Improving complexity of structured convex optimization problems using self-concordant barriers

IMAGE TECHNICAL REPORT 0102

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.

August 2001

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

Abstract

The purpose of this paper is to provide improved complexity results for several classes of structured convex optimization problems using to the theory of self-concordant functions developed in [11]. We describe the classical short-step interior-point method and optimize its parameters in order to provide the best possible iteration bound. We also discuss the necessity of introducing two parameters in the definition of self-concordancy and which one is the best to fix. A lemma from [3] is improved, which allows us to review several classes of structured convex optimization problems and improve the corresponding complexity results.

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

Contents

1 Introduction **3**

2 Self-concordancy and a polynomial short-step method **4**
2.1 Equivalent definitions 4
2.2 Optimal complexity of a short-step method 6

3 Using self-concordancy **10**
3.1 Barrier calculus 10
3.2 Fixing a parameter 11
3.3 A useful lemma 12

4 Application to structured convex problems **17**
4.1 Extended entropy optimization 18
4.2 Dual geometric optimization 19
4.3 l_p-norm optimization 19

5 Concluding remarks **20**

1 Introduction

Convex optimization deals with the minimization of a convex function on a convex set, see e.g. [15, 18]. It is well-known that a problem involving a nonlinear objective function can be rewritten using a linear objective function, an additional constraint and one more variable, so that one can consider without loss of generality the following formulation

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

where the feasible set $\mathcal{C} \subseteq \mathbb{R}^n$ is convex and vector $c \in \mathbb{R}^n$ defines the objective function.

We will also assume that $\mathcal{C}$ is full-dimensional, i.e. possesses a nonempty interior. This can be again be done without loss of generality, since a suitable change of variables can make the feasible set full-dimensional when the affine hull of $\mathcal{C}$ is a linear subspace different from $\mathbb{R}^n$ (namely, supposing the dimension of the affine hull of the feasible set is equal to $m < n$, one can use any affine transformation mapping $\operatorname{aff} \mathcal{C}$ to $\mathbb{R}^m$, see e.g. [15]).

Among the different types of algorithms that can be applied to solve problem (CL), *interior-point methods* have gained a lot of popularity in the last two decades, mainly because one can both give a polynomial bound on the number of arithmetic operations they need to reach a solution within a given accuracy and implement them to solve successfully real-world problems, especially in the fields of linear [1] (where they compare favourably with the simplex method on large-scale problems), quadratic and semidefinite optimization.

One traditional way to define $\mathcal{C}$ is to provide a list of convex constraints, i.e. $\mathcal{C} = \left\{ x \in \mathbb{R}^n \mid f_i(x) \leq 0 \ \forall 1 \leq i \leq m \right\}$ where each of the m functions $f_i : \mathbb{R}^n \mapsto \mathbb{R}$ is convex. However, interior-point methods rely instead on the notion of barrier function for the feasible set $\mathcal{C}$, according to the following definition.

Definition 1.1 *Let $\mathcal{C}$ be a full-dimensional convex subset of $\mathbb{R}^n$. A function $F : \operatorname{int} \mathcal{C} \mapsto \mathbb{R}$ is a* barrier function *for the set $\mathcal{C}$ if and only if F is three times continuously differentiable, strictly convex and $F(x)$ tends to $+\infty$ whenever x tends to $\partial\mathcal{C}$, the boundary of $\mathcal{C}$.*

Note that it is often possible to provide a suitable barrier function for a convex set $\mathcal{C}$ defined by convex constraints, using the *logarithmic barrier* [7] defined as

$$F : \operatorname{int} \mathcal{C} \mapsto \mathbb{R} : x \mapsto F(x) = -\sum_{i=1}^{m} \log(-f_i(x)) \ .$$

We have however to make sure the resulting function satisfies all the properties of a barrier function (for example, defining $\mathcal{C} = \mathbb{R}_+$ with a single convex constraint using $f_1 : \mathbb{R}_+ \mapsto \mathbb{R} : x \mapsto -x^x$ leads to $F(x) = -x \log x$, which does not possess the barrier property).

Problem (CL) is thus replaced by a family of unconstrained minimization problems relying on the barrier function

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

parameterized by a *barrier parameter* $\mu \in \mathbb{R}_{++}$ (see the classical monograph [6]). Most interior-point methods then follow the so-called central path, i.e. the set of minimizers $x(\mu)$ of each of these problems[1], while decreasing μ to zero, which under some mild conditions leads them to an optimal solution of (CL).

[1] Uniqueness of these minimizers is guaranteed by the strict convexity of F, while their existence requires some additional assumptions, see [6].

Interior-point methods rely on Newton's method to compute these minimizers, so that the following question becomes of interest: is it possible to choose a barrier function F such that Newton's method is provably efficient in solving subproblems (CL_μ), leading to a polynomial algorithm for problem (CL) ? This crucial question is thoroughly answered by the theory of self-concordancy, first developed by Nesterov and Nemirovsky [11], which we will introduce in the next Section 2.

