Lecture 15

Self concordant Barriers

- Fast Recap of Interior point method
- what properties of the logarithmic Barrier function were crucial.
- self concordant functions
- self concordant Barrier
- properties.

- Recap of IPM

original LP

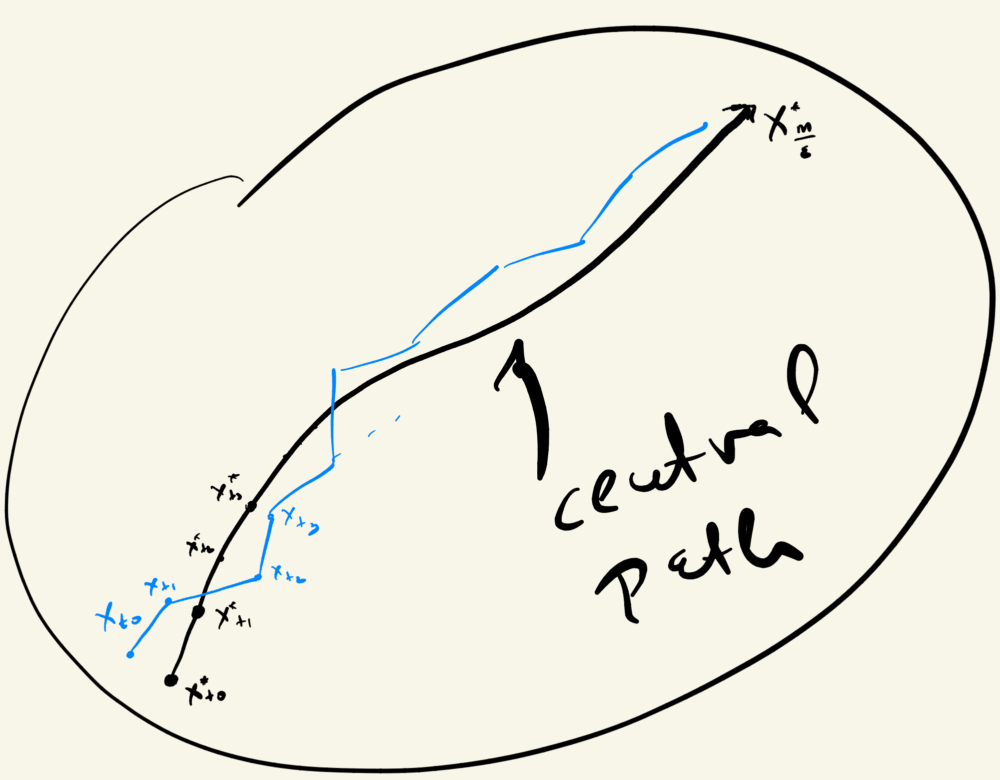
$\min \langle c, x \rangle$

s.t $Ax \geq b$

$, x \in \mathbb{R}^n \quad A = \begin{pmatrix} - a_1^T - \\ - a_2^T - \\ \vdots \\ - a_m^T - \end{pmatrix} \in \mathbb{R}^{m \times n} \quad , b \in \mathbb{R}^m$

$\downarrow$ unconstrained problem

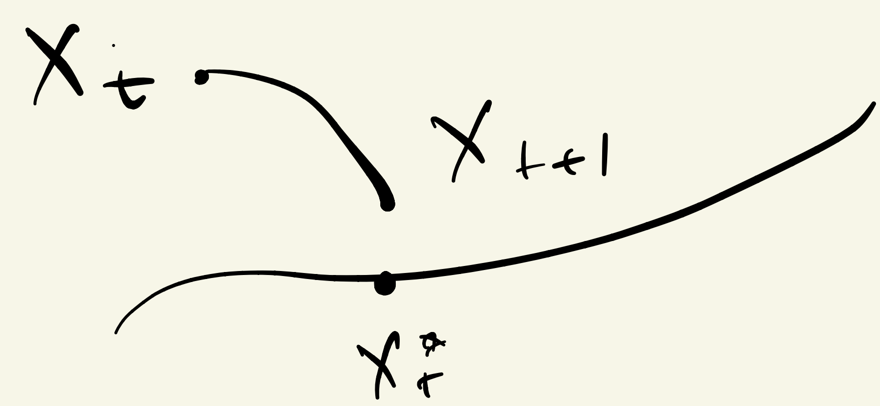
$$\min f_t(x) = t \langle c, x \rangle - \underbrace{\varphi(x)}_{= \sum_{i=1}^m \log(\langle a_i, x \rangle - b_i)}$$



what properties were crucial?

