid x \in C \text{ and } f(x) \leq t\right\} \;.$$

We will thus consider in the rest of this paper the problem

$$\inf_{x \in \mathbb{R}^n} c^T x \quad \text{s.t.} \quad x \in C \;. \tag{CL}$$

It is interesting to ask ourselves how one can specify the data of a problem cast in such a form, i.e. how one can describe its objective function and feasible set. While specifying the objective function is easily done by providing vector c, describing the convex set C, which is responsible for the *structure* of problem (CL), can be done in several manners.

a. The traditional way to proceed in nonlinear optimization is to provide list of convex constraints defining C, i.e.

$$C = \left\{x \in \mathbb{R}^n \mid f_i(x) \leq 0 \;\forall i \in I = \{1, 2, \ldots, m\}\right\} \;, \tag{1.1}$$

where each of the m functions $f_i : \mathbb{R}^n \mapsto \mathbb{R}$ is convex. This guarantees the convexity of C, as an intersection of convex level sets.

b. An alternative approach consists in considering the domain of a convex function. More precisely, we require the interior of C to be equal to the domain of a convex function. Extending the real line $\mathbb{R}$ with the quantity $+\infty$, we introduce the convex function $F : \mathbb{R}^n \mapsto \mathbb{R} \cup \{+\infty\}$ and define C as the closure of its effective domain, i.e.

$$C = \operatorname{cl} \operatorname{dom} F = \operatorname{cl} \{x \in \mathbb{R}^n \mid F(x) < +\infty\} \ .$$

Most of the time, we will require in addition F to be a *barrier function* for the set C, according to the following definition.

Definition 1.1 *A function F is a* barrier function *for the convex set C if and only if it satisfies the following assumptions:*

(a) F is smooth (three times continuously differentiable for our purpose),

(b) F is strongly convex, i.e. $\nabla^2 F$ is positive definite,

(c) $F(x)$ tends to $+\infty$ whenever x tends to ∂C, the boundary of C (this is the barrier property).

Note 1.1 *We also note that it is often possible to provide a suitable barrier function F for a convex set C given by a functional description* (1.1) *using the* logarithmic barrier *[5] defined as*

$$F : \mathbb{R}^n \mapsto \mathbb{R} : x \mapsto F(x) = -\sum_{i \in I} \log(-f_i(x)) \ ,$$

where we define $\log x = +\infty$ *whenever $x \leq 0$. We have indeed to check that f is strongly convex and is a barrier function for C, which is not always the case (for example, in the case of $C = \mathbb{R}_+$, taking $f_1(x) = |x| - x$ does not lead to a strongly convex F while $f_1(x) = -x^x$ leads to $F(x) = -x \log x$, which does not possess the barrier property).*

c. It may also be worthwhile to consider the special case where C can be described as the intersection of a convex cone $K \subseteq \mathbb{R}^n$ and an affine subspace $b + L$ (where L is a linear subspace)

$$C = K \cap (b + L) = \{x \in K \mid x - b \in L\} \ .$$

The resulting class of problems is known as conic optimization, and can be easily shown to be equivalent to convex optimization [8] (in practice, subspace $b + L$ would be defined with a set of linear equalities).

Special treatment for the linear constraints, i.e. their representation as an intersection with an affine subspace, can be justified by the fact that these constraints are easier to handle than general nonlinear constraints. In particular, let us mention that it is usually easy for algorithms to preserve feasibility with respect to these constraints, and that they cannot cause a nonzero duality gap, i.e. strong duality is valid without a Slater-type assumption for linear optimization. However, we will not need to use this approach in this paper.

1.2 Interior-point methods

Among the different types of algorithms that can be applied to solve problem (CL), the so-called *interior-point methods* have gained a lot of popularity in the last two decades. This is mainly due to the following facts:

- ⋄ it is not only possible to prove convergence of these methods to an optimal solution but also to give a polynomial bound on the number of arithmetic operations needed to reach a solution within a given accuracy,
- ⋄ these methods can be implemented and applied successfully to solve real-world problems, especially in the fields of linear (where they compare favourably with the simplex method), quadratic and semidefinite optimization.

A fundamental ingredient in the elaboration of these methods is the above-mentioned notion of barrier function F for the set C. Namely, let us consider the following parameterized family of unconstrained minimization problems:

$$\inf_{x \in \mathbb{R}^n} \frac{c^T x}{\mu} + F(x) \ , \tag{CL$_\mu$}$$

where parameter μ belongs to $\mathbb{R}_{++}$ and is called the *barrier parameter*. The constraint $x \in C$ of the original problem (CL) has been replaced by a penalty term $F(x)$ in the objective function, which tends to $+\infty$ as x tends to the boundary of C and whose purpose is to avoid that the iterates leave the feasible set (see the classical monograph [4]). Assuming existence of a minimizer $x(\mu)$ for each of these problems (strong convexity of F ensures uniqueness of such a minimizer), we call the set $\{x(\mu) \mid \mu > 0\} \subseteq C$ the central path for problem (CL).

It is intuitively clear that as μ tends to zero, the first term proportional to the original objective $\frac{c^T x}{\mu}$ becomes preponderant in the sum, which implies that the central path converges to a solution that is optimal for the original problem. The principle behind interior-point methods will thus be to follow this central path until an iterate that is sufficiently close to the optimum is found.

However, two questions remain pending: how do we compute $x(\mu)$ and how do we choose a suitable barrier F. The first question is readily answered: interior-point methods rely on Newton's method to compute these minimizers, which leads us to a refined version of the second question: is it possible to choose a barrier function F such that Newton's method is provably efficient in solving subproblems (CL$_\mu$) and has an algorithmic complexity that can be estimated. This crucial question is thoroughly answered by the remarkable theory of self-concordant functions, first developed by Nesterov and Nemirovsky [8], which we will present in the next section.

1.3 Purpose and organization of the paper

The purpose of this paper is to give a self-contained introduction to the theory of self-concordant functions and to apply it to several classes of structured convex optimization problems. Section 2 introduces a definition of self-concordant functions and presents several equivalent conditions. A short-step interior-point method using these functions is then presented along with an explanation of how the proof of polynomiality works. Our contribution at this stage is the computation of the best possible iteration bound for this method.

Section 3 deals with the construction of self-concordant functions. Scaling and addition of self-concordant functions are considered, as well as a discussion on the utility of two parameters in the definition of self-concordancy and how to fix one of them in the best possible way. We then present an improved version of a lemma from [2]. This lemma is the main tool used in Section 4, where we review several classes of structured convex optimization problems and prove self-concordancy of the corresponding logarithmic barriers, improving the complexity results found in[2]. We conclude in Section 5 with some comments.

2 Self-concordancy

2.1 Definitions

Let us start with a definition of a self-concordant function.

Definition 2.1 *A function $F : C \mapsto R$ is called* (κ, ν)-self-concordant *for the convex set $C \subseteq \mathbb{R}^n$ if and only if F is a barrier function according to Definition 1.1 and the following two conditions hold for all $x \in \operatorname{int} C$ and $h \in \mathbb{R}^n$:*

$$\nabla^3 F(x)[h,h,h] \leq 2\kappa \left(\nabla^2 F(x)[h,h]\right)^{\frac{3}{2}} , \tag{2.1}$$
$$\nabla F(x)^T (\nabla^2 F(x))^{-1} \nabla F(x) \leq \nu \tag{2.2}$$

(note that the square root in (2.1) is well defined since its argument $\nabla^2 F(x)[h,h]$ is positive because of the requirement that F is convex).

This definition does not exactly match the original definition of a self-concordant barrier in [8], but merely corresponds to the notion of strongly non-degenerate κ^{-2}-self-concordant barrier functional with parameter ν, that is general enough for our purpose.

Note 2.1 *We would like to point out that no absolute value is needed in* (2.1)*: while some authors usually require the apparently stronger condition*

$$\left|\nabla^3 F(x)[h,h,h]\right| \leq 2\kappa \left(\nabla^2 F(x)[h,h]\right)^{\frac{3}{2}} , \tag{2.3}$$

this is not needed since it suffices to notice that inequality (2.1) *also has to hold in the direction opposite to h, which gives*

$$\nabla^3 F(x)[-h,-h,-h] \leq 2\kappa \left(\nabla^2 F(x)[-h,-h]\right)^{\frac{3}{2}} \Leftrightarrow -\nabla^3 F(x)[h,h,h] \leq 2\kappa \left(\nabla^2 F(x)[h,h]\right)^{\frac{3}{2}}$$

(using the fact that the n^{th}-order differential is homogeneous with degree n), which combined with (2.1) *gives condition* (2.3)*.*

It is possible to reformulate conditions (2.1) and (2.2) into several equivalent inequalities that may prove easier to handle in some cases. However, before we list them, we would like to make a few comments about the use of inner products in our setting, following the line of thought of Renegar's monograph [10].

