Convex Optimization in Machine Learning and Computational Finance Lecture 11: Interior-Point Methods

Dr. Michel Baes, Pr. Patrick Cheridito

RiskLab / ETH Zürich

Quick recall of last week's lecture

- Smoothness accelerates the gradient method (**GM**).
- **GM** is no longer optimal for smooth problems.
- Nesterov's accelerated method is optimal
  and works in $\mathcal{O}(\sqrt{L/\epsilon})$ for L-smooth functions.
- Newton method (if f'' is available)
  works much faster (quadratically), but only locally.
  No usable description of the quadratic convergence zone.
  The method can be undefined (non-invertible Hessian).
- The Mirror Descent method generalizes **GM**
  by addressing an internal contradiction of the method.
  Complexity is the same as for **GM**.
  We can use other (better) norms/scalar products.
- Multiplicative Weights (a mirror-descent method)
  allows to combine weak guesses optimally.

I. Making Newton's Method work

A quick recall on Newton's Method

We approximate epi$f$ by the epigraph of a quadratic function, and we minimize it:

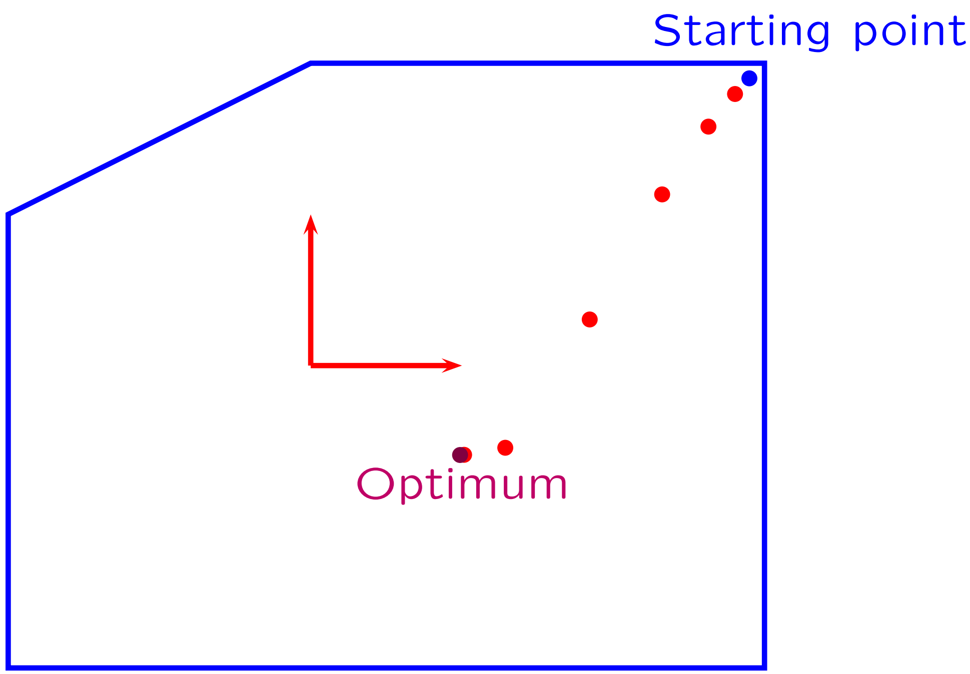
$$f(y) \rightsquigarrow f(x_k)+\langle f'(x_k), y-x_k\rangle+\frac{1}{2}\langle f''(x_k)(y-x_k), y-x_k\rangle \to \min_{y\in\mathbb{R}^n}$$

$$x_{k+1} = x_k - f''(x_k)^{-1} f'(x_k).$$

A seemingly insignificant observation: Newton's method is affine invariant

$f(x, y) = -\ln[(3-x)(2+x)] - \ln[(2+y)(2-y)] - \ln(x/2 - y + 2)$,

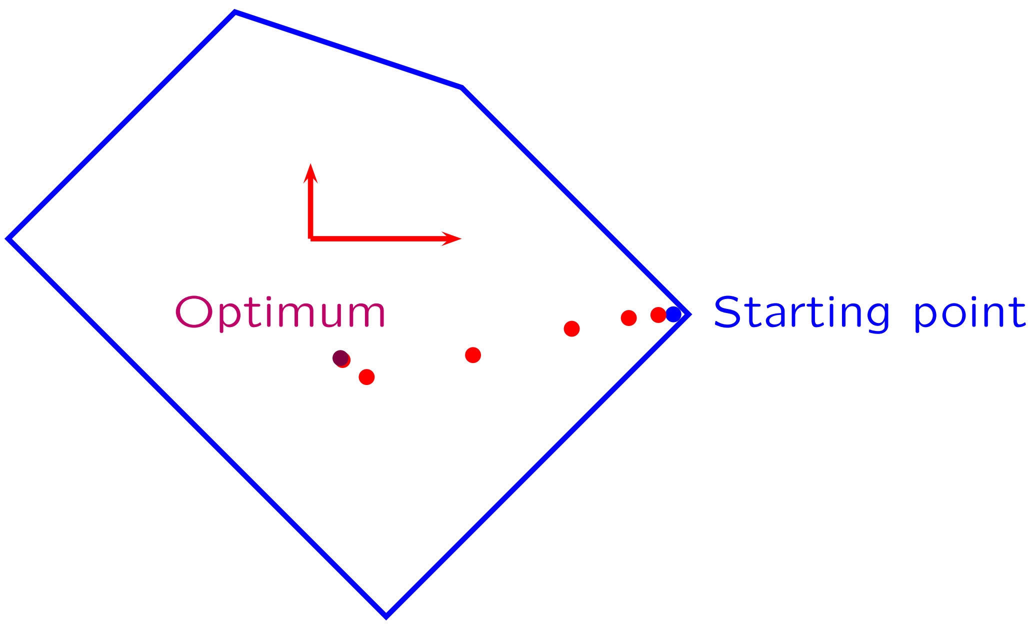
where $f(t) := +\infty$ if $t \leq 0$.



A seemingly insignificant observation: Newton's method is affine invariant

$$f(x,y) = -\ln[(6-2x-y)(4-2x-y)] - \ln[(4+2x+y)(4+2x-y)] - \ln(8-3y-2x),$$

where $f(t) := +\infty$ if $t \leq 0$.



A seemingly insignificant observation: Newton's method is affine invariant

Let f be a twice differentiable function.
Let $x_0 \in \mathbb{R}^n$ and $x_{k+1} := x_k - f''(x_k)^{-1} f'(x_k)$
while $f''(x_k)$ is still invertible (say $k \le K$).