2 Self-concordancy and a polynomial short-step method

2.1 Equivalent definitions

Let us first present a definition of a self-concordant barrier[2].

Definition 2.1 *Let $\mathcal{C}$ be a full-dimensional convex subset of $\mathbb{R}^n$. A function $F : \operatorname{int}\mathcal{C} \mapsto R$ is called a* (κ,ν)-self-concordant *barrier for $\mathcal{C}$ if and only if F is a barrier function for $\mathcal{C}$ and the following two conditions hold for all $x \in \operatorname{int}\mathcal{C}$ and $h \in \mathbb{R}^n$:*

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

$$\nabla F(x)^T (\nabla^2 F(x))^{-1} \nabla F(x) \leq \nu \tag{2.2}$$

(the square root in (2.1) is well defined because of the requirement that F is convex).

We point out that no absolute value is needed in the left-hand side of (2.1): the apparently stronger condition

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

required by some authors is not needed since inequality (2.1) also has to hold in the direction opposite to h, which gives (using the fact that the n^{th}-order differential is homogeneous with degree n)

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

and shows that conditions (2.1) and (2.3) are equivalent.

Given a barrier function F and a point x belonging to its domain, it will be convenient to introduce the Newton step $n(x)$ and the intrinsic inner product $\langle\cdot,\cdot\rangle_x$[3]:

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

The intrinsic norm $\|\cdot\|_x$ will also be used according to the usual definition $\|a\|_x = \sqrt{\langle a,a\rangle_x}$.

Conditions (2.1) and (2.2) admit several equivalent reformulations, sometimes easier to work with, based on the restriction of F along a line. Let $x \in \operatorname{int}\mathcal{C}$ and $h \in \mathbb{R}^n$, defining the line $\{x+th \mid t \in \mathbb{R}\}$ and let us introduce the one-dimensional function $F_{x,h} : \mathbb{R} \mapsto \mathbb{R} : t \mapsto F(x+th)$.

[2] This definition does not exactly match the definition of a self-concordant barrier originally stated in [11], but corresponds in fact to the notion of strongly non-degenerate κ^{-2}-self-concordant barrier functional with parameter ν. Our main restriction with respect to the original theory of [11] is the requirement that F be strictly convex, equivalent to the notion of non-degeneracy in [11], which is essentially needed for technical reasons and does not really reduce the generality of the approach (see the discussion at the end of [14, Chapter 2.5]).

[3] Renegar [14] points out that while the gradient and Hessian of F, i.e. its first- and second-order differentials do depend from the choice of reference inner product used in their definition, the intrinsic norm and inner products as well as the Newton step are in fact independent from this choice, so that their use leads to an essentially coordinate-free development of the theory.

Theorem 2.1 *The following four conditions are equivalent:*

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

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

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

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

Theorem 2.2 *The following four conditions are equivalent:*

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

Proof. We first show that condition (2.5b) implies condition (2.5a).

$$\begin{aligned}
\left(\nabla F(x)^T(\nabla^2 F(x))^{-1}\nabla F(x)\right)^2 &= F'_{x,n(x)}(0)^2\\
&\leq \nu F''_{x,n(x)}(0) = \nu \nabla F(x)^T(\nabla^2 F(x))^{-1}\nabla F(x)\ ,
\end{aligned}$$

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

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

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

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

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 [11, 10, 14]. Conditions (2.4d) and (2.5d) are less commonly seen (they were however mentioned in [2]).

2.2 Optimal complexity of a short-step method

Interior-point methods for convex optimization rely on a barrier function and the associated central path to solve problem (CL). Iterates x_k will be required to lie in a prescribed neighbourhood of the central path (since computing points that lie exactly on the central is too costly) while the barrier parameter μ_k will tend to zero.

A good measure of the proximity to the central path would be $\|x_k - x(\mu_k)\|_{x_k}$, but the fact it involves the unknown central point $x(\mu_k)$ makes it difficult to use in the analysis. We will instead rely on the following elegant proximity measure: we define $\delta(x,\mu)$, a measure of the proximity of x to the central point $x(\mu)$, to be the intrinsic norm of $n_\mu(x)$, the Newton step trying to minimize the objective in problem (CL_μ) (aiming thus at $x(\mu_k)$), i.e.

$$n_\mu(x) = -(\nabla^2 F(x))^{-1}(\frac{c}{\mu} + \nabla F(x)) = -\frac{1}{\mu}(\nabla^2 F(x))^{-1}c + n(x) \text{ and } \delta(x,\mu) = \|n_\mu(x)\|_x$$

(note that both $n_\mu(x)$ and $\|n_\mu(x)\|_x$ are independent of the coordinate system).