It is indeed important to realize that the definitions of gradient and Hessian, i.e. first-order and second-order differentials are in fact dependent from inner product that is being used. Nevertheless, in most texts, it is customary to use the dot product[1] as standard inner

[1] The dot product of two vector x and y whose coordinates are $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ and $(\beta_1, \beta_2, \ldots, \beta_n)$ in a given coordinate system is equal to $\sum_{i=1}^n \alpha_i \beta_i$.

product. This has the disadvantage to make all developments *a priori* dependent from the coordinate system. However, Renegar notices that it is possible to develop the theory of self-concordant functions in a completely coordinate-free manner, i.e. independently of a reference inner product. The is due to the fact that the two principal objects in this theory are indeed independent from the coordinate system: the Newton step $n(x)$ and the intrinsic inner product $\langle\cdot,\cdot\rangle_x$. Given a barrier function F and a point x belonging to its domain, these two objects are defined according to:

$$n(x) = -(\nabla^2 F(x))^{-1}\nabla F(x) \quad \text{and} \quad \langle\alpha,\beta\rangle_x = \langle\alpha,\nabla^2 F(x)\beta\rangle \ .$$

It is also convenient to introduce the intrinsic norm $\|\cdot\|_x$ based on the intrinsic inner product $\langle\cdot,\cdot\rangle_x$ according to the usual definition $\|a\|_x = \sqrt{\langle a,a\rangle_x}$.

Let $x \in \operatorname{int} C$ and $h \in \mathbb{R}^n$ and let us introduce the one-dimensional function $F_{x,h} : \mathbb{R} \mapsto \mathbb{R} : t \mapsto F(x+th)$, the restriction of F along the line $\{x+th \mid t \in \mathbb{R}\}$. We are now in position to state several reformulations of conditions (2.1) and (2.2), grouped in the following two theorems:

Theorem 2.1 *The following four conditions are equivalent:*

$$\begin{aligned}
\nabla^3 F(x)[h,h,h] &\le 2\kappa\left(\nabla^2 F(x)[h,h]\right)^{\frac{3}{2}} \textit{ for all } x \in \operatorname{int} C \textit{ and } h \in \mathbb{R}^n && (2.4a)\\
F'''_{x,h}(0) &\le 2\kappa F''_{x,h}(0)^{\frac{3}{2}} \textit{ for all } x \in \operatorname{int} C \textit{ and } h \in \mathbb{R}^n && (2.4b)\\
F'''_{x,h}(t) &\le 2\kappa F''_{x,h}(t)^{\frac{3}{2}} \textit{ for all } x+th \in \operatorname{int} C \textit{ and } h \in \mathbb{R}^n && (2.4c)\\
\left(-\frac{1}{\sqrt{F''_{x,h}(t)}}\right)' &\le \kappa \textit{ for all } x+th \in \operatorname{int} C \textit{ and } h \in \mathbb{R}^n \ . && (2.4d)
\end{aligned}$$

Proof. Since $F_{x,h}(t) = F(x+th)$, we can write

$$F'_{x,h}(t) = \nabla F(x+th)[h],\ F''_{x,h}(t) = \nabla^2 F(x+th)[h,h] \text{ and } F'''_{x,h}(t) = \nabla^3 F(x+th)[h,h,h] \ .$$

Condition (2.4b) is thus simply condition (2.4a) written differently. Moreover, condition (2.4c) is equivalent to condition (2.4a) written for $x+th$ instead of x. Finally, we note that

$$\left(-\frac{1}{\sqrt{F''_{x,h}(t)}}\right)' \le \kappa \Leftrightarrow \frac{1}{2}F''_{x,h}(t)^{-\frac{3}{2}}F'''_{x,h}(t) \le \kappa \Leftrightarrow F'''_{x,h}(t) \le 2\kappa F''_{x,h}(t)^{\frac{3}{2}} \ ,$$

which shows that (2.4d) and (2.4c) are equivalent. □

Theorem 2.2 *The following four conditions are equivalent:*

$$\begin{aligned}
\nabla F(x)^T(\nabla^2 F(x))^{-1}\nabla F(x) &\le \nu \textit{ for all } x \in \operatorname{int} C && (2.5a)\\
F'_{x,h}(0)^2 &\le \nu F''_{x,h}(0) \textit{ for all } x \in \operatorname{int} C \textit{ and } h \in \mathbb{R}^n && (2.5b)\\
F'_{x,h}(t)^2 &\le \nu F''_{x,h}(t) \textit{ for all } x+th \in \operatorname{int} C \textit{ and } h \in \mathbb{R}^n && (2.5c)\\
\left(-\frac{1}{F'_{x,h}(t)}\right)' &\ge \frac{1}{\nu} \textit{ for all } x+th \in \operatorname{int} C \textit{ and } h \in \mathbb{R}^n \ . && (2.5d)
\end{aligned}$$

Proof. Proving these equivalences is a little more involved than for the previous theorem. We start by showing that condition (2.5b) implies condition (2.5a). We can write

$$\begin{aligned}
\nabla F(x)^T(\nabla^2 F(x))^{-1}\nabla F(x) &= \nabla F(x)[(\nabla^2 F(x))^{-1}\nabla F(x)] = F'_{x,(\nabla^2 F(x))^{-1}\nabla F(x)}(0) \\
&\le \sqrt{\nu}\sqrt{F''_{x,(\nabla^2 F(x))^{-1}\nabla F(x)}(0)} \text{ using condition (2.5b)} \\
&= \sqrt{\nu}\sqrt{\nabla^2 F(x)[(\nabla^2 F(x))^{-1}\nabla F(x),(\nabla^2 F(x))^{-1}\nabla F(x)]} \\
&= \sqrt{\nu}\sqrt{\nabla F(x)^T(\nabla^2 F(x))^{-1}\nabla^2 F(x)(\nabla^2 F(x))^{-1}\nabla F(x)} \\
&= \sqrt{\nu}\sqrt{\nabla F(x)^T(\nabla^2 F(x))^{-1}\nabla F(x)}\ ,
\end{aligned}$$

which implies condition (2.5a). Considering now the reverse implication, we have

$$\begin{aligned}
F'_{x,h}(0)^2 &= (\nabla F(x)[h])^2 = \left(\nabla F(x)^T h\right)^2 \\
&= \left(\nabla F(x)^T(\nabla^2 F(x))^{-1}\nabla^2 F(x)h\right)^2 = \langle(\nabla^2 F(x))^{-1}\nabla F(x), h\rangle_x^2 \\
&\le \left\|(\nabla^2 F(x))^{-1}\nabla F(x)\right\|_x^2 \|h\|_x^2 \text{ (using the Cauchy-Schwarz inequality)} \\
&= \left(\nabla F(x)^T(\nabla^2 F(x))^{-1}\nabla^2 F(x)(\nabla^2 F(x))^{-1}\nabla F(x)\right)\left(h^T\nabla^2 F(x)h\right) \\
&\le \nu\nabla^2 F(x)[h,h] \text{ using condition (2.5a)} \\
&= \nu F''_{x,h}(0)\ .
\end{aligned}$$

Condition (2.5c) is condition (2.5b) written for $x + th$ instead of x, and we finally note that

$$\left(-\frac{1}{F'_{x,h}(t)}\right)' \ge \frac{1}{\nu} \Leftrightarrow F'_{x,h}(t)^{-2}F''_{x,h}(t) \ge \frac{1}{\nu} \Leftrightarrow \nu F''_{x,h}(t) \ge F'_{x,h}(t)^2\ ,$$

which shows that (2.5d) and (2.5c) are equivalent. □

The first three reformulations for each condition are well-known and can be found for example in [8, 7, 10]. Conditions (2.4d) and (2.5d) are less commonly used and were independently presented in [1].

2.2 Short-step method

As outlined in the introduction, interior-point methods for convex optimization rely on a barrier function and the associated central path to solve problem (CL). Ideally, we would like our iterates to be a sequence of points on the central path $x(\mu_0)$, $x(\mu_1), \ldots,$ $x(\mu_k), \ldots$ for a sequence of barrier parameters μ_k tending to zero (and thus $x(\mu_k)$ tending to an optimal solution).

We already mentioned that Newton's method, applied to problems (CL_μ), will be the workhorse to compute those minimizers. However, it would be too costly to compute each of these points with high accuracy, so that interior-point methods require instead their iterates to lie in a prescribed neighbourhood of the central path and its exact minimizers.

Let x_k, the k^{th} iterate, be an approximation of $x(\mu_k)$. A good proximity measure would be $\|x_k - x(\mu_k)\|$ or, to be independent from the coordinate system, $\|x_k - x(\mu_k)\|_{x_k}$. However, this quantity involve the unknown central point $x(\mu_k)$, and is therefore difficult to work with. Nevertheless, another elegant proximity measure can be used for that purpose. Let us define

$n_\mu(x)$ to be the Newton step trying to minimize the objective in problem (CL_μ), which is thus aiming at $x(\mu_k)$. Since this objective is equal to $F_\mu(x) = \frac{c^T x}{\mu} + F(x)$, we have

$$n_\mu(x) = -(\nabla^2 F_\mu(x))^{-1} \nabla F_\mu(x) = -(\nabla^2 F(x))^{-1}(\frac{c}{\mu} + \nabla F(x)) = -\frac{1}{\mu}(\nabla^2 F(x))^{-1} c + n(x) \; .$$

Let us now define $\delta(x, \mu)$, a measure of the proximity of x to the central point $x(\mu)$, as the intrinsic norm of the newton step $n_\mu(x)$, i.e. $\delta(x,\mu) = \|n_\mu(x)\|_x$. This quantity is indeed a good candidate to measure how far x lies from the minimizer $x(\mu)$, since the Newton step at x targeting $x(\mu)$ is supposed to be a good approximation of $x(\mu) - x$. The goal of a short-step interior-point method will be to trace the central path approximately, ensuring that the proximity $\delta(x_k, \mu_k)$ is kept below a predefined bound for each iterate.