Let $A \in \mathbb{R}^{n \times n}$ invertible and $\phi(y) := f(Ay)$.
Let $y_0 := A^{-1} x_0$ and $y_{k+1} := y_k - \phi''(y_k)^{-1} \phi'(y_k)$.

y_k is well-defined and $y_k = A^{-1} x_k$ for all $k \le K$.

Proof:

$$\langle \phi'(y), h \rangle = \lim_{t \downarrow 0} \frac{f(A(y+th)) - f(Ay)}{t} = \langle f'(Ay), Ah \rangle = \langle A^T f'(Ay), h \rangle,$$

thus $\phi'(y) = A^T f'(Ay)$. Similarly, one can show that $\phi''(y) = A^T f''(Ay) A$, which exists iff $f''(Ay)$ exists. Thus $y_{k+1} = y_k - (A^T f''(Ay_k) A)^{-1} A^T f'(Ay_k) = y_k - A^{-1} f''(Ay_k) f'(Ay_k)$. A recursive argument shows that $y_k = A^{-1} x_k$.

Kantorovich's convergence result is not affine invariant

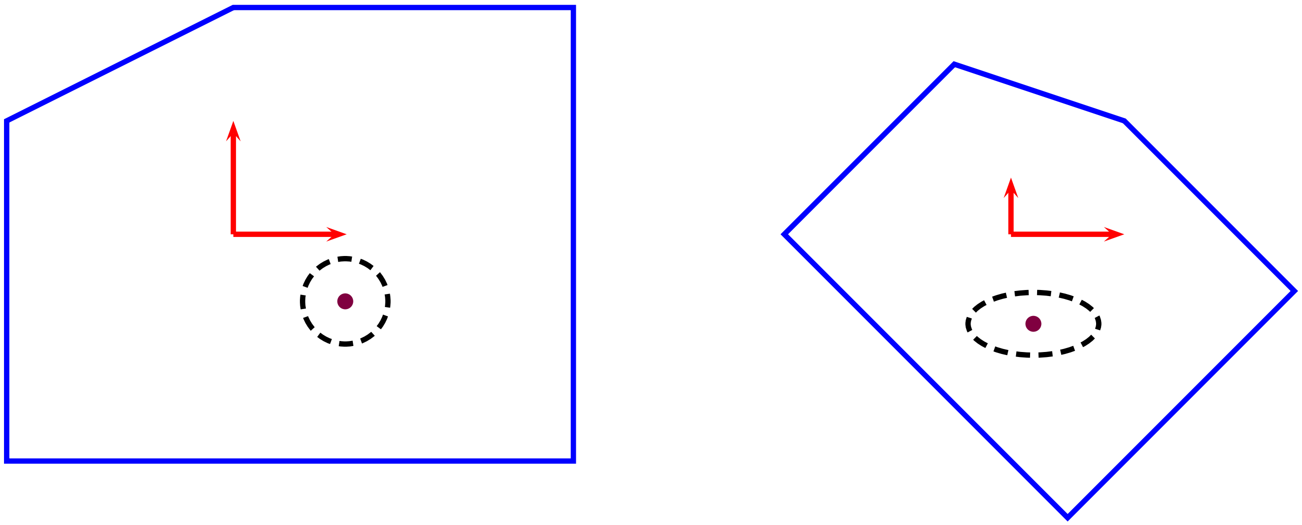
Suppose that $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ satisfies:

- f is twice differentiable.
- $f''(x^*) \succeq lI \succ 0$.
- The Hessian f'' is Lipschitz continuous w.r.t. $||\cdot||_2$:
  $\forall\, x, y \; |||f''(x) - f''(y)|||_2 \le M||x - y||_2$,
  where $|||A|||_2 := \sup\{||Ax||_2 : ||x||_2 \le 1\}$.

Theorem 1 (Kantorovich)
If $||x_0 - x^*||_2 \le 2l/3M$, the iterates x_k of Newton's method are well-defined and:

$$||x_{k+1} - x^*||_2 \le \frac{M||x_k - x^*||_2{}^2}{2(l - M||x_k - x^*||_2)} \le \frac{3M}{2l}||x_k - x^*||_2{}^2.$$

The convergence zone predicted by Kantorovich is not affine invariant



Something must be inappropriate in Kantorovich's assumptions.

How to revisit Kantorovich's result?

Suppose that $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ satisfies:

- f is twice differentiable.
- $f''(x^*) \succeq lI \succ 0$.
- The Hessian f'' is Lipschitz continuous w.r.t. $||\cdot||_2$: $\forall\, x, y \; |||f''(x) - f''(y)|||_2 \leq M||x-y||_2$, where $|||A|||_2 := \sup\{||Ax||_2 : ||x||_2 \leq 1\}$.

The last one is not affine-invariant, because of the **norm**.

The solution of Nesterov and Nemirovski:
use the norm $||h||_x := \langle f''(x)h, h\rangle^{1/2}$ instead.
It is affine invariant: $||h||_x = \langle \phi''(y)A^{-1}h, A^{-1}h\rangle^{1/2}$, where $\phi(y) := f(Ay)$.

Note: Its dual is $||h||_{x,*} = \langle f''(x)^{-1}h, h\rangle^{1/2}$.

How to correct Kantorovich's result?

Suppose that $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ satisfies:

- f is twice differentiable.
- $f''(x^*) \succeq lI \succ 0$.
- The Hessian f'' is Lipschitz continuous w.r.t. $||\cdot||_2$:
  $\forall\, x, y \; |||f'''(x)[y-x]|||_2 \le M||y-x||_2$,
  where $|||A|||_2 := \sup\{||Ax||_2 : ||x||_2 \le 1\}$.

The last one is not affine-invariant, because of the **norm**.

The solution of Nesterov and Nemirovski:

use the norm $||h||_x := \langle f''(x)h, h\rangle^{1/2}$ instead.

It is affine invariant: $||h||_x = \langle \phi''(y)A^{-1}h, A^{-1}h\rangle^{1/2}$,

where $\phi(y) := f(Ay)$.

Note: Its dual is $||h||_{x,*} = \langle f''(x)^{-1}h, h\rangle^{1/2}$.

Self-concordant functions

The property:

$$\exists M > 0 : \ \forall\, x, y \in \mathrm{dom} f, h \in \mathbb{R}^n : \ |f'''(x)[y-x, h, h]| \le M \|x-y\|_2 \|h\|_2^2$$

will be converted into:

$$\exists M > 0 : \ \forall\, x, y \in \mathrm{dom} f, h \in \mathbb{R}^n : \ |f'''(x)[y-x, h, h]| \le M \|x-y\|_x \|h\|_x^2$$

Equivalently [Book of Nesterov, Nemirovski, Appendix 1]:

$$\exists M > 0 : \ \forall\, x \in \mathrm{dom} f, h \in \mathbb{R}^n \ |f'''(x)[h, h, h]| \le M \|h\|_x^3.$$

We call such functions *self-concordant*.

