Self-concordant barriers for structured convex problems

François Glineur

Francois.Glineur@fpms.ac.be

Faculté Polytechnique de Mons

Interior-Point Methods Workshop

Budapest, Hungary

July 14, 2000

Outline

Introduction

- ◇ Convex optimization
- ◇ Barrier function

Self-concordancy

- ◇ A matter of definitions
- ◇ Two approaches

A useful lemma

- ◇ Statement
- ◇ Proof

Structured convex problems

- ◇ Extended entropy optimization
- ◇ Geometric optimization
- ◇ l_p-norm optimization

Concluding remarks

Introduction

Convex optimization

Let $f_0 : \mathbb{R}^n \mapsto \mathbb{R}$ be a convex function, $C \subseteq \mathbb{R}^n$ be a convex set : optimize a vector $x \in \mathbb{R}^n$

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

Properties

- ◇ All local optima are global
- ◇ Lagrange duality ⇒ nice dual problem

Different ways to describe C:

a. Nonlinear functions

Let $f_i : \mathbb{R}^n \mapsto \mathbb{R}$ be m convex functions, $i = 1, 2, \ldots, m$, we define C by

$$C = \{x \in \mathbb{R}^n \mid f_i(x) \leq 0 \text{ for all } i = 1, 2, \ldots, m\}$$

(intersection of convex level sets), which gives

$$\inf_{x \in \mathbb{R}^n} f_0(x) \quad \text{s.t.} \quad f_i(x) \leq 0 \text{ for all } i = 1, 2, \ldots, m \tag{NLO}$$

b. Conic setting

Let two vectors $b, c \in \mathbb{R}^n$, a matrix $A \in \mathbb{R}^{m\times n}$ and a convex cone $K \subseteq \mathbb{R}^n$ and define

- ◇ $f_0(x) = c^T x$, a linear objective
- ◇ $L = \{x \in \mathbb{R}^n \mid Ax = b\}$, an affine subspace
- ◇ $C = K \cap L$, the intersection of a convex cone and an affine subspace

which gives

$$\inf_{x\in\mathbb{R}^n} c^T x \quad \text{s.t.} \quad x \in K \text{ and } Ax = b \qquad \text{(CO)}$$

Barrier function

$F : \mathbb{R}^n \mapsto \mathbb{R}$ is a barrier function for C if

- ◇ it is smooth
- ◇ it is strongly convex on $\operatorname{int} C$
- ◇ $F(x) \to +\infty$ when $x \to \partial C$

Idea: approximate a constrained problem by a *family* of unconstrained problems based on F

Let $\mu \in \mathbb{R}_{++}$ be a parameter and consider

$$\inf_{x \in \mathbb{R}^n} f_0(x) + \mu F(x) \qquad (\mathrm{P}_\mu)$$

We have

$$x^*_\mu \to x^* \text{ when } \mu \searrow 0$$

where

- x^*_μ is the (unique) solution of (P_μ)
- x^* is a solution of the original problem (P)

Ingredients

- A method for unconstrained optimization
- A barrier function

Interior-point methods choose

- Newton's method
- Logarithmic barrier function

When C is defined with nonlinear functions:

$$F(x) = -\sum_{i=1}^{n} \ln(-f_i(x))$$

(other types of unconstrained approximations are possible, e.g. the method of centers)

Self-concordant barriers

A matter of definitions

Question: What is a good barrier, i.e. a barrier for which Newton's method is efficient ?

Answer: A *self-concordant* barrier

A function $F : \operatorname{int} C \mapsto \mathbb{R}$ is called (κ, ν)-self-concordant on C iff

- ◇ F is convex
- ◇ F is three times differentiable
- ◇ $F(x) \to +\infty$ when $x \to \partial C$
- ◇ the following two conditions hold

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

$$\nabla F(x)^T (\nabla^2 F(x))^{-1} \nabla F(x) \le \nu$$

for all $x \in \operatorname{int} C$ and $h \in \mathbb{R}^n$

[Nesterov & Nemirovsky, 1988]

Alternative definition.