We are now in position to sketch a short-step algorithm. Given a problem of type (CL), a barrier function F for C, an upper bound on the proximity measure $\tau > 0$, a decrease parameter $0 < \theta < 1$ and an initial iterate x_0 such that $\delta(x_0, \mu_0) < \tau$, we set $k \leftarrow 0$ and perform the following main loop:

$$\begin{aligned}
&\mu_{k+1} \leftarrow \mu_k(1-\theta) \\
&x_{k+1} \leftarrow x_k + n_{\mu_{k+1}}(x_k) \\
&k \leftarrow k+1
\end{aligned}$$

The key is to choose parameters τ and θ such that $\delta(x_k, \mu_k) < \tau$ implies $\delta(x_{k+1}, \mu_{k+1}) < \tau$, so that proximity to the central path is preserved. This is the moment where the self-concordancy of the barrier function F comes into play. Indeed, it is precisely this property that will guarantee that such a choice is always possible.

2.3 Optimal complexity

In order to relate the two proximities $\delta(x_k, \mu_k)$ and $\delta(x_{k+1}, \mu_{k+1})$, it is useful to introduce an intermediate quantity $\delta(x_k, \mu_{k+1})$, the proximity from an iterate to its next target on the central path. We have the following two properties:

Theorem 2.3 *Let F be a barrier function satisfying* (2.2), $x \in \operatorname{dom} F$ *and* $\mu^+ = (1-\theta)\mu$. *We have*

$$\delta(x, \mu^+) \leq \frac{\delta(x,\mu) + \theta\sqrt{\nu}}{1-\theta} \; .$$

Proof. We have

$$\begin{aligned}
& & \mu^+ n_{\mu^+}(x) - \mu^+ n(x) = -(\nabla^2 F(x))^{-1} c = \mu n_\mu(x) - \mu n(x) \\
\text{(dividing by } \mu) &\Leftrightarrow & (1-\theta) n_{\mu^+}(x) - (1-\theta) n(x) = n_\mu(x) - n(x) \\
&\Leftrightarrow & (1-\theta) n_{\mu^+}(x) = n_\mu(x) - \theta n(x) \\
&\Rightarrow & (1-\theta) \left\|n_{\mu^+}(x)\right\|_x \leq \|n_\mu(x)\|_x + \theta \|n(x)\|_x \\
&\Rightarrow & (1-\theta)\delta(x, \mu^+) \leq \delta(x,\mu) + \theta\sqrt{\nu} \; ,
\end{aligned}$$

which implies the desired inequality, where we used to derive the last implication the fact that

$$\begin{aligned}
\|n(x)\|_x = \sqrt{\langle n(x), n(x)\rangle_x} &= \sqrt{\nabla F(x)^T (\nabla^2 F(x))^{-1} \nabla^2 F(x) (\nabla^2 F(x))^{-1} \nabla F(x)} \\
&= \sqrt{\nabla F(x)^T (\nabla^2 F(x))^{-1} \nabla F(x)} \leq \sqrt{\nu} \; ,
\end{aligned}$$

because of condition (2.2). $\square$

Theorem 2.4 *Let F be a barrier function satisfying (2.1) and $x \in \operatorname{dom} F$. Let us suppose $\delta(x,\mu) < \frac{1}{\kappa}$. We have that $x + n_\mu(x) \in \operatorname{dom} F$ and*

$$\delta(x + n_\mu(x), \mu) \leq \frac{\kappa\delta(x,\mu)^2}{(1-\kappa\delta(x,\mu))^2} .$$

(this proof is more technical and is omitted here ; it can be found in [8, 7, 10]).

Note 2.2 *It is now clear why the self-concordancy property relies on two separate conditions: one of them is responsible for the control of the increase of the proximity measure when the target on the central path is updated (Theorem 2.3), while the other guarantees that the proximity to the target can be restored, i.e. sufficiently decreased when taking a Newton step (Theorem 2.4).*

Assuming for the moment that τ and θ can be chosen such that the proximity to central path is preserved at each iteration, we see that the number of iterations needed to attain a certain value of the barrier parameter μ_e depends solely on the ratio $\frac{\mu_0}{\mu_e}$ and the value of θ. Namely, since $\mu_k = (1-\theta)^k \mu_0$, it is readily seen that this number of iterations is equal to

$$\left\lceil \log_{(1-\theta)} \frac{\mu_e}{\mu_0} \right\rceil = \left\lceil \frac{1}{\log 1-\theta} \log \frac{\mu_e}{\mu_0} \right\rceil . \tag{2.6}$$

Given a (κ,ν)-self-concordant function, we are now going to find a suitable pair of parameters τ and θ . Moreover, we will optimize this choice of parameters, i.e. try to provide the greatest reduction for the parameter μ at each iteration, in other words maximize θ in order to get the lowest possible total iteration count. Letting $\delta = \delta(x_k, \mu_k)$, $\delta' = \delta(x_k, \mu_{k+1})$ and $\delta^+ = \delta(x_{k+1}, \mu_{k+1})$ and assuming $\delta \leq \tau$, we have to satisfy $\delta^+ \leq \tau$ with the greatest possible value for θ.

Let us assume first that $\delta' < \frac{1}{\kappa}$. Using Theorem 2.4, we find that

$$\delta^+ \leq \frac{\kappa\delta'^2}{(1-\kappa\delta')^2}$$

and therefore require that

$$\frac{\kappa\delta'^2}{(1-\kappa\delta')^2} \leq \tau .$$

This is equivalent to

$$\left(\frac{\kappa\delta'}{1-\kappa\delta'}\right)^2 \leq \kappa\tau \Leftrightarrow \left(\frac{1}{\kappa\delta'} - 1\right)^2 \geq \frac{1}{\kappa\tau} \Leftrightarrow \frac{1}{\kappa\delta'} \geq 1 + \frac{1}{\sqrt{\kappa\tau}}$$

(this also shows that the assumption $\kappa\delta' < 1$ we made in the beginning was valid). Using now Theorem 2.3, we know that

$$\delta' \leq \frac{\delta + \theta\sqrt{\nu}}{1-\theta} \Rightarrow \delta' \leq \frac{\tau + \theta\sqrt{\nu}}{1-\theta} \Leftrightarrow \frac{1}{\kappa\delta'} \geq \frac{1-\theta}{\kappa\tau + \theta\kappa\sqrt{\nu}}$$

and thus require that

$$\frac{1-\theta}{\kappa\tau+\theta\kappa\sqrt{\nu}} \geq 1+\frac{1}{\sqrt{\kappa\tau}} .$$

Letting $C=\kappa\sqrt{\nu}$ and $\beta=\sqrt{\kappa\tau}$ we have

$$\frac{1-\theta}{\beta^2+\theta C} \geq 1+\frac{1}{\beta} \Leftrightarrow 1-\theta \geq (1+\frac{1}{\beta})(\beta^2+C\theta) \Leftrightarrow 1-\beta-\beta^2 \geq \theta\left(1+C+\frac{C}{\beta}\right) ,$$

which means finally that we have to choose θ such that

$$\theta \leq \frac{1-\beta-\beta^2}{1+C+\frac{C}{\beta}} \tag{2.7}$$

in order to guarantee $\delta^+ \leq \tau$. We are now in position to optimize the value of θ, i.e. find the value of β that maximizes this upper bound. However, this value is likely to depend on C (and thus on κ and ν) in a complex way. We are therefore going to work with the following slightly worse upper bound, which has the advantage of allowing the optimization of β independently of C (we use the fact that $C=\kappa\sqrt{\nu}\geq 1$, see [8])

$$\theta \leq \frac{1}{C}\Big(\frac{1-\beta-\beta^2}{2+\frac{1}{\beta}}\Big) = \frac{f(\beta)}{C} \quad \Big(\leq \frac{1-\beta-\beta^2}{1+C+\frac{C}{\beta}}\Big) .$$

It is now straightforward to maximize $f(\beta)$: computing the derivative shows there is a unique maximizer when $\beta \approx 0.273$ (the exact value is the real root of $1-2\beta-5\beta^2-4\beta^3$) and our upper bound in (2.7) becomes $\frac{0.65}{1+4.66C}$. Translating back into our original quantities τ, κ and ν we find that we can choose

$$\tau = \frac{\beta^2}{\kappa} \approx \frac{1}{13.42\kappa} \quad \text{and} \quad \theta = \frac{1-\beta-\beta^2}{1+C+\frac{C}{\beta}} \approx \frac{1}{1.53+7.15\kappa\sqrt{\nu}} , \tag{2.8}$$

which is the best result obtainable if we want β to be independent from κ and ν. This improves several results from the literature, e.g. $\theta=\frac{1}{9\kappa\sqrt{\nu}}$ in [7] and $\theta=\frac{1}{1+8\kappa\sqrt{\nu}}$ in [10].