Example: $f(t) = -\ln(t)$ for $t > 0$.

$$f'(t) = -1/t, \quad f''(t) = 1/t^2, \quad f'''(t) = -2/t^3.$$

We have $\|h\|_t^2 = h^2/t^2$, and $|f'''(t)[h, h, h]| = 2(|h|/t)^3$.

Thus $M := 2$ works in the definition.

Building self-concordant functions

Denote $f \in \mathcal{SC}(M)$ if f is self-concordant with constant M.

- For $\gamma > 0$, we have $f \in \mathcal{SC}(M) \Rightarrow \gamma f \in \mathcal{SC}(M/\sqrt{\gamma})$.
  Therefore, we can assume $M := 2$ wlog.

Proof: Let $x \in \mathrm{dom} f, h \in \mathbb{R}^n$. Then $|\gamma f'''(x)[h,h,h]| \le \frac{M}{\sqrt{\gamma}} \langle \gamma f''(x)h, h\rangle^{3/2}$.

- If $f \in \mathcal{SC}(M_f)$, $g \in \mathcal{SC}(M_g)$, then $f + g \in \mathcal{SC}(\max\{M_f, M_g\})$.

Proof: Note that $\lambda^{3/2} + (1-\lambda)^{3/2} \le 1$ for $\lambda \in [0,1]$.
Then
$$|(f'''(x) + g'''(x))[h,h,h]| \le M_f \langle f''(x)h, h\rangle^{3/2} + M_g \langle g''(x)h, h\rangle^{3/2}$$
$$\le \max\{M_f, M_g\} \langle (f''(x) + g''(x))h, h\rangle^{3/2}.$$

- If $f \in \mathcal{SC}(M)$ and $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ is affine,
  then $f \circ \mathcal{A} \in \mathcal{SC}(M)$.

Proof: Indeed, if $\phi(x) := f(Ax)$, then $\phi''(x)[h,h] = f''(Ax)[Ah, Ah]$ and $\phi'''(x)[h,h,h] = f'''(Ax)[Ah, Ah, Ah]$.

Illustration: As $-\ln(t) \in \mathcal{SC}(2)$,
we have $\phi(x) := -\sum_i \ln(a_i^T x - b_i) \in \mathcal{SC}(2)$.

Some elementary building blocks

The following functions are self-concordant.

- Quadratic functions are in $\mathcal{SC}(0)$.
- $t \in \mathbb{R}_{++} \mapsto -\ln(t) \in \mathcal{SC}(2)$.
- $X \in \mathbb{S}^n_{++} \mapsto -\ln(\det(X)) \in \mathcal{SC}(2)$.
- $(t, x) \in \mathcal{L}^n_{++} \mapsto -\ln(t^2 - ||x||_2^2) \in \mathcal{SC}(2)$.
- $t \in [-\pi/2; \pi/2] \mapsto -\ln(\cos(t)) \in \mathcal{SC}(2)$.
- More will come...

The proofs are often a bit tedious. See next slide.

The barrier for Semidefinite Programming

Let $F(X) := -\ln(\det(X))$ for $X \in \mathbb{S}^n_{++}$. Let $X \in \mathbb{S}^n_{++}$ and $H \in \mathbb{S}^n$:

$$\langle F'(X), H\rangle_F = \lim_{t\downarrow 0} \frac{-\ln[\det(X+tH)/\det(X)]}{t} = \lim_{t\downarrow 0} \frac{-\ln\det(I+tX^{-1/2}HX^{-1/2})}{t}.$$

Set $A := X^{-1/2}HX^{-1/2}$. We continue as follows:

$$\langle F'(X), H\rangle_F = \lim_{t\downarrow 0} \frac{-\ln\det(I+tA)}{t} = \lim_{t\downarrow 0} -\sum_{i=1}^{n} \frac{\ln(1+t\lambda_i(A))}{t} = -\sum_{i=1}^{n} \frac{1}{\lambda_i(A)}.$$

Thus $\langle F'(X), H\rangle_F = -\langle I, A\rangle_F = -\langle X^{-1}, H\rangle_F$ and $F'(X) = -X^{-1}$.
Observe that $(X+tH)(X^{-1} - tX^{-1}HX^{-1}) = I + \mathcal{O}(t^2)$, thus

$$F''(X)H = \lim_{t\downarrow 0} \frac{(X+tH)^{-1} - X^{-1}}{t} = \lim_{t\downarrow 0} \frac{X^{-1} - tX^{-1}HX^{-1} - X^{-1}}{t} = X^{-1}HX^{-1},$$

Finally, $F'''(X)[H,H] = -2X^{-1}HX^{-1}HX^{-1}$ by the chain rule.

Then $|F'''(X)[H,H,H]| \leq 2\langle F''(X)H, H\rangle_F^{3/2}$ results

from $\sqrt[3]{\sum_i \lambda_i(A)^3} \leq \sqrt{\sum_i \lambda_i(A)^2}$, and $F \in \mathcal{SC}(2)$. ∎

II. What is nice about self-concordant functions?

An ellipsoid that's always in $\mathrm{dom} f$

Let $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{+\infty\} \in \mathcal{SC}(2)$ and $x \in \mathrm{int}\,\mathrm{dom} f$.

1. The *Dikin ellipsoid* $B_x(x,1) := \{y \in \mathbb{R}^n : ||y-x||_x < 1\}$ is contained in $\mathrm{dom} f$.
2. If $y \in B_x(x,1)$, then $\frac{||y-x||_x}{1-||y-x||_x} \geq ||y-x||_y \geq \frac{||y-x||_x}{1+||y-x||_x}$.

Proof:

Let us fix $x \in \mathrm{int}\,\mathrm{dom} f$ and $h \in \mathbb{R}^n$, $h \neq 0$ so that $y := x+h \in \mathrm{int}\,\mathrm{dom} f$.

1. Consider the function:

$$\phi(t) := \langle f''(x+th)h, h\rangle^{-1/2} \quad \Rightarrow \quad \phi'(t) = -\frac{f'''(x+th)[h,h,h]}{2\langle f''(x+th)h,h\rangle^{3/2}},$$

thus $-1 \leq \phi'(t) \leq 1$ for every t such that $x+th \in \mathrm{dom} f$.

Moreover $\phi(t) \to 0$ when $x+th \to \mathrm{bd}(\mathrm{dom} f)$.