The principle lying behind a short-step interior-point method is to trace the central path approximately, ensuring that the proximity $\delta(x_k,\mu_k)$ at each iteration is kept below a predefined bound. Given a problem of type (CL), a barrier function F for the feasible set $\mathcal{C}$, 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:

(1) $\mu_{k+1} \leftarrow \mu_k(1-\theta)$ (decrease the barrier parameter)
(2) $x_{k+1} \leftarrow x_k + n_{\mu_{k+1}}(x_k)$ (take a Newton step)
(3) $k \leftarrow k+1$

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 at each iteration. Indeed, it is precisely the self-concordancy of the barrier function that will guarantee that such a choice is possible. Our goal now is to find optimal values for this pair of parameters, i.e. leading to a minimum number of iterations.

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

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

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

$$\|n(x)\|_x = \sqrt{\nabla F(x)^T(\nabla^2 F(x))^{-1}\nabla F(x)} \leq \sqrt{\nu}\,,$$

because of condition (2.2). □

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 [10, 11, 14]).

Note 2.1 *The reason why the self-concordancy property relies on two separate conditions becomes clear: 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 barrier, we are now going to find a suitable pair of parameters τ and θ providing the greatest reduction for the parameter μ at each iteration, in other words maximizing 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

$$\delta^+ \leq \frac{\kappa\delta'^2}{(1-\kappa\delta')^2} \quad \text{and therefore require that} \quad \frac{\kappa\delta'^2}{(1-\kappa\delta')^2} \leq \tau .$$

This is equivalent to

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

(which 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 $\beta = \sqrt{\kappa\tau}$ and $\Gamma = \kappa\sqrt{\nu}$ (we call this quantity the *complexity value* of the barrier since it plays an important role in the complexity analysis), we have

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

which finally means that maximum the value of θ that guarantees $\delta^+ \leq \tau$ is given by

$$\theta = \frac{1-\beta-\beta^2}{1+\Gamma+\frac{\Gamma}{\beta}} \,. \tag{2.7}$$

Computing the derivative of this quantity with respect to β shows that it is maximum when β satisfies $2\beta^3(1+\Gamma)+\beta^2(1+4\Gamma)+2\beta\Gamma-\Gamma=0$ or equivalently when

$$\Gamma = -\frac{\beta^2(2\beta+1)}{2\beta^3+4\beta^2+2\beta-1} \,, \tag{2.8}$$

the corresponding graph being depicted on Figure 1. Possible values for Γ ranging from 1 to $+\infty$ (proof of the nontrivial fact that Γ is never less than 1 can be found in [11]), straightforward computations show that the corresponding values of β range from $\beta_- \approx 0.273$ to $\beta^+ \approx 0.297$ (satisfying respectively $4\beta_-^3+5\beta_-^2+2\beta_--1=0$ and $2\beta_+^3+4\beta_+^2+2\beta_+-1=0$). Unfortunately, the corresponding explicit analytical expression $\beta = h(\Gamma)$ is too complicated[4] to be stated here. However, leaving β as a function of Γ, we can evaluate the maximal value of θ as follows: plugging the value of Γ given by (2.8) into (2.7) gives after some manipulations the simple expression $\theta = 1-2\beta-4\beta^2-2\beta^3$, which using again (2.8) provides the equality

$$\theta = \frac{\beta^2(2\beta+1)}{\Gamma} \,, \tag{2.9}$$

which will prove useful later.

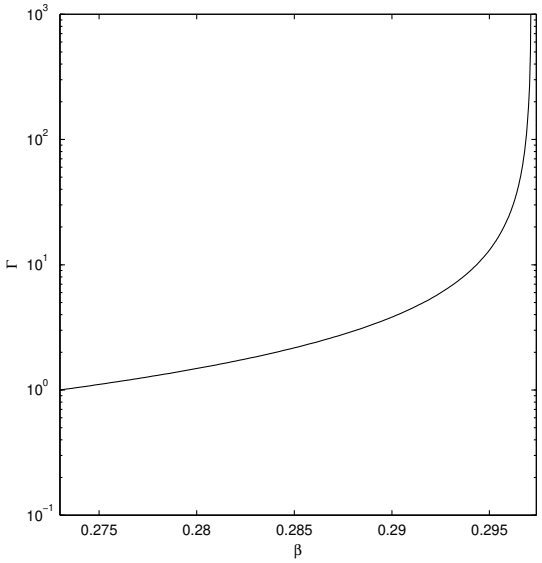

Figure 1: Graph of the relation $\beta = h(\Gamma)$.

Before we state our final complexity result, we have to discuss termination of the algorithm. The most practical stopping criterion is a small target value μ_e for the barrier parameter, which directly relates to the iteration bound given in (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

[4] It is indeed the real root of a cubic equation and can be computed without effort for any value of Γ.