Let $x \in \operatorname{int} C$ and $h \in \mathbb{R}^n$ and define

$$F_{x,h}(t) : \mathbb{R} \mapsto \mathbb{R} : t \mapsto F(x + th)$$

Replace conditions involving differentials by

$$\begin{aligned} F'''_{x,h}(0) &\leq 2\kappa F''_{x,h}(0)^{\frac{3}{2}} \\ F'_{x,h}(0)^2 &\leq \nu F''_{x,h}(0) \end{aligned}$$

for all $x \in \operatorname{int} C$ and $h \in \mathbb{R}^n$

Geometric interpretation

Using local Hessian-based norms

- ◇ Variation of the Hessian bounded
- ◇ Size of the Newton step bounded

Scaling

Let $\lambda \in \mathbb{R}_+$ be a positive scalar

$$F \text{ is } (\kappa, \nu)\text{-SC} \Leftrightarrow \lambda F \text{ is } (\frac{\kappa}{\sqrt{\lambda}}, \lambda\nu)\text{-SC}$$

Summation

Let F_1 be (κ_1, ν_1)-SC and F_2 be (κ_2, ν_2)-SC

$$F_1 + F_2 \text{ is } (\max\{\kappa_1, \kappa_2\}, \nu_1 + \nu_2)\text{-SC}$$

Main result

Self-concordant barrier $\Rightarrow$ polynomial number of iterations to solve (P)

Let F be a (κ, ν)-self-concordant barrier for C and let $x_0 \in \operatorname{int} C$ be a starting point, a *short-step interior-point* algorithm can solve problem (P) up to ϵ accuracy within

$$O\left(\kappa\sqrt{\nu}\log\frac{c^T x_0 - p^*}{\epsilon}\right) \text{ iterations,}$$

such that at each iteration the self-concordant barrier and its first and second derivatives have to be evaluated and a linear system has to be solved in $\mathbb{R}^n$.

Complexity *invariant* w.r.t. to scaling of F

Corollary

F, ∇F and $\nabla^2 F$ polynomially computable $\Rightarrow$ problem (P) can be solved in polynomial time.

Existence

There is a universal SC barrier with parameter $(\kappa = 1, \nu = O(n))$ (not necessarily computable)

Two approaches

Scaling doesn't change complexity $\Rightarrow$ one of the two constants κ and ν can be arbitrarily fixed

If there exists a (κ, ν)-SC barrier F for C then it can be scaled to get a

- $\diamond$ $(\kappa\sqrt{\nu}, 1)$-SC barrier or a
- $\diamond$ $(1, \kappa^2\nu)$-SC barrier

Comparison

When dealing with problem (NLO), it was typical to

- $\diamond$ consider each f_i separately
- $\diamond$ use the first scaling ($\nu = 1$)

Indeed, if

$$F_i : \mathbb{R}^n \mapsto \mathbb{R} : x \mapsto -\ln(-f_i(x))$$

satisfies the first condition with $\kappa = \kappa_i$ then

$$F_i \text{ is } (\kappa_i, 1)\text{-self-concordant}$$

because the second ν condition is automatically satisfied with $\nu = 1$ if f_i is convex.

This implies in the end that

$$F = \sum_{i=1}^{m} F_i \text{ is } (\kappa, m)\text{-SC with } \kappa = \max_{i=1,\ldots,m} \kappa_i$$

and that problem (NLO) can be solved in

$$O(\sqrt{m} \max_{i=1,\ldots,m} \kappa_i) = O(\sqrt{m}\, \|\kappa\|_\infty) \text{ iterations}$$

However, second scaling ($\kappa = 1$) is **superior** !

Indeed, we have then that F_i is $(1, \kappa_i^2)$-SC which implies that

$$F = \sum_{i=1}^{m} F_i \text{ is } (1, \nu)\text{-SC with } \nu = \sum_{i=1}^{m} \kappa_i^2$$

and that problem (NLO) can be solved in

$$O(\sqrt{\sum_{i=1}^{m} \kappa_i^2}) = O(\|\kappa\|_2) \text{ iterations}$$

which is always better since

$$\|\kappa\|_2 \le \sqrt{m}\, \|\kappa\|_\infty$$

(strict inequality when κ_i's not all equal)

A useful lemma

Proving self-concordancy not always an easy task → helpful lemma :