As $-1 \leq \phi'(t) \leq 1$, we have $\phi(t) > 0$ for $-\phi(0) < t < \phi(0)$.

But $\phi(0) = 1/||h||_x = 1/||y-x||_x$.

2. The first inequality is $\phi(1) \geq \phi(0) - 1$, since $\phi(0) > 1$.

The second one is $\phi(1) \leq \phi(0) + 1$. ■

The Hessian does not vary too much

In Kantorovich's setting, f'' is Lipschitz continuous
  i.e. f'' does not vary much for points that are close.
We have the same for self-concordant functions,
  but the bounds are **far** better.

Let $f \in \mathcal{SC}(2)$, $x \in \text{int}\,\text{dom} f$, and $y \in B_x(x, 1)$.
Let $r := ||y - x||_x$. Then

$$(1 - r)^2 f''(x) \preceq f''(y) \preceq f''(x)/(1 - r)^2.$$

The proof is not completely obvious...

Let $f \in \mathcal{SC}(2)$, $x \in \operatorname{int}\operatorname{dom} f$, and $y \in B_x(x,1)$.
Let $r := ||y-x||_x$. Then

$$(1-r)^2 f''(x) \preceq f''(y) \preceq f''(x)/(1-r)^2.$$

Proof:

Let $x \in \operatorname{int}\operatorname{dom} f$, $h \in \mathbb{R}^n$, $h \neq 0$. Let

$$\psi(t) := \langle f''(x+t(y-x))h,h\rangle = \langle f''(x_t)h,h\rangle, \quad t \in [0,1]$$

$$\begin{aligned} |\psi'(t)| &= |f'''(x_t)[y-x,h,h]| \le 2||y-x||_{x_t}||h||_{x_t}^2 = 2||y-x||_{x_t}\psi(t) \\ &= \frac{2\psi(t)}{t}||x_t-x||_{x_t} \le \frac{2\psi(t)}{t}\frac{||x_t-x||_x}{1-||x_t-x||_x} = \frac{2\psi(t)||y-x||_x}{1-t||y-x||_x} = \frac{2\psi(t)r}{1-tr}. \end{aligned}$$

because of the inequality $||z-x||_z \le \frac{||z-x||_x}{1-||z-x||_x}$ shown earlier. Thus:

$$|\psi'(t)|/\psi(t) = [\ln(\psi(t))]' \le 2r/(1-tr) = -2[\log(1-tr)]'.$$

Integrating on $[0,1]$, we get the second inequality.

The first one is proved similarly. ∎

The Hessian does not vary much, so what?

$$(1-r)^2 f''(x) \preceq f''(y) \preceq f''(x)/(1-r)^2.$$

$\Rightarrow\; f(y) \geq f(x) + \langle f'(x), y-x\rangle + r - \ln(1+r)$.

$\Rightarrow\; f(y) \leq f(x) + \langle f'(x), y-x\rangle - r - \ln(1-r)$.

Let $\lambda(x) := ||f'(x)||_{x,*}$ be the *Newton decrement*.

$$x_{k+1} = x_k - \frac{1}{1+\lambda(x_k)} f''(x_k)^{-1} f'(x_k).$$

If f is bounded from below, the algorithm (*damped Newton method*) always converges.

Proof: Note that $||x_{k+1} - x_k||_{x_k} = \lambda(x_k)/(1+\lambda(x_k)) < 1$. As:

$$f(x_{k+1}) \leq f(x_k) + \langle f'(x_k), x_{k+1} - x_k\rangle - \frac{\lambda(x_k)}{1+\lambda(x_k)} - \ln\left(1 - \frac{\lambda(x_k)}{1+\lambda(x_k)}\right),$$

we have $f(x_{k+1}) \leq f(x_k) - \lambda(x_k) + \ln(1+\lambda(x_k)) < f(x_k)$. ■

You don't need x^* to check that x is in the quadratic convergence zone

$$x_{k+1} = x_k - f''(x_k)^{-1} f'(x_k).$$

Assume $\lambda(x_k) < 1$. Then

$$\lambda(x_{k+1}) \leq \frac{\lambda(x_k)^2}{(1 - \lambda(x_k))^2}.$$

Note 1: In particular, if $\lambda(x_k) < \frac{3-\sqrt{5}}{2} \simeq 0.3819$,
then $\lambda(x_{k+1}) < \lambda(x_k)$.

Note 2: Globally convergent algorithm:
If $\lambda(x_k) > 0.3$, use the damped Newton method:
we can guarantee $f(x_{k+1}) < f(x_k) - 0.3 + \ln(1 + 0.3)$
$\simeq f(x_k) - 0.037$
Otherwise, use the standard Newton method.

The proof is fairly non-trivial

Proof: Let $r_k := ||x_{k+1} - x_k||_{x_k}$. Then $\lambda(x_k) = r_k$. Also:

$$\lambda(x_{k+1})^2 = \langle f''(x_{k+1})^{-1} f'(x_{k+1}), f'(x_{k+1})\rangle \leq \left(\frac{1}{1-r_k}\right)^2 \langle f''(x_k)^{-1} f'(x_{k+1}), f'(x_{k+1})\rangle.$$

By definition of x_{k+1}

$$f'(x_{k+1}) = f'(x_{k+1}) - f'(x_k) - f''(x_k)(x_{k+1} - x_k) = G(x_{k+1} - x_k),$$

where

$$f''(x_k)\left(\frac{r_k^2}{3} - r_k\right) \preceq G := \int_0^1 [f''(x_k + t(x_{k+1} - x_k)) - f''(x_k)]dt \preceq f''(x_k)\frac{r_k}{1-r_k},$$

as a consequence of the inequalities between Hessians.
Since $r_k^2/3 - r_k \leq r_k/(1-r_k)$, we get:

$$\lambda(x_{k+1})^2 = ||f'(x_{k+1})||_{x_k*}^2 \leq \left(\frac{r_k}{1-r_k}\right)^2 \langle f''(x_k)^{-1} f'(x_k), f'(x_k)\rangle = \frac{r_k^2 \lambda(x_k)^2}{(1-r_k)^2}.$$

Summary - Self-concordant functions: the right thing for Newton's method

These functions have many properties, among which:

1- If $||f'(x)||_x^* := \sqrt{\langle f'(x), f''(x)^{-1} f'(x) \rangle} \leq \frac{3-\sqrt{5}}{2}$,
then x is in the quadratic convergence zone
(automatic test, no x^* needed)

2- for every $x \in \mathrm{dom} f$, $B_x(x,1) \subseteq \mathrm{dom} f$
[Generalizes Karmarkar's Method, the first efficient method for LP]

3- Damped Newton method always converges
when the step-size is $h := 1/(1 + ||f'(x)||_{x,*})$

III. Solving convex optimization problems with self-concordant functions

III. Solving convex optimization problems with self-concordant barriers

An inconsistency in Black-Box methods for convex programming

How do Black-Box methods deal with convexity?