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 $c^T x^*$

$$c^T x_e - c^T x^* \leq \frac{\mu_e}{1 - 3\kappa\tau} \kappa\sqrt{\nu} \tag{2.10}$$

(proof of this fact is omitted here and can easily be obtained combining Theorems 2.2.5 and 2.3.3 in [14]). 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* $\mathcal{C}$, *a constant* $\beta = h(\kappa\sqrt{\nu})$ *and an initial iterate* x_0 *such that* $\delta(x_0, \mu_0) < \frac{\beta^2}{\kappa}$, *one can find a solution with accuracy* ϵ *in*

$$\left\lceil \Big(\frac{\kappa\sqrt{\nu}}{\beta^2(2\beta+1)} - \frac{1}{2} \Big) \log \frac{\mu_0 \kappa\sqrt{\nu}}{\epsilon(1-3\beta^2)} \right\rceil \quad \textit{iterations.}$$

Proof. Using the optimal value $\beta = h(\kappa\sqrt{\nu})$, the maximal value for θ given by (2.7), the fact that $\kappa\tau = \beta^2$ and the bound on the objective accuracy in (2.10), we find that the stopping threshold on the barrier parameter μ_e must satisfy

$$\frac{\mu_e}{1-3\beta^2} \kappa\sqrt{\nu} \leq \epsilon \Leftrightarrow \mu_e \leq \frac{\epsilon(1-3\beta^2)}{\kappa\sqrt{\nu}} \; .$$

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

$$\begin{aligned} \frac{1}{\log(1-\theta)} \log \frac{\mu_e}{\mu_0} &\leq \frac{1}{\log(1-\theta)} \log \frac{\epsilon(1-3\beta^2)}{\mu_0 \kappa\sqrt{\nu}} = -\frac{1}{\log(1-\theta)} \log \frac{\mu_0 \kappa\sqrt{\nu}}{\epsilon(1-3\beta^2)} \\ &\leq \big(\frac{1}{\theta} - \frac{1}{2}\big) \log \frac{\mu_0 \kappa\sqrt{\nu}}{\epsilon(1-3\beta^2)} = \Big(\frac{\kappa\sqrt{\nu}}{\beta^2(2\beta+1)} - \frac{1}{2} \Big) \log \frac{\mu_0 \kappa\sqrt{\nu}}{\epsilon(1-3\beta^2)} \; , \end{aligned}$$

as announced (the second inequality uses the fact that $\frac{1}{\log(1-\theta)} \geq \frac{1}{2} - \frac{1}{\theta}$, which can be easily derived using the Taylor series of $\log x$ around 1, while the equality uses equation (2.9)). □

This result is not very explicit since it still involves the constant β. However, using the fact that β is lower bounded by $\beta_- \approx 0.273$, one can easily derive the following corollary:

Corollary 2.1 *Given a convex optimization problem* (CL), *a* (κ, ν)*-self-concordant barrier* F *for* $\mathcal{C}$ *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*

$$\left\lceil \big(8.68\kappa\sqrt{\nu} - \frac{1}{2} \big) \log \frac{1.29\mu_0 \kappa\sqrt{\nu}}{\epsilon} \right\rceil \quad \textit{iterations.}$$

It is also interesting to note that the value of the coefficient of $\kappa\sqrt{\nu}$ in the leading factor of the number of iterations can be further lowered if a larger lower bound on $\kappa\sqrt{\nu}$ is assumed. Indeed, coefficient 8.68 becomes 7.90 if $\kappa\sqrt{\nu} \geq 2$, 7.43 if $\kappa\sqrt{\nu} \geq 5$, 7.27 if $\kappa\sqrt{\nu} \geq 10$ and tends to 7.10 when $\kappa\sqrt{\nu}$ tends to $+\infty$. This improves several similar results from the literature, e.g. $9\kappa\sqrt{\nu}$ in [10] and $1 + 8\kappa\sqrt{\nu}$ in [14].

Moreover, one can compare these values with the corresponding ones valid in the special case of linear optimization. Recalling that $\kappa\sqrt{\nu}$ is equal to $\sqrt{n}$ in the case of a linear optimization problem involving n nonnegative variables and equipped with the standard logarithmic

barrier, we see that a similar short-step algorithm described in [17, Theorem II.24][5] achieves an iteration bound featuring a leading factor equal to $3\sqrt{n}$, noticeably lower than our result for the general convex case.

3 Using 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 contained in [11] states that every convex set in $\mathbb{R}^n$ admits a $(\mathcal{K}, n)$-self-concordant barrier, where $\mathcal{K}$ is a universal constant (independent of n). However, the universal barrier exhibited in their proof is defined as a volume integral over an n-dimensional convex body, and is therefore very 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 very well take a number of arithmetic operations that grows exponentially with n, which would lead to an exponential total algorithmic complexity for the short-step method, despite the polynomial iteration bound.