Before we conclude this section with a global complexity result, let us say a few words about termination of the algorithm. The most practical stopping criteria is a small target value μ_e for the barrier parameter, which gives the iteration bound (2.6). Our final iterate x_e will thus satisfy $\delta(x_e,\mu_e)\leq\tau$, which tells us it is not too far from $x(\mu_e)$, itself not too far from the optimum since μ_e is small. Indeed, using again the self-concordancy property of F, it is possible to derive the following bound on the accuracy of the final objective $c^T x_e$, i.e. its deviation from the optimal objective $c^T x^*$

$$c^T x_e - c^T x^* \leq \frac{\mu_e}{1-3\kappa\tau}\kappa\sqrt{\nu} \tag{2.9}$$

(proof of this fact is omitted here and can easily be obtained combining Theorems 2.2.5 and 2.3.3 in [10]). We are now ready to state our final complexity result:

Theorem 2.5 *Given a convex optimization problem* (CL), *a* (κ,ν)*-self-concordant barrier* F *for* C *and an initial iterate* x_0 *such that* $\delta(x_0,\mu_0) < \frac{1}{13.42\kappa}$, *one can find a solution with accuracy* ϵ *in*

$$\left\lceil (1.03+7.15\kappa\sqrt{\nu})\log\frac{1.29\mu_0\kappa\sqrt{nu}}{\epsilon} \right\rceil \quad \textit{iterations.}$$

Proof. Using our optimal values for θ and τ from (2.8) and the bound on the objective accuracy in (2.9), we find that the stopping threshold on the barrier parameter μ_e must satisfy

$$\frac{\mu_e}{1-3/13.42}\kappa\sqrt{\nu} \le \epsilon \Leftrightarrow 1.29\mu_e\kappa\sqrt{\nu} \le \epsilon \Leftrightarrow \mu_e \le \frac{\epsilon}{1.29\kappa\sqrt{\nu}}\ .$$

Plugging this value into (2.6) we find that the total number of iterations can be bounded by (omitting the rounding bracket for clarity)

$$\begin{aligned}\frac{1}{\log(1-\theta)}\log\frac{\mu_e}{\mu_0} &\le \frac{1}{\log(1-\theta)}\log\frac{\epsilon}{1.29\mu_0\kappa\sqrt{\nu}} \\ &= -\frac{1}{\log(1-\theta)}\log\frac{1.29\mu_0\kappa\sqrt{\nu}}{\epsilon} \\ &\le \left(\frac{1}{\theta}-\frac{1}{2}\right)\log\frac{1.29\mu_0\kappa\sqrt{\nu}}{\epsilon} \\ &= (1.03+7.15\kappa\sqrt{\nu})\log\frac{1.29\mu_0\kappa\sqrt{\nu}}{\epsilon}\ ,\end{aligned}$$

as announced (the third line uses the inequality $\frac{1}{\log(1-\theta)} \ge \frac{1}{2}-\frac{1}{\theta}$, which can be easily derived using the Taylor series of $\log x$ around 1). □

3 Proving self-concordancy

3.1 Barrier calculus

The previous section has made clear that the self-concordancy property of the barrier function F is essential to derive a polynomial bound on the number of iterations of the short-step method. Moreover, smaller values for parameters κ and ν imply a lower total complexity. The next question we may ask ourselves is how to find self-concordant barriers (ideally with low parameters).

An impressive result in [8] states that every convex set in $\mathbb{R}^n$ admits a (K,n)-self-concordant barrier, where K is a universal constant (independent of n). However, the universal barrier they provide in their proof is defined as a volume integral over an n-dimensional convex body, and is therefore difficult to evaluate in practice, even for simple sets in low-dimensional spaces. Another potential problem with this approach is that evaluating this barrier (and/or its gradient and Hessian) might take a number of arithmetic operations that grows exponentially with n, which would lead to an exponential algorithmic complexity for the short-step method, despite the polynomial iteration bound.

Another approach to find self-concordant function is to combine basic self-concordant functions using operations that are known to preserve self-concordancy (this approach is called *barrier calculus* in [8]). We are now going to describe two of these self-concordancy preserving operations, positive scaling and addition, and examine how the associated parameters are affected in the process.

Let us start with positive scalar multiplication.

Theorem 3.1 *Let F be a (κ,ν)-self-concordant barrier for $C \subseteq \mathbb{R}^n$ and $\lambda \in \mathbb{R}_{++}$ a positive scalar. Then (λF) is also a self-concordant barrier for C with parameters $(\frac{\kappa}{\sqrt{\lambda}}, \lambda\nu)$.*

Proof. It is clear that (λF) is also a barrier function (i.e. smoothness, strong convexity and the barrier property are obviously preserved by scaling). Looking at the restrictions $(\lambda F)_{x,h} = \lambda F_{x,h}$, we also have that

$$(\lambda F)'_{x,h} = \lambda F'_{x,h},\ (\lambda F)''_{x,h} = \lambda F''_{x,h} \text{ and } (\lambda F)'''_{x,h} = \lambda F'''_{x,h} \ .$$

Since F is (κ, ν)-self-concordant, we have (using Theorems 2.1 and 2.2)

$$F'''_{x,h}(0) \leq 2\kappa F''_{x,h}(0)^{\frac{3}{2}} \text{ and } F'_{x,h}(0)^2 \leq \nu F''_{x,h}(0) \text{ for all } x \in \operatorname{int} C \text{ and } h \in \mathbb{R}^n \ .$$

This is equivalent to

$$\lambda F'''_{x,h}(0) \leq 2\frac{\kappa}{\sqrt{\lambda}}(\lambda F''_{x,h}(0))^{\frac{3}{2}} \text{ and } (\lambda F'_{x,h}(0))^2 \leq \lambda\nu\lambda F''_{x,h}(0) \text{ for all } x \in \operatorname{int} C \text{ and } h \in \mathbb{R}^n \ ,$$

which is precisely stating that (λF) is $(\frac{\kappa}{\sqrt{\lambda}}, \lambda\nu)$-self-concordant. □

This theorem show that self-concordancy is preserved by positive scalar multiplication, but that parameters κ and ν are both modified. It is interesting to note that these parameters do not occur individually in the iteration bound of Theorem 2.5 but are rather always appearing together in the expression $\kappa\sqrt{\nu}$. This quantity, which we will call the *complexity value* of the barrier, is solely responsible for the polynomial iteration bound. Looking at what happens to it when F is scaled by λ, we find that the scaled complexity value is equal to $\frac{\kappa}{\sqrt{\nu}}\sqrt{\lambda\nu} = \kappa\sqrt{\nu}$, i.e. that the complexity value is invariant to scaling. This means *in fine* that scaling a self-concordant barrier does not influence the algorithmic complexity of the associated short-step method, a property than could reasonably be expected from the start.

Let us now examine what happens when two self-concordant barriers are added.

Theorem 3.2 *Let F be a (κ_1, ν_1)-self-concordant barrier for $C_1 \subseteq \mathbb{R}^n$ and G be a (κ_2, ν_2)-self-concordant barrier for $C_2 \subseteq \mathbb{R}^n$. Then $(F + G)$ is a self-concordant barrier for $C_1 \cap C_2$ (provided this intersection is nonempty) with parameters* $(\max\{\kappa_1, \kappa_2\}, \lambda_1 + \lambda_2)$.

Proof. It is straightforward to see that $(F + G)$ is a barrier function for $C_1 \cap C_2$. Looking at the restrictions $(F + G)_{x,h}$, we also have that

$$(F+G)'_{x,h} = F'_{x,h} + G'_{x,h},\ (F+G)''_{x,h} = F''_{x,h} + G''_{x,h} \text{ and } (F+G)'''_{x,h} = F'''_{x,h} + G'''_{x,h} \ .$$

We can write thus

$$\begin{aligned} (F+G)'''_{x,h} = F'''_{x,h} + G'''_{x,h} &\leq 2\kappa_1 {F''_{x,h}}^{\frac{3}{2}} + 2\kappa_2 {G''_{x,h}}^{\frac{3}{2}} \\ &\leq 2\max\{\kappa_1, \kappa_2\}({F''_{x,h}}^{\frac{3}{2}} + {G''_{x,h}}^{\frac{3}{2}}) \\ &\leq 2\max\{\kappa_1, \kappa_2\}(F''_{x,h} + G''_{x,h})^{\frac{3}{2}} = 2\max\{\kappa_1, \kappa_2\}(F+G)''_{x,h} \end{aligned}$$

(where we used for the third inequality the easily proven fact $x^{\frac{3}{2}} + y^{\frac{3}{2}} \leq (x+y)^{\frac{3}{2}}$ for $x, y \in \mathbb{R}_{++}$) and

$$\begin{aligned} \left|(F+G)'_{x,h}\right| = \left|F'_{x,h} + G'_{x,h}\right| &\leq \left|F'_{x,h}\right| + \left|G'_{x,h}\right| \\ &\leq \sqrt{\nu_1}\sqrt{F''_{x,h}} + \sqrt{\nu_2}\sqrt{G''_{x,h}} \\ &\leq \sqrt{\nu_1 + \nu_2}\sqrt{F''_{x,h} + G''_{x,h}} = \sqrt{\nu_1 + \nu_2}\sqrt{(F+G)''_{x,h}} \end{aligned}$$

(where we used for the third inequality the Cauchy-Schwarz inequality applied to vectors $(\sqrt{\nu_1}, \sqrt{\nu_2})$ and $(\sqrt{F''_{x,h}}, \sqrt{G''_{x,h}})$), which is precisely stating that $(F+G)$ is $(\max\{\kappa_1, \kappa_2\}, \nu_1 + \nu_2)$-self-concordant. □

3.2 Fixing a parameter

As mentioned above, scaling a barrier function with a positive scalar does not affect its self-concordancy, i.e. its suitability as a tool for convex optimization, and leaves its complexity value unchanged. One can thus make the decision to fix one of the two parameters κ and ν arbitrarily and only work with the corresponding subclass of barrier, without any real loss of generality. We describe now two choices of this kind that have been made in the literature.