- First, you realize that your problem is convex
  (or even strongly convex).
  Thus, you investigate its **global** properties.

- Then you hide your problem in a mysterious **black box**.
  You only interact with it through an oracle
  that gives you **local** information.
  (if x is the current point, gives $f(x)$,
  and/or $f'(x)$, and/or $f''(x)$ . . .)

They act as if they didn't know the problem is convex.

By the way, how do you check convexity ?

- Directly from the definition.

Try this one if you are patient enough: for $x \in \mathbb{R}^n_{++}$,

$$f(x) := \max \left\{ \exp(||x||_2^2), \lambda_{\max} \left(\sum_{i=1}^{n} x_i A_i \right) - \ln(x_1) \right\} + 5x_n^4,$$

with $A_1, \ldots, A_n \in \mathbb{S}^N_+$.

- By using the structure of the function.

You know several "simple" convex functions: $t^2, \exp(t), \ldots$
and several operations that preserve convexity: max, $+, \ldots$

And after all this work, you give this beautiful structure to a Black-Box method that does not care about it!

By the way, how do you check convexity ?

- Directly from the definition

Try this one if you are patient enough: for $x \in \mathbb{R}^n_{++}$,

$$f(x) := \max\left\{\exp(||x||_2^2), \lambda_{\max}\left(\sum_{i=1}^{n} x_i A_i\right) - \ln(x_1)\right\} + 5x_n^4.$$

with $A_1, \ldots, A_n \in \mathbb{S}^N_+$.

- By using the structure of the function.

You know several "simple" convex functions: $t^2, \exp(t), \ldots$
and several operations that preserve convexity: max, $+$, ...

But Interior-Point Methods use this structure explicitly to construct a *barrier* for the feasible set.

Dealing with equality constraints in Newton's method

Not an issue!

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a twice diff. convex function.

- Equality constraints for convex problems must be linear.
- Wlog, no redundant constraint: A has full row rank.

$$\min\{f(x) : Ax = b\}$$

Without constraints, Newton method linearizes $f'(x^*) = 0$. Indeed, around x_k, this linearization is:

$$f'(x_k) + f''(x_k)(\bar{x} - x_k) = 0 \quad \Leftrightarrow \quad \bar{x} = x_k - f''(x_k)^{-1} f'(x_k).$$

Linear equality constraints are easy

$$\min\{f(x) : Ax = b\}$$

With constraints, KKT conditions read as:

$$f'(x^*) + A^T u^* = 0, \quad Ax^* = b.$$

Linearizing around (x_k, u_k), we get:

$$f'(x_k) + f''(x_k)(\bar{x} - x_k) + A^T \bar{u} = 0, \quad A\bar{x} = b.$$

$$\begin{pmatrix} f''(x_k) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{u} \end{pmatrix} = \begin{pmatrix} f''(x_k)x_k - f'(x_k) \\ b \end{pmatrix}.$$

Dealing with equality constraints in Newton's method

$$\begin{pmatrix} f''(x_k) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{u} \end{pmatrix} = \begin{pmatrix} f''(x_k)x_k - f'(x_k) \\ b \end{pmatrix}.$$

► If $f''(x_k) \in \mathbb{S}^n_{++}$, the system above has a unique solution.

Proof: Let $\begin{pmatrix} f''(x_k) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Taking the scalar product of the first equation with x gives

$$\langle f''(x_k)x, x \rangle + \langle A^T u, x \rangle = \langle f''(x_k)x, x \rangle = 0,$$

and $x = 0$. Thus $A^T u = 0$ and $u = 0$ because A has full row rank. The matrix is invertible and the Newton's system is solvable. ■

Set constraints and Newton's method

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a twice diff. convex function.
Let $Q \subseteq \mathbb{R}^n$ be a closed convex set with nonempty interior.

$$\min\{f(x) : x \in Q\} = \min_x f(x) + \chi_Q(x),$$

where $\chi_Q(x) = 0$ if $x \in Q$ and $\chi_Q(x) = +\infty$ otherwise.

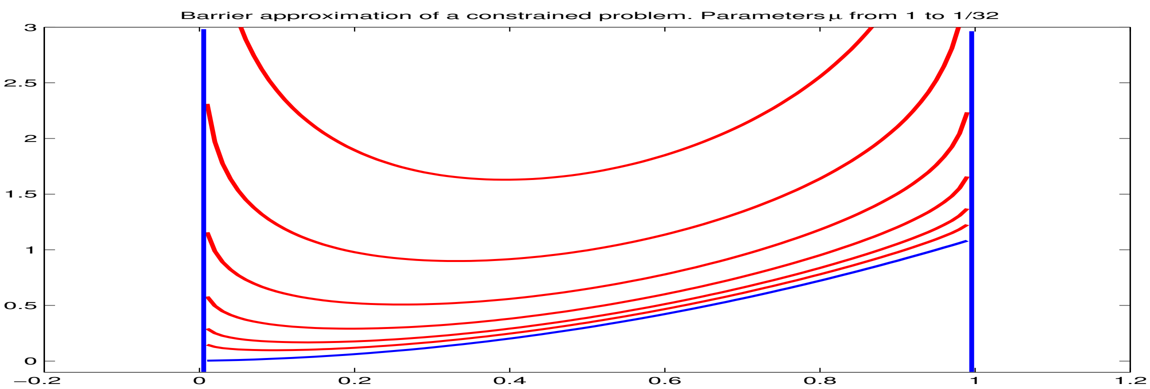
- $f(x)+\chi_Q(x)$ is no longer twice differentiable on $\operatorname{dom} f \cap Q$.
- A *barrier for Q* is a function $F_Q : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$:
  convex (preferably strictly);
  twice differentiable, with $\operatorname{int} Q$ as domain;
  $\Rightarrow \lim_{x \to \mathrm{bd}(Q)} F_Q(x) = +\infty$.

The barrier method

for approximating the feasible set

$$\min\{f(x) : x \in Q\} = \min_x f(x) + \chi_Q(x),$$

Let F_Q be a barrier for Q.
We approximate $f(x) + \chi_Q(x)$ by $f(x) + \mu F_Q(x)$ for $\mu \downarrow 0$.