Another approach to find self-concordant barriers, termed *barrier calculus* in [11], consists in combining basic self-concordant barriers using operations that are known to preserve self-concordancy. We now proceed to briefly 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 $\mathcal{C} \subseteq \mathbb{R}^n$ and $\lambda \in \mathbb{R}_{++}$ a positive scalar. Then (λF) is also a self-concordant barrier for $\mathcal{C}$ with parameters $(\frac{\kappa}{\sqrt{\lambda}}, \lambda\nu)$.*

Proof. It is clear that (λF) is also a barrier function (i.e. smoothness, strict convexity and the barrier property are obviously preserved by scaling). Since F is (κ, ν)-self-concordant, we have for all $x \in \operatorname{int}\mathcal{C}$ and $h \in \mathbb{R}^n$ (using Theorems 2.1 and 2.2)

$$F'''_{x,h}(0) \leq 2\kappa F''_{x,h}(0)^{\frac{3}{2}} \text{ and } F'_{x,h}(0)^2 \leq \nu F''_{x,h}(0)$$

which is clearly equivalent to

$$(\lambda F)'''_{x,h}(0) \leq 2\frac{\kappa}{\sqrt{\lambda}}\big((\lambda F)''_{x,h}(0)\big)^{\frac{3}{2}} \text{ and } \big((\lambda F)'_{x,h}(0)\big)^2 \leq \lambda\nu(\lambda F)''_{x,h}(0)\ ,$$

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

[5] Using a primal-dual algorithm, [17, Remark II.54] further lowers this bound to achieve a leading factor equal to $\sqrt{n}$ as soon as $n \geq 8$.

together in the expression $\kappa\sqrt{\nu}$. This quantity, which we 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 $\mathcal{C}_1 \subseteq \mathbb{R}^n$ and G be a (κ_2, ν_2)-self-concordant barrier for $\mathcal{C}_2 \subseteq \mathbb{R}^n$. Then $(F + G)$ is a self-concordant barrier for $\mathcal{C}_1 \cap \mathcal{C}_2$ (provided this intersection is nonempty) with parameters $(\max\{\kappa_1, \kappa_2\}, \nu_1 + \nu_2)$.*

Proof. It is straightforward to see that $(F + G)$ is a barrier function for $\mathcal{C}_1 \cap \mathcal{C}_2$. We write

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

and

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

which is exactly 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 [3, 4, 9, 16] 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 list convex constraints and rely on the logarithmic barrier described in the introduction. The following lemma will prove useful.

Lemma 3.1 *Let $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{+\infty\}$ be a convex function, $\mathcal{C} = \{x \in \mathbb{R}^n \mid f(x) \leq 0\}$ and define the logarithmic barrier $F : \operatorname{int}\mathcal{C} \mapsto \mathbb{R} \cup \{+\infty\} : x \mapsto -\log(-f(x))$. 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 the following quantities for $x \in \operatorname{int}\mathcal{C}$, $h \in \mathbb{R}^n$ 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 $\mathcal{C}$), which states that F satisfies the second self-concordancy condition (2.2) with $\nu = 1$. □

Since the logarithmic barrier corresponding to a set defined by several convex constraints 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 $\nu = m$, 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. [10, 11, 14]), and we 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 $(\kappa_i, 1)$-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 may 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 becoming completely irrelevant cannot 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 really depending on the parameters of all the terms of the logarithmic barrier.*

3.3 A useful lemma

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 a lemma that can help us in that regard. Let us first introduce two auxiliary functions r_1 and r_2, whose graphs are depicted in Figure 2:

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

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 $+\infty$.