First choice. Some authors [2, 12, 6, 3] choose to work with the second parameter ν fixed to one. However, this choice is not made explicitly but results from the particular structure of the barrier functions that are considered. Indeed, these authors consider convex optimization problems whose feasible sets are given by a functional description like (1.1), i.e.

$$\min_{x \in \mathbb{R}^n} c^T x \quad \text{s.t.} \quad f_i(x) \leq 0 \ \forall i \in I \ .$$

In order to apply the interior-point method methodology, a barrier function is needed and it is customary to use the logarithmic barrier as described in Note 1.1

$$F : \mathbb{R}^n \mapsto \mathbb{R} : x \mapsto F(x) = -\sum_{i \in I} \log(-f_i(x)) \ .$$

The following lemma will prove useful.

Lemma 3.1 *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a convex function and define $F : \mathbb{R}^n \mapsto \mathbb{R} \cup \{+\infty\} : x \mapsto -\log(-f(x))$, whose effective domain is the set $C = \{x \in \mathbb{R}^n \mid f(x) < 0\}$. We have that F satisfies the second condition of self-concordancy (2.2) with parameter $\nu = 1$.*

Proof. Using the equivalent condition (2.5b) of Theorem 2.2, we have to evaluate for $x \in \operatorname{int} C$, $h \in \mathbb{R}^n$ and $t = 0$

$$F'_{x,h}(t) = -\frac{\nabla f(x+th)[h]}{f(x+th)} \text{ and } F''_{x,h}(t) = \frac{\nabla f(x+th)[h]^2 - \nabla^2 f(x+th)[h,h] f(x+th)}{f(t+th)^2} ,$$

which implies

$$F'_{x,h}(0)^2 = \frac{\nabla f(x)[h]^2}{f(x)^2} \leq \frac{\nabla f(x)[h]^2 - \nabla^2 f(x)[h,h] f(x)}{f(x)^2} = F''_{x,h}(0)$$

(where we have used the fact that $\nabla^2 f(x)[h,h] \geq 0$ because f is convex and $f(x) \leq 0$ because x belongs to the feasible set C), which implies that F satisfies the second self-concordancy condition (2.2) with $\nu = 1$. □

Since the complete logarithmic barrier is a sum of terms for which this lemma is applicable, we can use Theorem 3.2 to find that it satisfies the same condition with $\nu = |I| = m$, the number of constraints.

This means that we only have to check the first condition (2.1) involving κ to establish self-concordancy for the logarithmic barrier. Assuming that each individual term $-\log(-f_i(x))$ can be shown to satisfy it with $\kappa = \kappa_i$, we have that the whole logarithmic barrier is $(\max_{i\in I}\{\kappa_i\}, m)$-self-concordant, which leads to a complexity value equal to $\|\kappa\|_\infty \sqrt{m}$, where we have defined $\kappa = (\kappa_1, \kappa_2, \ldots, \kappa_m)$.

Second choice. Another arbitrary choice of self-concordance parameters that one encounters frequently in the literature consists in fixing $\kappa = 1$ in the first self-concordancy condition (2.1). This approach has been used increasingly in the recent years (see e.g. [8, 10, 7]), and we propose to give here a justification of its superiority over the alternative presented above.

Let us consider the same logarithmic barrier, and suppose again that each individual term $F_i : x \mapsto -\log(-f_i(x))$ has been shown to satisfy the first self-concordancy condition (2.1) with $\kappa = \kappa_i$. Our previous discussion implies thus that F_i is $(\kappa, 1)$-self-concordant. Multiplying now F_i with κ_i^2, Theorem 3.1 implies that $\kappa_i^2 F_i$ is $(1, \kappa_i^2)$-self-concordant. The corresponding complete scaled logarithmic barrier

$$\tilde{F} : x \mapsto -\sum_{i\in I} \kappa_i^2 \log(-f_i(x))$$

is then $(1, \sum_{i\in I}\kappa_i^2)$-self-concordant by virtue of Theorem 3.2, which leads finally to a complexity value equal to $\sqrt{\sum_{i\in I}\kappa_i^2} = \|\kappa\|_2$. This quantity is always lower than the complexity value for the standard logarithmic barrier considered above because of the well-known norm inequality $\|\kappa\|_2 \leq \sqrt{m}\,\|\kappa\|_\infty$, which proves the superiority of this second approach (the only case where they are equivalent is when all parameters κ_i's are equal).

Note 3.1 *The fundamental reason why the first approach is less efficient is that it makes us combine barriers with different κ parameters, with the consequence that only the largest value $\max_{i\in I}\{\kappa_i\}$ appears in the final complexity value (the other smaller values become completely irrelevant and do not influence the final complexity at all). The second approach avoids this situation by ensuring that κ is always equal to one, which means that κ's are equal for each combination and that the final complexity is well depending on the parameters of all the terms of the logarithmic barrier.*

3.3 Two useful lemmas

We have seen so far how to construct self-concordant barrier by combining simpler functionals but still have no tool to prove self-concordancy of these basic barriers. The purpose of this section is to present two lemmas that can help us in that regard.

The first one deals with the second condition of self-concordancy with logarithmically homogeneous barriers [8].

Lemma 3.2 *Let us suppose F is a logarithmically homogeneous function with parameter α, i.e.*

$$F(tx) = F(x) - \alpha \log t\ . \tag{3.1}$$

We have that F satisfies the second condition of self-concordancy (2.2) with parameter $\nu = \alpha$.

Proof. This fact admits the following straightforward proof. We start by differentiating both sides of (3.1) with respect to t, to find

$$\nabla F(tx)[x] = -\frac{\alpha}{t} \ .$$

Fixing $t = 1$ gives

$$\nabla F(x)[x] = \nabla F(x)^T x = -\alpha \ . \tag{3.2}$$

Differentiating this last equality again, this time with respect to x, leads to

$$\nabla F(x) + \nabla^2 F(x) x = 0 \Leftrightarrow \nabla F(x) = -\nabla^2 F(x) x \ . \tag{3.3}$$

Looking now at the left-hand side in (2.2) we have

$$\nabla F(x)^T (\nabla^2 F(x))^{-1} \nabla F(x) = -\nabla F(x)^T (\nabla^2 F(x))^{-1} \nabla^2 F(x) x = -\nabla F(x)^T x = \alpha$$

(using successively (3.3) and (3.2)), which implies immediately that F satisfies the second condition of self-concordancy (2.2) with $\nu = \alpha$. □

The second lemma we are going to present deals with the first self-concordancy condition. Let us first introduce two auxiliary functions r_1 and r_2, whose graphs are depicted in Figure 3.3:

$$r_1 : \mathbb{R} \mapsto \mathbb{R} : \gamma \mapsto \max\{1, \frac{\gamma}{\sqrt{3 - 2/\gamma}}\} \text{ and } r_2 : \mathbb{R} \mapsto \mathbb{R} : \gamma \mapsto \max\{1, \frac{\gamma + 1 + 1/\gamma}{\sqrt{3 + 4/\gamma + 2/\gamma^2}}\} \ .$$

Both of these functions are equal to 1 for $\gamma \leq 1$ and strictly increasing for $\gamma \geq 1$, with the asymptotic approximations $r_1(\gamma) \approx \frac{\gamma}{\sqrt{3}}$ and $r_2(\gamma) \approx \frac{\gamma+1}{\sqrt{3}}$ when γ tends to $+\infty$.

Lemma 3.3 *Let us suppose F is a convex function with effective domain $C \subseteq \mathbb{R}^n_+$ and that there exists a constant γ such that*

$$\nabla^3 F(x)[h,h,h] \leq 3\gamma \nabla^2 F(x)[h,h] \sqrt{\sum_{i=1}^n \frac{h_i^2}{x_i^2}} \text{ for all } x \in \operatorname{int} C \text{ and } h \in \mathbb{R}^n \ . \tag{3.4}$$

We have that

$$F_1 : C \mapsto \mathbb{R} : x \mapsto F(x) - \sum_{i=1}^n \log x_i$$

satisfies the first condition of self-concordancy (2.1) *with parameter $\kappa_1 = r_1(\gamma)$ on its domain C and*

$$F_2 : C \times \mathbb{R} \mapsto \mathbb{R} : (x,u) \mapsto -\log\big(u - F(x)\big) - \sum_{i=1}^n \log x_i$$

satisfies the first condition of self-concordancy (2.1) *with parameter $\kappa_2 = r_2(\gamma)$ on its domain* $\operatorname{epi} F = \{(x,u) \mid F(x) \leq u\}$.