Statement

[Den Hertog et al.] Let $F : \mathbb{R}^n \mapsto \mathbb{R}$ be a convex three times differentiable function on C. If there is a $\gamma \in \mathbb{R}_+$ such that

$$\nabla^3 F(x)[h,h,h] \leq 3\gamma \nabla^2 F(x)[h,h] \sqrt{\sum_{i=1}^n \frac{h_i^2}{x_i^2}}$$

then the following barrier functions

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

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

satisfy the first condition on κ with

$\kappa_1 = \gamma + 1$ for F_1 on C

$\kappa_2 = \gamma + 1$ for F_2 on $\operatorname{epi} F = \{(x,u) \mid F(x) \leq u\}$

Improvement

It is possible to lower κ_1 and κ_2 : let

$$r_1 : \mathbb{R} \mapsto \mathbb{R} : \gamma \mapsto \begin{cases} 1 & \text{when } \gamma \leq 1 \\ \frac{\gamma}{\sqrt{3-2/\gamma}} & \text{when } \gamma \geq 1 \end{cases}$$

and

$$r_2 : \mathbb{R} \mapsto \mathbb{R} : \gamma \mapsto \begin{cases} 1 & \text{when } \gamma \leq 1 \\ \frac{\gamma+1+1/\gamma}{\sqrt{3+4/\gamma+2/\gamma^2}} & \text{when } \gamma \geq 1 \end{cases}$$

We have in fact $\kappa_1 = r_1(\gamma)$ and $\kappa_2 = r_2(\gamma)$

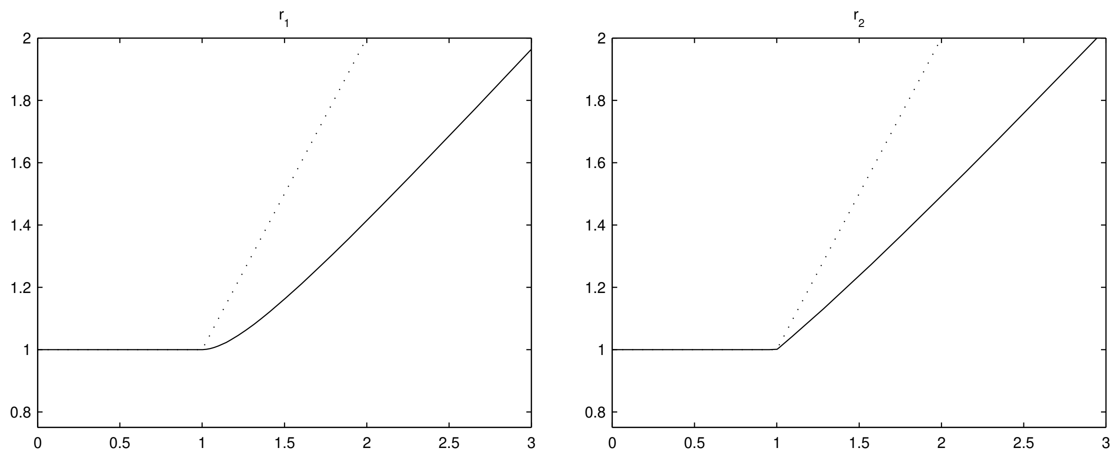

Figure 1: Graphs of functions r_1 and r_2

Note that $r_{1,2}(\gamma) \geq 1$ with equality when $\gamma \leq 1$ and $r_{1,2}(\gamma) \sim \frac{\gamma}{\sqrt{3}}$ when $\gamma \to +\infty$

Proof

For the case of barrier F_1 and parameter κ_1: let

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

After some manipulations we obtain

$$\nabla^2 F_1[h,h] = a^2 + b^2$$

and

$$\nabla^3 F_1(x)[h,h,h] \le 3\gamma a^2 b + 2b^3$$

We wish to show that

$$\nabla^3 F_1(x)[h,h,h] \le 2\kappa_1 (\nabla^2 F_1(x)[h,h])^{\frac{3}{2}}$$

which means we have

$$\kappa_1 = \sup_{a,b} \frac{3\gamma a^2 b + 2b^3}{2(a^2+b^2)^{\frac{3}{2}}}$$

which can be shown (after some elementary calculus) to be equal to $r_1(\gamma)$

A similar proof holds for F_2 and κ_2

Structured convex problems

Extended entropy optimization

$$\min c^T + \sum_{i=1}^{n} g_i(x_i) \quad \text{s.t.} \quad Ax = b \text{ and } x \geq 0$$

where scalar functions $g_i : \mathbb{R} \mapsto \mathbb{R}$ satisfy

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

(which implies convexity)