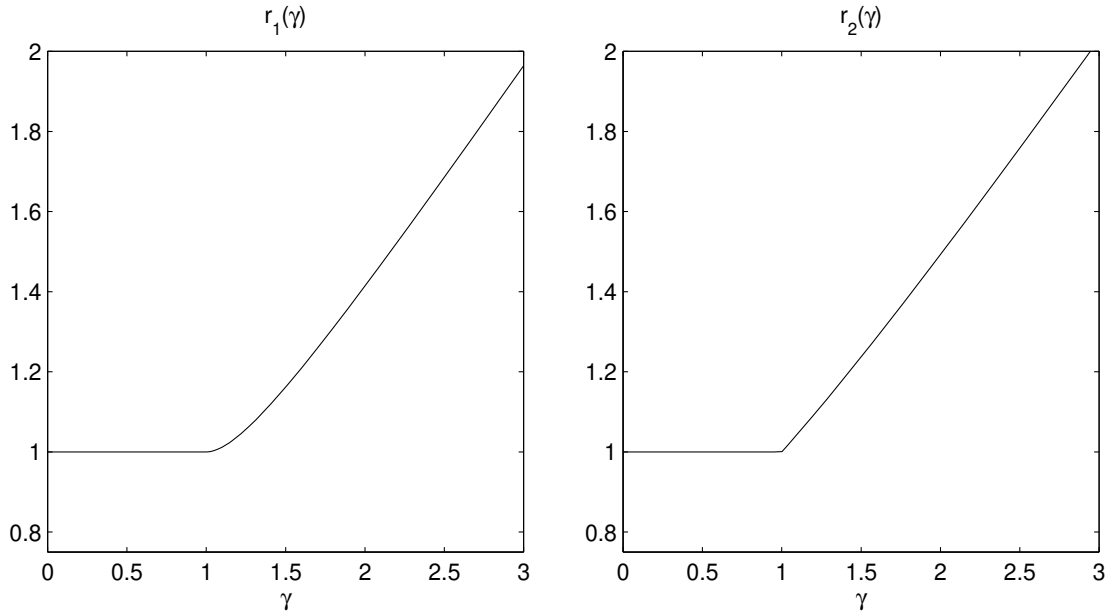

Figure 2: Graphs of functions r_1 and r_2

Lemma 3.2 *Let us suppose F is a convex function defined on $\operatorname{int}\mathcal{C} \subseteq \mathbb{R}^n_{++}$ and that there exists a constant γ such that*

$$\nabla^3 F(x)[h,h,h] \leq 3\gamma \nabla^2 F(x)[h,h] \sqrt{\sum_{i=1}^n \frac{h_i^2}{x_i^2}} \text{ for all } x \in \operatorname{int}\mathcal{C} \text{ and } h \in \mathbb{R}^n . \tag{3.1}$$

We have that

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

satisfies the first condition of self-concordancy (2.1) *with parameter $\kappa_1 = r_1(\gamma)$ on its domain and, defining* $\operatorname{epi} F = \{(x,u) \in \mathcal{C} \times \mathbb{R} |\ F(x) \leq u\}$

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

satisfies the first condition of self-concordancy (2.1) *with parameter $\kappa_2 = r_2(\gamma)$ on its domain.*

Note 3.2 *This lemma originates from a similar result is proved in [3] with parameters* κ_1 *and* κ_2 *both equal to* $1+\gamma$*. The second result is improved in [10], with* κ_2 *equal to* $\max\{1,\gamma\}$*, as a special case of a more general compatibility theory developed in [11]. 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* F_2*, with an asymptotical ratio of* $\sqrt{3}$ *when* γ *tends to* $+\infty$*.*

Proof. We follow the lines of [3] 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 $a \in \mathbb{R}_+$ and $b \in \mathbb{R}_+$ 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 ensures that a is real), we rewrite condition (3.1) as $\nabla^3 F(x)[h,h,h] \leq 3\gamma a^2 b$. Combining it with the fact that

$$\left| \left(\sum_{i=1}^n \frac{h_i^3}{x_i^3} \right)^{\frac{1}{3}} \right| \leq \left(\sum_{i=1}^n \frac{|h_i|^3}{|x_i|^3} \right)^{\frac{1}{3}} \leq \left(\sum_{i=1}^n \frac{h_i^2}{x_i^2} \right)^{\frac{1}{2}} = b\,, \tag{3.2}$$

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 derivative 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 $\frac{\gamma}{2}\cos^2\theta + (1-\gamma)\sin^2\theta = 0$. 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 $\operatorname{int} \operatorname{epi} F$, 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 again 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

$$\begin{aligned}
\nabla^3 F_2(\tilde{x})[\tilde{h},\tilde{h},\tilde{h}] &= -\frac{\nabla^3 G(\tilde{x})[\tilde{h},\tilde{h},\tilde{h}]}{G(\tilde{x})} + 3\frac{\nabla^2 G(\tilde{x})[\tilde{h},\tilde{h}]\nabla G(\tilde{x})[\tilde{h}]}{G(\tilde{x})^2} - 2\frac{\nabla G(\tilde{x})[\tilde{h}]^3}{G(\tilde{x})^3} - 2\sum_{i=1}^n \frac{h_i^3}{x_i^3} \\
&= -\frac{\nabla^3 G(\tilde{x})[\tilde{h},\tilde{h},\tilde{h}]}{G(\tilde{x})} + 3a^2c + 2c^3 - 2\sum_{i=1}^n \frac{h_i^3}{x_i^3} \\
&\leq -\frac{\nabla^3 G(\tilde{x})[\tilde{h},\tilde{h},\tilde{h}]}{\nabla^2 G(\tilde{x})[\tilde{h},\tilde{h}]}\frac{\nabla^2 G(\tilde{x})[\tilde{h},\tilde{h}]}{G(\tilde{x})} + 3a^2c + 2c^3 + 2b^3 \text{ using again (3.2)} \\
&= -\frac{\nabla^3 F(x)[h,h,h]}{\nabla^2 F(x)[h,h]}\frac{\nabla^2 G(\tilde{x})[\tilde{h},\tilde{h}]}{G(\tilde{x})} + 3a^2c + 2c^3 + 2b^3 \\
&\leq 3\gamma a^2 b + 3a^2 c + 2c^3 + 2b^3 \text{ using condition (3.1)} \, .
\end{aligned}$$