The barrier method

for approximating the feasible set

Let $x^* \in \arg\min\{f(x) : x \in Q\} = \arg\min_x f(x) + \chi_Q(x)$,
and $x^*(\mu) \in \arg\min_x f(x) + \mu F_Q(x)$.

If F_Q is bounded from below and f continuous,
$f(x^*(\mu)) \to f(x^*)$ for $\mu \downarrow 0$.

Proof: Let M be the lower bound of F_Q. For all $\bar{x} \in \operatorname{int} Q$:

$$
\begin{aligned}
f(\bar{x}) &= \lim_{\mu \downarrow 0} f(\bar{x}) + \mu(F_Q(\bar{x}) - M) \\
&\geq \lim_{\mu \downarrow 0} f(x^*(\mu)) + \mu(F_Q(x^*(\mu)) - M) \\
&\geq \lim_{\mu \downarrow 0} f(x^*(\mu)) \geq f(x^*).
\end{aligned}
$$

Note: The curve $\mu \mapsto x^*(\mu)$ is called the *central path*.

General strategy of barrier methods

Determine a point on/close to the central path,
say $\widehat{x}(\mu_0) \approx x^*(\mu_0)$.

for $k \geq 1$:

Fix $\mu_{k+1} < \mu_k$.

Starting from $\widehat{x}(\mu_k)$,

find an approximation $\widehat{x}(\mu_{k+1})$ of $x^*(\mu_{k+1})$.

Increase k.

end

1. How to choose F_Q?

2. How fast can we decrease μ?

3. How close $\widehat{x}(\mu_k)$ must be to $x^*(\mu_k)$?

4. How to compute efficiently $\widehat{x}(\mu_{k+1})$ from $\widehat{x}(\mu_k)$?

5. When do we need to stop?

Self-concordant functions are the ideal barriers F_Q

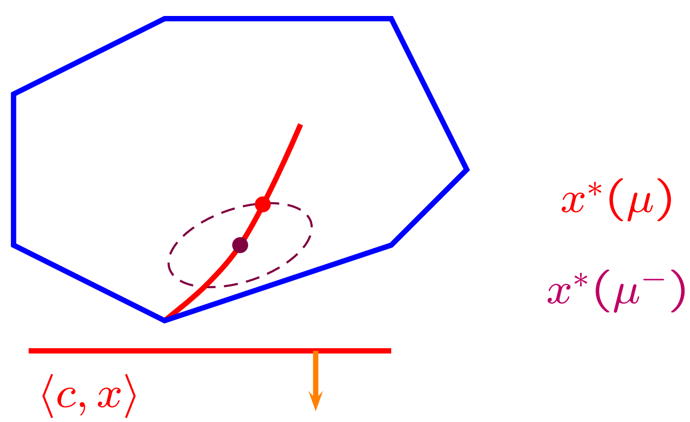
$$\min_x f(x) + \chi_Q(x) \;\to\; \min_x f(x) + \mu F_Q(x)$$

- We need to compute efficiently points close to the central path (with Newton's method).
- Obviously, we choose F_Q self-concordant: $F_Q \in \mathcal{SC}(2)$. If f is **linear** or **quadratic**, $f(x) + \mu F_Q(x) \in \mathcal{SC}(2/\sqrt{\mu})$.
- To make the objective linear, replace $\min\{f(x) : x \in Q\}$ by $\min\{t : x \in Q, (t, x) \in \text{epi} f\}$.
- **Important issue:** given $Q \subseteq \mathbb{R}^n$, how do we construct $F_Q \in \mathcal{SC}(2)$ such that $\text{dom} F_Q = \text{int}\, Q$? In a few slides!

Moving on the central path and staying efficient

$$\min_x \langle c, x\rangle + \chi_Q(x) \to \min_x \langle c, x\rangle + \mu F_Q(x)$$

Goal: Reduce μ as fast as possible (linearly), while staying in the quadratic convergence zone.



Moving on the central path and staying efficient

$$\min_x \langle c, x\rangle + \chi_Q(x) \to \min_x \langle c, x\rangle + \mu F_Q(x)$$

Goal: Reduce μ as fast as possible (linearly), while staying in the quadratic convergence zone.

To get a $\mathcal{SC}(2)$ objective, we scale it:

$$\arg\min_x \langle c, x\rangle + \mu F_Q(x) = \arg\min_x \langle c, x\rangle/\mu + F_Q(x)$$

Let $f(x;\mu) := \langle c, x\rangle/\mu + F_Q(x)$ and

$\lambda(x;\mu)$ be the Newton decrement of $f(x;\mu)$:

$\lambda(x;\mu)^2 := \langle f''(x;\mu)^{-1} f'(x;\mu), f'(x;\mu)\rangle$,

so that x is in the Newton quadratic convergence zone of $f(x;\mu)$ if $\lambda(x;\mu) < \bar{\lambda} := \frac{3-\sqrt{5}}{2}$.

Moving on the central path: Self-concordant barriers

We assume that we are exactly in $x^*(\mu)$.

For which $\mu_- := \frac{\mu}{1+\delta}$ is the point $x^*(\mu)$
in the Newton quadratic convergence zone of $f(x, \mu_-)$?

$$\begin{aligned}\lambda(x^*(\mu);\mu_-) &= \langle f''(x^*(\mu);\mu_-)^{-1} f'(x^*(\mu);\mu_-), f'(x^*(\mu);\mu_-)\rangle^{1/2} \\ &= \delta\langle F_Q''(x^*(\mu))^{-1} F_Q'(x^*(\mu)), F_Q'(x^*(\mu))\rangle^{1/2} < \bar{\lambda},\end{aligned}$$

because $c/\mu + F_Q'(x^*(\mu)) = 0$ by optimality of $x^*(\mu)$,
and $f'(x^*(\mu);\mu_-) = c(1+\delta)/\mu + F_Q'(x^*(\mu)) = \delta F_Q'(x^*(\mu))$.
Thus, we must assume that:

$$\exists \nu > 0 : \quad \langle F_Q''(x)^{-1} F_Q'(x), F_Q'(x)\rangle \leq \nu \quad \forall x \in \operatorname{int} Q.$$

We call such $F_Q \in \mathcal{SC}(2)$ a *ν-self-concordant* **barrier** *for Q*.

Self-concordant barriers: a few examples

$$\forall x \in \mathrm{dom} F,\ \forall h \in \mathbb{R}^n\ |F'''(x)[h,h,h]| \leq 2\langle F''(x)h, h\rangle^{3/2},$$

$$\forall x \in \mathrm{dom} F\ \langle F''(x)^{-1}F'(x), F'(x)\rangle \leq \nu.$$

F	Domain	ν
$-\log(t)$	$\mathbb{R}_{++}$	1
$-\ln(t^2 - \|\|x\|\|_2^2)$	$\mathbb{L}^n_{++}$	2
$-\ln\det(X)$	$\mathbb{S}^n_{++}$	n
$-\ln\cos(t)$	$]-\pi/2, \pi/2[$	1
$-\ln(\ln(t) - x) - \ln(t)$	epi exp	2

Note: None of them is bounded from below,
but we will only use restrictions of them.