Note 3.2 *A similar lemma is proved in [2], with parameters κ_1 and κ_2 both equal to $1 + \gamma$. The second result is improved in [7], with κ_2 equal to* $\max\{1,\gamma\}$*, as a special case of a more general compatibility theory developed in [8]. However, it is easy to see that our result is better. Indeed, our parameters are strictly lower in all cases for F_1 and as soon as $\gamma > 1$ for r_2), with an asymptotical ratio of $\sqrt{3}$ when γ tends to $+\infty$.*

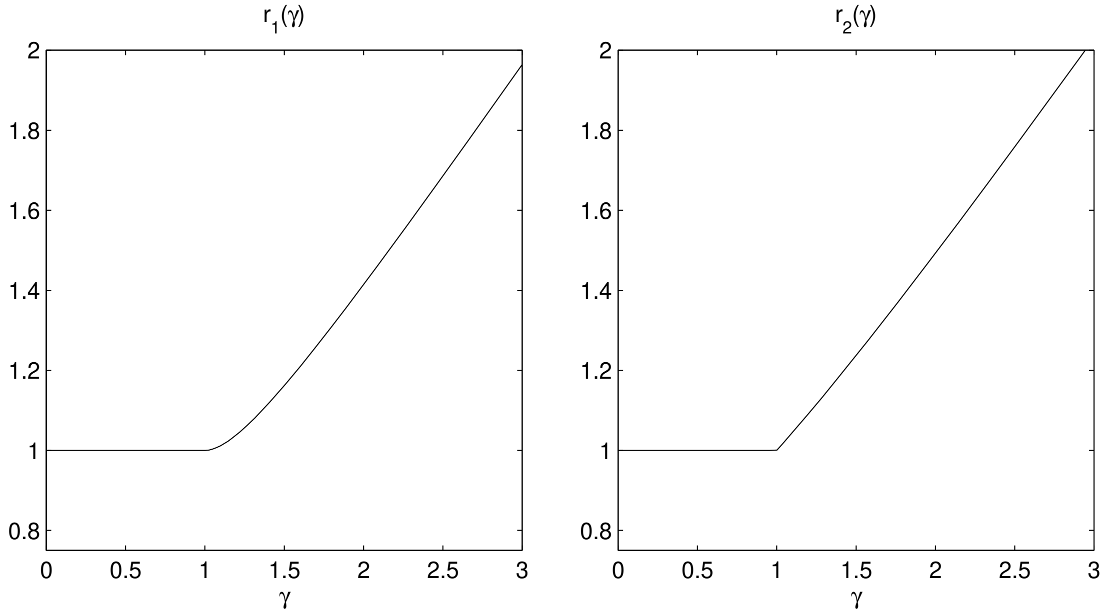

Figure 1: Graphs of functions r_1 and r_2

Proof. We follow the lines of [2] and start with F_1: computing its second and third differentials gives

$$\nabla^2 F_1(x)[h,h] = \nabla^2 F(x)[h,h] + \sum_{i=1}^n \frac{h_i^2}{x_i^2} \text{ and } \nabla^3 F_1(x)[h,h,h] = \nabla^3 F(x)[h,h,h] - 2\sum_{i=1}^n \frac{h_i^3}{x_i^3} \ .$$

Introducing two auxiliary variables $a \geq 0$ and $b \geq 0$ such that

$$a^2 = \nabla^2 F[h,h] \quad \text{ and } \quad b^2 = \sum_{i=1}^n \frac{h_i^2}{x_i^2}$$

(convexity of F guarantees that a is real), we can rewrite inequality (3.4) as $\nabla^3 F_1(x)[h,h,h] \leq 3\gamma a^2 b$. Combining it with the fact that

$$\left| \left(\sum_{i=1}^n \frac{h_i^3}{x_i^3} \right)^{\frac{1}{3}} \right| \leq \left(\sum_{i=1}^n \frac{|h_i|^3}{|x_i|^3} \right)^{\frac{1}{3}} \leq \left(\sum_{i=1}^n \frac{h_i^2}{x_i^2} \right)^{\frac{1}{2}} = b \ , \tag{3.5}$$

where the second inequality comes from the well-known relation $\|\cdot\|_3 \leq \|\cdot\|_2$ applied to vector $(\frac{h_1}{x_1}, \ldots, \frac{h_n}{x_n})$, we find that

$$\frac{\nabla^3 F_1(x)[h,h,h]}{2(\nabla^2 F_1(x)[h,h])^{\frac{3}{2}}} \leq \frac{3\gamma a^2 b + 2b^3}{2(a^2+b^2)^{\frac{3}{2}}} \ .$$

According to (2.1), finding the best parameter κ for F_1 amounts to maximize this last quantity depending on a and b. Since $a \geq 0$ and $b \geq 0$ we can write $a = r\cos\theta$ and $b = r\sin\theta$ with

$r \geq 0$ and $0 \leq \theta \leq \frac{\pi}{2}$, which gives

$$\frac{3\gamma a^2 b + 2b^3}{2(a^2+b^2)^{\frac{3}{2}}} = \frac{3\gamma}{2}\cos^2\theta\sin\theta + \sin^3\theta = h(\theta) \ .$$

The derivative of h is

$$h'(\theta) = \frac{3\gamma}{2}\cos^3\theta - 3\gamma\sin^2\theta\cos\theta + 3\cos\theta\sin^2\theta = 3\cos\theta\left(\frac{\gamma}{2}\cos^2\theta + (1-\gamma)\sin^2\theta\right) \ .$$

When $\gamma \leq 1$, this quantity is clearly always nonnegative, which implies that the maximum is attained for the largest value of θ, which gives $h_{max} = h(\frac{\pi}{2}) = 1 = r_1(\gamma)$. When $\gamma < 1$, we easily see that h has a maximum when $\frac{\gamma}{2}\cos^2\theta + (1-\gamma)\sin^2 = 0$. This condition is easily seen to imply $\sin^2\theta = \frac{\gamma}{3\gamma-2}$, and h_{max} becomes

$$\begin{aligned} h_{max} = 3\frac{\gamma}{2}\cos^2\theta\sin\theta + \sin^3\theta &= \big(3(\gamma-1)+1\big)\sin^3\theta \\ &= (3\gamma-2)\left(\frac{\gamma}{3\gamma-2}\right)^{\frac{3}{2}} = \frac{\gamma}{\sqrt{3-2/\gamma}} = r_1(\gamma) \ . \end{aligned}$$

A similar but slightly more technical proof holds for F_2. Letting $\tilde{x} = (x,u), \tilde{h} = (h,v)$ and $G(\tilde{x}) = F(x) - u$, we have that $F_2(\tilde{x}) = -\log(-G(\tilde{x}) - \sum_{i=1}^n \log x_i$. G is easily shown to be convex and negative on $\operatorname{epi} F$, the domain of F_2. Since F and G only differ by a linear term, we also have that $\nabla^2 F(x)[h,h] = \nabla^2 G(\tilde{x})[\tilde{h},\tilde{h}]$ and $\nabla^3 F(x)[h,h,h] = \nabla^3 G(\tilde{x})[\tilde{h},\tilde{h},\tilde{h}]$. Looking now at the second differential of F_2 we find

$$\nabla^2 F_2(\tilde{x})[\tilde{h},\tilde{h}] = \frac{\nabla G(\tilde{x})[h]^2}{G(\tilde{x})^2} - \frac{\nabla^2 G(\tilde{x})[\tilde{h},\tilde{h}]}{G(\tilde{x})} + \sum_{i=1}^n \frac{h_i^2}{x_i^2} \ .$$

Let us define for convenience $a \in \mathbb{R}_+$, $b \in \mathbb{R}_+$ and $c \in \mathbb{R}$ with

$$a^2 = -\frac{\nabla^2 G(\tilde{x})[\tilde{h},\tilde{h}]}{G(\tilde{x})}, \ b^2 = \sum_{i=1}^n \frac{h_i^2}{x_i^2} \text{ and } c = -\frac{\nabla G(\tilde{x})[h]}{G(\tilde{x})}$$

(convexity of G and the fact that it is negative on the domain of F_2 guarantee that a is real), which implies $\nabla^2 F_2(x)[\tilde{h},\tilde{h}] = a^2 + b^2 + c^2$. We can now evaluate the third differential

$$\begin{aligned} \nabla^3 F_2(\tilde{x})[\tilde{h},\tilde{h},\tilde{h}] &= -\frac{\nabla^3 G(\tilde{x})[\tilde{h},\tilde{h},\tilde{h}]}{G(\tilde{x})} + 3\frac{\nabla^2 G(\tilde{x})[\tilde{h},\tilde{h}]\nabla G(\tilde{x})[\tilde{h}]}{G(\tilde{x})^2} - 2\frac{\nabla G(\tilde{x})[\tilde{h}]^3}{G(\tilde{x})^3} - 2\sum_{i=1}^n \frac{h_i^3}{x_i^3} \\ &= -\frac{\nabla^3 G(\tilde{x})[\tilde{h},\tilde{h},\tilde{h}]}{G(\tilde{x})} + 3a^2c + 2c^3 - 2\sum_{i=1}^n \frac{h_i^3}{x_i^3} \\ &\leq -\frac{\nabla^3 G(\tilde{x})[\tilde{h},\tilde{h},\tilde{h}]}{\nabla^2 G(\tilde{x})[\tilde{h},\tilde{h}]}\frac{\nabla^2 G(\tilde{x})[\tilde{h},\tilde{h}]}{G(\tilde{x})} + 3a^2c + 2c^3 + 2b^3 \text{ using again (3.5)} \\ &= -\frac{\nabla^3 F(x)[h,h,h]}{\nabla^2 F(x)[h,h]}\frac{\nabla^2 G(\tilde{x})[\tilde{h},\tilde{h}]}{G(\tilde{x})} + 3a^2c + 2c^3 + 2b^3 \\ &\leq 3\gamma a^2 b + 3a^2c + 2c^3 + 2b^3 \text{ using condition (3.4)} \ . \end{aligned}$$