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. □

Because it does not involve a fractional power of hessian (namely $\nabla^2 F(x)[h, h]$) to the power $\frac{3}{2}$), condition (3.1) is usually much easier to check than the original inequality (2.1), as will be demonstrated in the next section. 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 3.2. Then the second barrier*

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

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 3.2, 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* (3.1) *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 $x \to 0$).

We point out that since condition (3.1) is invariant with respect to positive scaling of F, the results from Lemma 3.2 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.

To conclude this section, we present an extension of Lemma 3.2.

Definition 3.1 *A convex function F defined on $\operatorname{int}\mathcal{C} \subseteq \mathbb{R}^n_{++}$ is called $[\gamma; I]$-compatible if and only if there exists a constant γ and a set $I \subseteq \{1, 2, \ldots, n\}$ such that*

$$\nabla^3 F(x)[h,h,h] \leq 3\gamma \nabla^2 F(x)[h,h] \sqrt{\sum_{i \in I} \frac{h_i^2}{x_i^2}} \text{ for all } x \in \operatorname{int}\mathcal{C} \text{ and } h \in \mathbb{R}^n \ . \tag{3.3}$$

This definition is a simple extension of the condition that was underlying Lemma 3.2, where we only require a subset of the n variables to enter the sum under the square root. A straightforward adaptation of Lemma 3.2 and Corollary 3.1 leads to the following:

Lemma 3.3 *Let F be a $[\gamma; I]$-compatible convex function defined on $\operatorname{int}\mathcal{C} \subseteq \mathbb{R}^n_{++}$. We have that*

$$F_1 : \operatorname{int}\mathcal{C} \mapsto \mathbb{R} : x \mapsto F(x) - \sum_{i \in I} \log x_i$$

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

$$F_2 : \operatorname{int}\operatorname{epi} F \mapsto \mathbb{R} : (x,u) \mapsto -\log\big(u - F(x)\big) - \sum_{i \in I} \log x_i$$

satisfies the first and second conditions of self-concordancy (2.1)-(2.2) with parameter $(\kappa_2, \nu_2) = (r_2(\gamma), |I| + 1)$ on its domain.

An interesting property is that the sum of two $[\gamma; I]$-compatible functions is still $[\gamma; I]$-compatible. More precisely, we have the following theorem.

Theorem 3.3 *Let F_1 be a $[\gamma_1; I_1]$-compatible convex function defined on $\operatorname{int}\mathcal{C}_1 \subseteq \mathbb{R}^n_{++}$ and F_2 be a $[\gamma_2; I_2]$-compatible convex function defined on $\operatorname{int}\mathcal{C}_2 \subseteq \mathbb{R}^n_{++}$. Then $F_1 + F_2$ is a $[\max\{\gamma_1, \gamma_2\}; I_1 \cup I_2]$-compatible convex function on $\operatorname{int}\mathcal{C}_1 \cap \operatorname{int}\mathcal{C}_2$.*

Proof. Using the definition of $[\gamma; I]$-compatibility, we write for all $x \in \operatorname{int}\mathcal{C}$ and $h \in \mathbb{R}^n$

$$\begin{aligned}
\nabla^3(F_1+F_2)(x)[h,h,h] &\leq 3\gamma_1 \nabla^2 F_1(x)[h,h] \sqrt{\textstyle\sum_{i \in I_1} \frac{h_i^2}{x_i^2}} + 3\gamma_2 \nabla^2 F_2(x)[h,h] \sqrt{\textstyle\sum_{i \in I_2} \frac{h_i^2}{x_i^2}} \\
&\leq 3\max\{\gamma_1,\gamma_2\} \Big(\nabla^2 F_1(x)[h,h] \sqrt{\textstyle\sum_{i \in I_1} \frac{h_i^2}{x_i^2}} + \nabla^2 F_2(x)[h,h] \sqrt{\textstyle\sum_{i \in I_2} \frac{h_i^2}{x_i^2}} \Big) \\
&\leq 3\max\{\gamma_1,\gamma_2\} \nabla^2(F_1+F_2)(x)[h,h] \sqrt{\textstyle\sum_{i \in I_1 \cup I_2} \frac{h_i^2}{x_i^2}} \ ,
\end{aligned}$$

which is the desired property. □

4 Application to structured convex problems

In this section we rely on the work in [3], where several classes of structured convex optimization problems are shown to admit a self-concordant logarithmic barrier. However, Lemma 3.2 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 $g_i : \mathbb{R}_+ \mapsto \mathbb{R} : z \mapsto g_i(z)$ are required to satisfy

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

(which by the way implies their convexity). This class of problems is studied in [8, 13]. Classical entropy optimization results as a special case when $g_i(z) = z\ln z$ (in that case, it is straightforward to see that condition (4.1) holds with $\kappa_i = 1$). Another simplex example of problems in this class involve functions $g_i(z) = z^{l_i}$, which is easily shown to satisfy condition (4.1) with $\kappa_i = l_i - 2$.