Computing some barrier parameters

- For $F(t) = -\log(t)$,
  we have $\langle F''(t)^{-1}F'(t), F'(t)\rangle = t^2 \cdot \left(-\frac{1}{t}\right)^2 = 1$.
- For $F(X) = -\log\det(X)$,
  $\langle F''(t)^{-1}F'(t), F'(t)\rangle = \mathrm{Tr}(X(-X^{-1})X(-X^{-1})) = n$.
- For $F(t,x) := -\ln(t^2 - ||x||_2^2)$:
  let $(\tau, h) \in \mathbb{R} \times \mathbb{R}^n$ and $\phi(\lambda) := (t + \lambda\tau)^2 - ||x + \lambda h||_2^2$.
  Then $F'(t,x)[(\tau;h)] = -\frac{\phi'(0)}{\phi(0)}$
  and $F''(t,x)[(\tau;h),(\tau;h)] = -\frac{\phi''(0)}{\phi(0)} + \left(\frac{\phi'(0)}{\phi(0)}\right)^2$.
  We need to check $2\phi(0)\phi''(0) \le \phi'(0)^2$,
  i.e. $(\tau^2 - ||h||^2)(t^2 - ||x||^2) \le (t\tau - \langle x, h\rangle)^2$.
  It is enough to assume $\tau > 0$ and $\langle x, h\rangle = ||x|| \cdot ||h||$,
  for which the inequality is clearly true. ∎

Building self-concordant barriers

Let $f \in \mathcal{SCB}(\nu; Q)$ if f is a ν-self-concordant barrier for Q.

$$\forall x \in \mathrm{dom} f \ \langle f''(x)^{-1} f'(x), f'(x) \rangle \leq \nu$$

$\Leftrightarrow \quad \forall x \in \mathrm{dom} f, \quad \sup\{2\langle f'(x), h\rangle - \langle f''(x)h, h\rangle : h \in \mathbb{R}^n\} \leq \nu.$

(Use optimality conditions to compute the maximizer).

- If $f \in \mathcal{SCB}(\nu_f; Q_f), g \in \mathcal{SCB}(\nu_g; Q_g)$, then $f + g \in \mathcal{SCB}(\nu_f + \nu_g; Q_f \cap Q_g)$.
- If $f \in \mathcal{SCB}(\nu; Q)$ and $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ is affine, then $f \circ \mathcal{A} \in \mathcal{SCB}(\nu; Q_{\mathcal{A}})$, where $Q_{\mathcal{A}} := \{y \in \mathbb{R}^m : \mathcal{A}(y) \in Q\}$.

Illustration: $\phi(x) := -\sum_i \ln(a_i^T x - b_i) \in \mathcal{SCB}(m; P)$, where P is the *polytope* $\{x : a_i^T x \geq b_i, 1 \leq i \leq m\}$.

Building self-concordant barriers

Write $f \in \mathcal{SCB}(\nu; Q)$ if f is a ν-self-concordant barrier for Q.

- If $f \in \mathcal{SCB}(\nu_f; Q_f), g \in \mathcal{SCB}(\nu_g; Q_g)$,
  then $\phi(x, y) := f(x) + g(y) \in \mathcal{SCB}(\nu_f + \nu_g; Q_f \times Q_g)$.
- Suppose that K is a closed convex cone
  with nonempty interior and no straight line.
  Let $F \in \mathcal{SCB}(\nu, K)$ be *logarithmically homogenous*:
  $F(tx) = F(x) - \nu \ln(t)$ for all $t > 0$ and $x \in \operatorname{int} K$.
  Then $F_* \in \mathcal{SCB}(\nu, -K^*)$.
- **Universality of barriers** Every closed convex set in $\mathbb{R}^n$
  has a Cn-self-concordant barrier [Nesterov, Nemirovski].
  (But computing it at a point involves computing an n-D. integral)

General strategy of barrier methods

Determine a point on/close to the central path,
say $\hat{x}(\mu_0) \approx x^*(\mu_0)$.

for $k \geq 1$:

Fix $\mu_{k+1} < \mu_k$.
Starting from $\hat{x}(\mu_k)$,
find an approximation $\hat{x}(\mu_{k+1})$ of $x^*(\mu_{k+1})$.
Increase k.

end

1. How to choose F_Q? Use a ν-self-concordant barrier.

2. How fast can we decrease μ?

3. How close $\hat{x}(\mu_k)$ must be to $x^*(\mu_k)$?

4. How to compute efficiently $\hat{x}(\mu_{k+1})$ from $\hat{x}(\mu_k)$?

5. When do we need to stop?

A half-space containing the whole domain

Let $F \in \mathcal{SCB}(\nu; Q)$. For every $x, y \in \text{int}\, Q$, we have $\langle F'(x), y - x\rangle < \nu$.

Proof: Observe that for all $x \in \text{int}\, Q$, and $h \in \mathbb{R}^n$:

$$F \in \mathcal{SCB}(\nu; Q) \Rightarrow \sup_{\lambda} 2\langle F'(x), \lambda h\rangle - \langle F''(x)\lambda h, \lambda h\rangle \le \nu,$$

or $\langle F'(x), h\rangle^2 \le \nu \langle F''(x)h, h\rangle$ by optimizing on λ.
We fix $x, y \in \text{int}\, Q$. Let $\phi(t) := \langle F'(x + t(y - x)), y - x\rangle$ for all $t \in [0, 1]$. If $\phi(0) \le 0$, there is nothing to prove. Suppose $\phi(0) > 0$. Since $F \in \mathcal{SCB}(\nu; Q)$, we have $\phi(t)^2 \le \nu\phi'(t)$, therefore $\phi(t) > 0$ for $t \in [0, 1]$, and:

$$\frac{1}{\nu} \le \frac{\phi'(t)}{\phi(t)^2} = \left(-\frac{1}{\phi(t)}\right)' \quad \Rightarrow \quad \frac{t}{\nu} \le \frac{1}{\phi(0)} - \frac{1}{\phi(t)} < \frac{1}{\phi(0)}.$$

Thus, $\phi(0) = \langle F'(x), y - x\rangle < \nu/t$ for $t \in [0, 1]$.