According to (2.1), finding the best parameter κ for F_2 amounts to maximize the following ratio

$$\frac{\nabla^3 F_2(\tilde{x})}{2(\nabla^3 F_2(\tilde{x}))^{\frac{3}{2}}} \leq \frac{3\gamma a^2 b + 3a^2 c + 2c^3 + 2b^3}{2(a^2+b^2+c^2)^{\frac{3}{2}}} = \frac{\frac{3\gamma}{2} a^2 b + \frac{3}{2} a^2 c + c^3 + b^3}{(a^2+b^2+c^2)^{\frac{3}{2}}} .$$

Since this last quantity is homogeneous of degree 0 with respect to variables a, b and c, we can assume that $a^2 + b^2 + c^2 = 1$, which gives

$$\frac{3\gamma}{2} a^2 b + \frac{3}{2} a^2 c + c^3 + b^3 = \frac{3}{2} a^2 (\gamma b + c) + c^3 + b^3 = \frac{3}{2}(1 - b^2 - c^2)(\gamma b + c) + b^3 + c^3 .$$

Calling this last quantity $m(b,c)$, we can now compute its partial derivatives with respect to b and c and find

$$\frac{\partial m}{\partial b} = -\frac{3}{2}\left((3\gamma - 2)b^2 + \gamma c^2 + 2bc - \gamma\right) \text{ and } \frac{\partial m}{\partial c} = -\frac{3}{2}\left(b^2 + c^2 + 2bc\gamma - 1\right) .$$

We have now to equate those two quantities to zero and solve the resulting system. We can for example write $\frac{\partial m}{\partial b} - \gamma \frac{\partial m}{\partial c} = 0$, which gives $(g-1)b(b - c(\gamma+1)) = 0$, and explore the resulting three cases. The solutions we find are

$$(b,c) = (0, \pm 1) \quad \text{and} \quad \left(\frac{\gamma+1}{\sqrt{3\gamma^2 + 4\gamma + 2}}, \frac{1}{\sqrt{3\gamma^2 + 4\gamma + 2}}\right)$$

with an additional special case $b + c = 1$ when $\gamma = 1$. Plugging these values into $m(b,c)$, one finds after some computations the following potential maximum values

$$\pm 1 \quad \text{and} \quad \frac{\gamma^2 + \gamma + 1}{\sqrt{3\gamma^2 + 4\gamma + 2}} = \frac{\gamma + 1 + 1/\gamma}{\sqrt{3 + +4/\gamma + 2/\gamma^2}}$$

(and 1 in the special case $\gamma = 1$). One concludes that the maximum we seek is equal to $r_2(\gamma)$, as announced. □

While the lemma we have just proved is useful to tackle the first condition of self-concordancy (2.1), it does not say anything about the second condition (2.2). The following Corollary about the second barrier F_2 might prove useful in this respect.

Corollary 3.1 *Let F satisfy the assumptions of Lemma 3.3. Then the second barrier*

$$F_2 : C \times \mathbb{R} \mapsto \mathbb{R} : (x,u) \mapsto -\log\bigl(u - F(x)\bigr) - \sum_{i=1}^{n} \log x_i$$

is $(r_2(\gamma), n+1)$-self-concordant.

Proof. Since $G(x,u) = F(x) - u$ is convex, $-\log(u - F(x)) = -\log(-G(x,u))$ is known to satisfy the second self-concordancy condition (2.2) with $\nu = 1$ by virtue of Lemma 3.1. Moreover, it is straightforward to check that each term $-\log x_i$ also satisfies that second condition with parameter $\nu = 1$. Using the addition Theorem 3.2 and combining with the result of Lemma 3.3, we can conclude that F_2 is $(r_2(\gamma), n+1)$-self-concordant. □

Note 3.3 *We would like to point out that no similar result can hold for the first function* F_1*, since we know nothing about the status of the second self-concordancy condition* (2.2) *on its first term* $F(x)$*. Indeed, taking the case of* $F : \mathbb{R}_+ \mapsto \mathbb{R} : x \mapsto \frac{1}{x}$*, we can check that* $\nabla^2 F(x)[h,h] = 2\frac{h^2}{x^3}$ *and* $\nabla^3 F(x)[h,h,h] = -6\frac{h^3}{x^4}$*, which implies that condition* (3.4) *is holds with* $\gamma = 1$ *since*

$$-6\frac{h^3}{x^4} \leq 3 \times 2\frac{h^2}{x^3}\frac{|h|}{x} \Leftrightarrow -h^3 \leq h^2\,|h|$$

is satisfied. On the other hand, the second self-concordancy condition (2.2) *cannot hold for* $F_1 : \mathbb{R}_+ \mapsto R : x \mapsto \frac{1}{x} - \log x$*, since*

$$\nabla F(x)^T(\nabla^2 F(x))^{-1}\nabla F(x) = \frac{F_1'(x)^2}{F_1''(x)} = \frac{\frac{(x+1)^2}{x^4}}{\frac{(2+x)}{x^3}} = \frac{(x+1)^2}{x(x+2)}$$

does not admit an upper bound (it tends to $+\infty$ *when* $x \to 0$*).*

To conclude this section, we mention that since condition (3.4) is invariant with respect to positive scaling of F, the results from Lemma 3.3 hold for barriers $F_{\lambda,1}(x) = \lambda F(x) - \sum_{i=1}^n \log x_i$ and $F_{\lambda,2}(x,u) = -\log(u - \lambda F(x)) - \sum_{i=1}^n \log x_i$ where λ is a positive constant.

4 Application to structured convex problems

In this section we rely on the work in [2], where several classes of structured convex optimization problems are shown to admit a self-concordant logarithmic barrier. However, Lemma 3.3 will allow us to improve the self-concordancy parameters and lower the resulting complexity values.

4.1 Extended entropy optimization

Let $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$. We consider the following problem

$$\min_{x\in\mathbb{R}^n} \; c^T x + \sum_{i=1}^n g_i(x_i) \quad \text{s.t.} \quad Ax = b \text{ and } x \geq 0 \tag{EEO}$$

where scalar functions $g_i : \mathbb{R}_+ \mapsto \mathbb{R}$ are required to satisfy

$$\left|g_i'''(x)\right| \leq \kappa_i \frac{g_i''(x)}{x} \;\forall x \geq 0 \tag{4.1}$$

(which by the way implies their convexity). Classical entropy optimization results as a special case when $g_i(x) = x\ln x$ (in that case, it is straightforward to see that condition (4.1) holds with $\kappa_i = 1$).

Let us use Lemma 3.3 with $F_i : x_i \mapsto g_i(x_i)$ and $\gamma = \frac{\kappa_i}{3}$. Indeed, checking condition (3.4) amounts to write

$$h^3 g_i'''(x) \leq 3\frac{\kappa_i}{3}h^2 g_i''(x)\frac{|h|}{x} \Leftrightarrow \frac{h}{|h|}g_i'''(x) \leq \kappa_i \frac{g_i''(x)}{x}\;,$$

which is guaranteed by condition (4.1). Using the second barrier and Corollary 3.1, we find that

$$F_i : (x_i, u_i) \mapsto -\log\bigl(u_i - g_i(x_i)\bigr) - \log x_i$$

is $(r_2(\frac{\kappa_i}{3}), 2)$-self-concordant[2]. However, in order to use this barrier to solve problem (EEO), we need to reformulate it as

$$\min_{x \in \mathbb{R}^n} c^T x + \sum_{i=1}^n u_i \quad \text{s.t.} \quad Ax = b,\ g_i(x_i) \le u_i \text{ and } x \ge 0 \ ,$$

which is clearly equivalent. We are now able to write the complete logarithmic barrier

$$F : (x, u) \mapsto -\sum_{i=1}^n \log\big(u_i - g_i(x_i)\big) - \sum_{i=1}^n \log x_i \ ,$$

which is $(r_2(\frac{\max\{\kappa_i\}}{3}), 2n)$-self-concordant by virtue of Theorem 3.2. In light of Note 3.1, we can even do better with a different scaling of each term, to get

$$\tilde{F} : (x, u) \mapsto -\sum_{i=1}^n r_2(\frac{\kappa_i}{3})^2 \log\big(u_i - g_i(x_i)\big) - \sum_{i=1}^n r_2(\frac{\kappa_i}{3})^2 \log x_i$$

which is then $(1, \sqrt{2\sum_{i=1}^n r_2(\frac{\kappa_i}{3})^2})$-self-concordant. In the case of classical entropy optimization, these parameters become $(1, \sqrt{2n})$, since $r_2(\frac{1}{3}) = 1$.

4.2 Dual geometric optimization

Let $\{I_k\}_{k=1\ldots r}$ be a partition of $\{1, 2, \ldots, n\}$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$. The dual geometric optimization problem is

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} c^T x + \sum_{k=1}^r \left[\sum_{i \in I_k} x_i \log\Big(\frac{x_i}{\sum_{i \in I_k} x_i}\Big) \right] \\ &\text{s.t. } Ax = b \text{ and } x \ge 0 \end{aligned} \qquad \text{(DGO)}$$

It is shown in [2] that condition (3.4) holds for

$$F_k : (x_i)_{i \in I_k} \mapsto \sum_{i \in I_k} x_i \log\Big(\frac{x_i}{\sum_{i \in I_k} x_i}\Big)$$

with $\gamma = 1$, so that the corresponding second barrier in Lemma 3.4 is $(1, |I_k| + 1)$-self-concordant. Using the same trick as for problem (EEO), we introduce additional variables u_k to find that the following barrier

$$F : (x, u) \mapsto \sum_{k=1}^r -\log\Big(u_k - \sum_{i \in I_k} x_i \log\big(\frac{x_i}{\sum_{i \in I_k} x_i}\big)\Big) - \sum_{i=1}^n \log x_i$$

is a $(1, n + r)$-self-concordant barrier for a suitable reformulation of problem (DGO).