Let us use Lemma 3.2 with $F_i : x_i \mapsto g_i(x_i)$ and $\gamma = \frac{\kappa_i}{3}$. Indeed, checking condition (3.1) 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

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

is $(r_2(\frac{\kappa_i}{3}), 2)$-self-concordant[6]. However, in order to use this barrier to solve problem (EEO), we need to reformulate it as

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

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 $r_2(\frac{1}{3}) = 1$.

[6] This corrects the statement in [3] where it is mentioned that $g_i(x_i) - \log x_i$, i.e. the first barrier in Lemma 3.2, is self-concordant. As it is made clear in Note 3.3, this cannot be true in general.

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 [5] is

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

It is shown in [3] that condition (3.1) holds for

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

with $\gamma = 1$, so that the corresponding second barrier in Lemma 3.1 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

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

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

4.3 l_p-norm optimization

Let $\{I_k\}_{k=1\ldots r}$ be a partition of $\{1,2,\ldots,m\}$, $\eta \in \mathbb{R}^n$, $a_i \in \mathbb{R}^n$ for $1 \leq i \leq m$, $b_k \in \mathbb{R}^n$ for $1 \leq k \leq r$, $c \in \mathbb{R}^m$, $d \in \mathbb{R}^r$ and $p \in \mathbb{R}^m$ such that $p_i \geq 1$. The primal l_p-norm optimization problem [19] is

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

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

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

This problem can be reformulated as

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

where each of the m constraints involving an absolute value is indeed equivalent to a pair of linear constraints $a_i^T x - c_i \leq s_i$ and $c_i - a_i^T x \leq s_i$. Once again, a self-concordant barrier 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 (3.1) with $\gamma = \frac{2p_i - 1}{3p_i} < 1$, which implies in the same fashion as above that

$$-\log\big(t_i^{1/p_i} - s_i\big) - \log t_i$$

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

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

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 [3], where Lemma 3.2 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 also mention that very similar results hold for the dual l_p-norm optimization problem[7].

5 Concluding remarks

In this paper, we recalled a few basic facts about 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 2.1). 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 3.2 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: Lemma 3.1 deals with the second self-concordancy condition (2.2), while our improved Lemma 3.2 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

[7] However, we would like to point out that the nonlinear function involved in these developments is wrongly stated to satisfy condition (3.1) with $\gamma = \sqrt{2}(q_i+1)/(3q_i)$, while the correct value is $\sqrt{5q_i^2 - 2q_i + 2}/(3q_i)$.

the only quantity that really matters in the final complexity result, one could imagine to consider the following inequality

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

which is satisfied with $\Gamma = \kappa\sqrt{\nu}$ 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 [12], an attempt at generalizing self-concordant barriers.
- ◇ Following the same principles as for (2.4d) and (2.5d), condition (5.1) can be reformulated as

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

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

References

[1] E. D. Andersen, J. Gondzio, C. Mészáros, and X. Xu, *Implementation of interior-point methods for large scale linear programs*, Interior Point Methods of Mathematical Programming (T. Terlaky, ed.), Applied Optimization, vol. 5, Kluwer Academic Publishers, 1996, pp. 189–252.

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

[3] 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.

[4] 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.

[5] R. J. Duffin, E. L. Peterson, and C. Zener, *Geometric programming*, John Wiley & Sons, New York, 1967.

[6] 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.

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

[8] C. Han, P. Pardalos, and Y. Ye, *Implementation of interior-point algorithms for some entropy optimization problems*, Optimization Methods and Software **1** (1992), 71–80.

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

[10] ———, *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.

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

[12] 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.

[13] F. Potra and Y. Ye, *A quadratically convergent polynomial interior-point algorithm for solving entropy optimization problems*, SIAM Journal on Optimization **3** (1993), 843–860.

[14] J. Renegar, *A mathematical view of interior-point methods in convex optimization*, MPS/SIAM Series on Optimization, SIAM, New York, 2001.

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

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

[17] C. Roos, T. Terlaky, and J.-Ph. Vial, *Theory and algorithms for linear optimization. an interior point approach*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Chichester, UK, 1997.

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

[19] T. Terlaky, *On l_p programming*, European Journal of Operations Research **22** (1985), 70–100.