An exact algorithm converges linearly to the solution

$$\min_x \langle c, x\rangle + \chi_Q(x) \rightarrow \min_x \langle c, x\rangle + \mu F_Q(x)$$

Note: The exact solution of the original problem is $x^*(0)$.

Assume we can compute $x^*(\mu)$ exactly.
Recall that $c + \mu F'_Q(x^*(\mu)) = 0$. We have:

$$\langle c, x^*(\mu)\rangle - \langle c, x^*(0)\rangle = -\mu \langle F'_Q(x^*(\mu)), x^*(\mu) - x^*(0)\rangle \leq \mu\nu.$$

The convergence is **linear**.

ONE Newton step is enough to move further on the central path

$$\mu_- := \mu/(1+\delta), \quad x_- := x - F_Q''(x)^{-1}[c + \mu_- F_Q'(x)].$$

(Below, we can take $\beta := 0.1$, and $\delta := 1.4/(1+10\sqrt{\nu})$.)

Theorem 2 *Let* $\beta < \frac{3-\sqrt{5}}{2}$. *When* $\delta \leq \frac{\sqrt{\beta}-\beta-\beta\sqrt{\beta}}{(\beta+\sqrt{\nu})(\sqrt{\beta}+1)}$,

$$\lambda := \left\| \frac{c}{\mu} + F_Q'(x) \right\|_{x*} \leq \beta \quad \Rightarrow \quad \lambda_- := \left\| \frac{c}{\mu_-} + F_Q'(x_-) \right\|_{x_-*} \leq \beta.$$

Proof: Let $\hat{\lambda} := ||c/\mu_- + F_Q'(x)||_{x*}$. Then $\hat{\lambda} \leq \lambda + \delta ||c||_{x*}/\mu_-$. Note:

$$||c||_{x*} = ||f'(x;\mu_-) - \mu_- F_Q'(x)||_{x*} \leq \mu_- \lambda + \mu_- ||F_Q'(x)||_{x*} \leq \mu_-(\beta + \sqrt{\nu}).$$

Since $\langle c, x\rangle/\mu + F_Q(x) \in \mathcal{SC}(2)$, we have $\lambda_- \leq \left(\frac{\hat{\lambda}}{1-\hat{\lambda}}\right)^2 \leq \left(\frac{\lambda+\delta(\beta+\sqrt{\nu})}{1-\lambda-\delta(\beta+\sqrt{\nu})}\right)^2$.
As $\lambda \leq \beta$, this is smaller than β for our particular choice of δ.

Approximate points on the central path are good enough for linear convergence

Theorem 3 *Assume that* $x := \hat{x}(\mu)$ *is* centered, *that is,* $||c + \mu F'_Q(x)||_{x*} \leq \mu\beta$. *Then*

$$\langle c, x\rangle - \langle c, x^*(0)\rangle \leq \mu\left(\nu + \frac{(\beta + \sqrt{\nu})\beta}{(1-\beta)}\right).$$

Proof: Let $f(x;\mu) := \langle c, x\rangle + \mu F_Q(x)$.
We know that $\langle c, x^*(\mu)\rangle - \langle c, x^*(0)\rangle \leq \mu\nu$. The missing part is:

$$\langle c, x\rangle - \langle c, x^*(\mu)\rangle = \langle f'(x;\mu) - \mu F'_Q(x), x\rangle \leq (||f'(x;\mu)||_{x*} + \mu ||F'_Q(x)||_{x*})||x - x^*(\mu)||_x.$$

First, $||f'(x;\mu)||_{x*} = ||c + \mu F'_Q(x)||_{x*} \leq \mu\beta$. Also, $||F'_Q(x)||_{x*} \leq \sqrt{\nu}$. Finally, $||x - x^*(\mu)||_x \leq \frac{\beta}{1-\beta}$, as it can be deduced from the Hessian inequality for self-concordant functions (Exercise).

Here is the algorithm we have developed and its convergence speed

Find a ν-self-concordant barrier F_Q for Q.
Set $\beta := 0.1$, $\delta := 1.4/(1+10\sqrt{\nu})$.
Determine a point $\widehat{x}_0$ such that $||c + \mu_0 F'_Q(\widehat{x}_0)||_{\widehat{x}_0*} \leq \beta$.
for $k \geq 1$:
  Set $\mu_{k+1} := \mu_k/(1+\delta)$.
  Perform one Newton step from $\widehat{x}(\mu_k)$ for getting $\widehat{x}(\mu_{k+1})$.
  Increase k.
end

We have:

$$\langle c, x_k\rangle - \langle c, x^*(0)\rangle \leq \frac{\mu_0}{(1+\delta)^k}\left(\nu + \frac{(\beta+\sqrt{\nu})\beta}{(1-\beta)}\right) = \frac{\mu_0}{(1+\delta)^k}\left(\nu + \frac{\sqrt{\nu}}{9} + \frac{1}{90}\right) \leq \epsilon$$

if $\quad k = \mathcal{O}(\log(\mu_0\nu/\epsilon)(\log(1+\delta))^{-1}) = \mathcal{O}(\sqrt{\nu}\log(\mu_0\nu/\epsilon))$.

General strategy of barrier methods

Determine a point on/close to the central path,
say $\widehat{x}(\mu_0) \approx x^*(\mu_0)$.

for $k \geq 1$:

Fix $\mu_{k+1} < \mu_k$.

Starting from $\widehat{x}(\mu_k)$,

find an approximation $\widehat{x}(\mu_{k+1})$ of $x^*(\mu_{k+1})$.

Increase k.

end

1. How to choose F_Q? Use a ν-self-concordant barrier.

2. How fast can we decrease μ? Linearly: $\mu \to \mu/(1+\delta)$.

3. How close $\widehat{x}(\mu_k)$ must be to $x^*(\mu_k)$? Answered.

4. How to compute efficiently $\widehat{x}(\mu_{k+1})$ from $\widehat{x}(\mu_k)$? ✓

5. When do we need to stop? Answered.

But much better strategies are possible using duality!

Is this method running inside Sedumi?

Not exactly.

Some common points:

- Uses self-concordant barriers in a barrier method.
- Follows a central path.
- Performs Newton steps (with equality constraints).

Some differences:

- Solves simultaneously the primal and the dual;
  uses heavily primal-dual relations.
- Allows a much less restrictive condition
  for central path approximations (Larger quadratic
  convergence zone for *symmetric* barriers).
- Manages to decrease μ much faster by using a *predictor*
  (look for $x^*(\mu_{--})$)-*corrector* (to get back to the central path).
- Has a particular strategy for finding a starting point.
- Works only for LP, SOCP, and SDP.


That's all Folks!