[2] This corrects the statement in [2] where it is mentioned that $g_i(x_i) - \log x_i$, i.e. the first barrier in Lemma 3.3, is self-concordant. As it is made clear in Note 3.3, this cannot be true in general

4.3 l_p-norm optimization

Let $\{I_k\}_{k=1\ldots r}$ be a partition of $\{1, 2, \ldots, m\}$, $\eta \in \mathbb{R}^n$, $a_i \in \mathbb{R}^n$, $b_k \in \mathbb{R}^n$, $c \in \mathbb{R}^m$, $d \in \mathbb{R}^r$ and $p \in \mathbb{R}^m$ such that $p_i \geq 1$. The primal l_p-norm optimization problem is

$$\max_{x \in \mathbb{R}^n} \eta^T x \quad \text{s.t.} \quad f_i(x) \leq 0 \text{ for all } k = 1, \ldots, r \ , \tag{LPO}$$

where functions $f_i : \mathbb{R}^n \mapsto \mathbb{R}$ are defined according to

$$f_i : x \mapsto \sum_{i \in I_k} \frac{1}{p_i} \left| a_i^T x - c_i \right|^{p_i} + b_k^T x - d_k \ .$$

This problem can be reformulated as

$$\max_{x \in \mathbb{R}^n} \eta^T x \quad \text{s.t.} \quad \begin{cases} \left| a_i^T x - c_i \right| & \leq & s_i & \forall i = 1, \ldots, m \\ s_i & \leq & t_i^{1/p_i} & \forall i = 1, \ldots, m \\ \sum_{i \in I_k} \frac{t_i}{p_i} & \leq & d_k - b_k^T x & \forall k = 1, \ldots, r \end{cases}$$

where each of the m constraints involving an absolute value is indeed equivalent to a pair of linear constraints $a_i^T x - c_i \leq u_i$ and $c_i - a_i^T x \leq u_i$. Once again, a self-concordant function can be found for the difficult part of the constraints, i.e. the nonlinear inequality $s_i \leq t_i^{1/p_i}$. Indeed, it is straightforward to check that $f_i : t_i \mapsto -t_i^{1/p_i}$ satisfies condition (3.4) with $\gamma = \frac{2p_i - 1}{3p_i} < 1$, which implies in the same fashion as above that

$$-\log\left(t_i^{1/p_i} - u_i\right) - \log t_i$$

is $(1, 2)$-self-concordant. Combining with the logarithmic barrier for the linear constraints, we have that

$$-\sum_{i=1}^m \log(u_i - a_i^T x + c_i) - \sum_{i=1}^m \log(u_i + a_i^T x - c_i) - \sum_{i=1}^m \log\left(t_i^{1/p_i} - u_i\right) \ldots$$
$$\ldots - \sum_{i=1}^m \log t_i - \sum_{k=1}^r \log\left(d_k - b_k^T x - \sum_{i \in I_k} \frac{t_i}{p_i}\right)$$

is $(1, 4m + r)$-self-concordant for our reformulation of problem (LPO) (since each linear constraint is $(1, 1)$-self-concordant).

Let us mention that another reformulation is presented in [2], where Lemma 3.3 is applicable to the nonlinear constraint with parameter $\gamma = \frac{|p_i - 2|}{3}$, with the disadvantage of having a parameter that depends on p_i (although $r_2(\gamma)$ will stay at its lowest value as long as $p_i \leq 5$). We conclude this section by mentioning that very similar results hold for the dual l_p-norm optimization problem[3].

[3]However, we would like to point out that the nonlinear function involved in these developments is wrongly stated to satisfy condition (3.4) with $\gamma = \frac{\sqrt{2}(q_i + 1)}{3q_i}$, while a correct value is $\frac{\sqrt{5q_i^2 - 2q_i + 2}}{3q_i}$.

5 Concluding remarks

We have presented in this paper an introduction to the theory of self-concordant functions. We would like to point out that this very powerful framework relies on two different conditions (2.1) and (2.2) and the two corresponding parameters κ and ν, each with its own purpose (see the discussion in Note 2.2). However, the important quantity is the resulting complexity value $\kappa\sqrt{\nu}$, which is of the same order as the number of iterations that is needed to reduce the barrier parameter by a constant factor by the short-step interior-point algorithm.

It is possible to scale self-concordant barriers such that one of the parameters is arbitrarily fixed without any real loss of generality. We have shown that this is best done fixing parameter κ, considering the way the complexity value is affected when adding several self-concordant barriers. However, it is in our opinion better to keep two parameters all the time, in order to simplify the presentation (for example, Lemma 3.3 intrinsically deals with the κ parameter and would need a rather awkward reformulation to be written for parameter ν with κ fixed to 1).

Several important results help us prove self-concordancy of barrier functions: Lemmas 3.1 and 3.2 deal with the second self-concordancy condition (2.2), while our improved Lemma 3.3 pertains to the first self-concordancy condition (2.1). They are indeed responsible for most of the analysis carried out in Section 4, which is dedicated to several classes of structured convex optimization problems. Namely, it is proved that nearly all the nonlinear (i.e. corresponding to the nonlinear constraints) terms in the associated logarithmic barriers are self-concordant with $\kappa = 1$ (the exception being extended entropy optimization, which encompasses a very broad class of problems). We would also like to mention that since all the barriers that are presented are polynomially computable, as well as their gradient and Hessian, the short-step method applied to any of these problems would need to perform a polynomial number of arithmetic operations to provide a solution with a given accuracy.

To conclude, we would like to speculate on the possibility of replacing the two self-concordancy conditions by a single inequality. Indeed, since the complexity value $\kappa\sqrt{\nu}$ is the only quantity that really matters in the final complexity result, one could imagine to consider the following inequality

$$\frac{F'''_{x,h}(0)F'_{x,h}(0)}{F''_{x,h}(0)^2} \leq 2C \text{ for all } x \in \operatorname{int} C \text{ and } h \in \mathbb{R}^n \ , \tag{5.1}$$

which is satisfied with $C = \kappa\sqrt{\nu}$ for (κ, ν)-self-concordant barriers (to see that, simply multiply condition (2.4b) by the square root of condition (2.5b)). We point out the following two intriguing facts and leave their investigation for further research:

- ⋄ Condition (5.1) appears to be central in the recent theory of *self-regular* functions [9], an attempt at generalizing self-concordant functions.
- ⋄ Following the same principles as for (2.4d) and (2.5d), condition (5.1) can be reformulated as

$$\left(-\frac{F'_{x,h}(t)}{F''_{x,h}(t)}\right)' \leq 2C - 1 \ ,$$

where the quantity on the left-hand side is the derivative of the Newton step applied to the restriction $F_{x,h}$.

References

[1] J. Brinkhuis, Communication at the International Symposium on Mathematical Programming, Atlanta, August 2000.

[2] D. den Hertog, F. Jarre, C. Roos, and T. Terlaky, *A sufficient condition for self-concordance with application to some classes of structured convex programming problems*, Mathematical Programming, Series B **69** (1995), no. 1, 75–88.

[3] D. den Hertog, C. Roos, and T. Terlaky, *On the classical logarithmic barrier method for a class of smooth convex programming problems*, Journal of Optimization Theory and Applications **73** (1992), no. 1, 1–25.

[4] A. V. Fiacco and G. P. McCormick, *Nonlinear programming: Sequential unconstrained minimization techniques*, John Wiley & Sons, New York, 1968, Reprinted in *SIAM Classics in Applied Mathematics*, SIAM Publications, 1990.

[5] K. R. Frisch, *The logarithmic potential method of convex programming*, Tech. report, University Institute of Economics, Oslo, Norway, 1955.

[6] F. Jarre, *The method of analytic centers for smooth convex programs*, Dissertation, Institut für Angewandte Mathematik und Statistik, Universität Wärzburg, Germany, 1989.

[7] ———, *Interior-point methods for classes of convex programs*, Interior Point Methods of Mathematical Programming (T. Terlaky, ed.), Applied Optimization, vol. 5, Kluwer Academic Publishers, 1996, pp. 255–296.

[8] Y. E. Nesterov and A. S. Nemirovsky, *Interior-point polynomial methods in convex programming*, SIAM Studies in Applied Mathematics, SIAM Publications, Philadelphia, 1994.

[9] J. Peng, C. Roos, and T. Terlaky, *Self-regular proximities and new search directions for linear and semidefinite optimization*, Technical report, Department of Computing and Software, McMaster University, Hamilton, Ontario, Canada, March 2000, submitted to Mathematical Programming.

[10] J. Renegar, *A mathematical view of interior-point methods in convex optimization*, to be published by in the MPS/SIAM Series on Optimization, SIAM, New York, 2000.

[11] R. T. Rockafellar, *Convex analysis*, Princeton University Press, Princeton, N. J., 1970.

[12] C. Roos and T. Terlaky, *Nonlinear optimization*, Delft University of Technology, The Netherlands, 1998, Course WI387.

[13] J. Stoer and Ch. Witzgall, *Convexity and optimization in finite dimensions I*, Springer Verlag, Berlin, 1970.