Special case: classical entropy optimization when $g_i(x) = x \ln x \Rightarrow \kappa_i = 1$

Use Lemma with $F(x_i) = g_i(x_i)$ to prove that

$$-\ln\Big(t_i - g_i(x_i)\Big) - \ln(x_i) \;\text{ is }\; \Big(r_2(\frac{\kappa_i}{3}), 2\Big)\text{-SC}$$

Total complexity of extended entropy opt. is

$$O\Big(\sqrt{2\sum_{i=1}^{n} r_2(\frac{\kappa_i}{3})^2}\Big) \text{ iterations}$$

or

$$O(\sqrt{2n}) \text{ iterations for entropy optimization}$$

Geometric optimization

Let $\{I_k\}_{k=1\ldots r}$ be a partition of $\{1,2,\ldots,n\}$. The dual geometric optimization problem is

$$\min\ c^T x + \sum_{k=1}^{r}\left[\sum_{i\in I_k} x_i \ln(\frac{x_i}{\sum_{i\in I_k} x_i})\right]$$
$$\text{s.t. } Ax = b \text{ and } x \geq 0 \qquad \text{(DGO)}$$

Applying the Lemma with $\gamma = 1$ and

$$F(x_{I_k}) = \sum_{i\in I_k} x_i \ln(\frac{x_i}{\sum_{i\in I_k} x_i})$$

we find that the following barrier

$$-\ln\Big(t_k - \sum_{i\in I_k} x_i \ln(\frac{x_i}{\sum_{i\in I_k} x_i})\Big) - \sum_{i\in I_k} \ln x_i$$

is $(1, |I_k| + 1)$-self-concordant

$\rightarrow$ global complexity for (DGO) is

$$O(\sqrt{n+r}) \text{ iterations}$$

l_p-NORM OPTIMIZATION

Let $\{I_k\}_{k=1\ldots r}$ be a partition of $\{1,2,\ldots,m\}$ and $p_i \geq 1$. The primal l_p-norm optimization problem is

$$\max\ \eta^T x \quad \text{s.t.} \quad f_i(x) \leq 0 \text{ for all } k = 1,\ldots,r$$

with

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

which can be reformulated as

$$\max \eta^T x \quad \text{s.t.} \quad \left\{ \begin{array}{rcl} \left| a_i^T x - c_i \right| & \leq & s_i \ (\equiv 2\times \text{ lin.}) \\ s_i & \leq & t_i^{\frac{1}{p_i}} \\ \sum_{i \in I_k} \frac{t_i}{p_i} & \leq & d_k - b_k^T x \end{array} \right.$$

Using $F(t_i) = t_i^{\frac{1}{p_i}}$ and $\gamma = \frac{2-\frac{1}{p_i}}{3}$ we find that $-\ln\left(t_i^{\frac{1}{p_i}} - s_i\right) - \ln t_i$ is $(1,2)$-self-concordant

$\rightarrow$ global complexity is

$$O(\sqrt{4m+r}) \text{ iterations}$$

(a similar result holds for the dual)

Concluding remarks

- ◇ Definition of self-concordancy involves two parameters (κ, ν)
- ◇ Complexity of IPMs proportional to $\kappa\sqrt{\nu}$
- ◇ Scaling affects κ and ν but not complexity $\Rightarrow$ one of them can be arbitrarily fixed
- ◇ However, when combining (adding) SC barriers with different parameters, it is best to scale them such that all κ's are equal, e.g. $(\kappa_i, 1)$ gives $\|\kappa\|_2$ instead of $\sqrt{m}\,\|\kappa\|_\infty$

- ◇ Den Hertog et al's lemma has been improved when $\gamma \geq 1$
- ◇ This lemma alone is sufficient to prove self-concordancy of the logarithmic barrier for many classes of structured convex problems