① The Hessian is smooth. That property is needed for the Newton step to make sense

(i.e



comes closer to x_t^*)

$$\forall x,y \text{ s.t } \|x-y\|_{H^{-1}} \le \delta \Rightarrow \bigcirc H_y \preceq H_x \preceq \bigcirc H_y$$

② $\|\nabla\varphi(x)\|_{H^{-1}} \le \vartheta$ the barrier function gradient norm is bounded

$$\overset{②}{\Rightarrow} \tilde{t} \geq t\left(1+\frac{\delta}{\sqrt{\vartheta}}\right)$$

$$\Rightarrow \tilde{O}\left(\frac{\sqrt{\vartheta}}{\delta}\right) \text{ iterations}$$

$$\overset{①}{\Rightarrow} c^T x^t_{w/\varepsilon} - c^T x^* < \varepsilon$$

(Not a lot about LPs was used)

self-concordant functions

Informall definition = a class of functions for which Newton's method work.

In the quadratic convergence close to the OPT proof of the Newton method we required

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq M \|x-y\|_2$$

which is not good for 2 reasons:

① It does not hold for functions that $\to +\infty$ in the boundary (barrier functions)

② It is not affine invariant

Definition

$f: \mathbb{R} \to \mathbb{R}$ is self-concordant if

$$|f'''(x)| \leq 2 f''(x)^{3/2}$$

it permits to $\to +\infty$ in the boundary but not too fast.

Observations

- constant 2 is not important

If $|f'''(x)| \leq K f''(x)^{3/2} \Rightarrow$

$\Rightarrow \tilde{f}(x) = \frac{K^2}{4} f(x)$ is self-concordant

- affine invariant

$$g(y) = f(ay+b) = f(x)$$

$$g'''(y) = a^3 f'''(ay+b) = a^3 f'''(x)$$

$$g''(y) = a^2 f''(x) \qquad |a^3 f'''(x)| \leq 2 \left(a^2 f''(x)\right)^{3/2} = $$

$$= 2 \cdot a^3 f''(x)^{3/2}$$

Definition of self concordant function in higher dimensions

(one dimension $\longrightarrow$ higher dimension)

$f: \mathbb{R}^n \to \mathbb{R}$ is self concordant only if $g(t) = f(x+tv)$ is self-concordant for every x, v (in every line)

$\Rightarrow \forall v, x \quad |\nabla^3 f(x)[v][v][v]| \leq 2\left(v^T \nabla^2 f(x) v\right)^{3/2}$

$= 2 \|v\|^3_{\nabla^2 f(x)}$

$\sum_{i,j,k} \frac{\partial^3 f(x)}{\partial x_i \partial x_j \partial x_k} v_i v_j v_k$

examples

$f(x) = -\log x$ self-concordant

$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$ self concordant

$f(x) = -\log\det X \longleftarrow$ self-concordant

$X \in S^n_{++}$

properties

If x, y are close changing the quadratic norm does not affect the distance

$\|y-x\|_{\nabla^2 f(x)} \ll 1$ then also

$\|y-x\|_{\nabla^2 f(y)}$ is small

proof

$$\varphi(t) = \frac{1}{\sqrt{v^T \nabla^2 f(x+tu) v}} = \frac{1}{\|u\|_{\nabla^2 f(x+tu)}}$$

$$|\varphi'(t)| = \left| -\frac{1}{2} \frac{\nabla^3 f(x+tu)[u][u][u]}{u^T \nabla^2 f(x+tu) u} \right| \leq 1$$

$\forall t$

take $u = y-x$. Then

$$\varphi(0) = \frac{1}{\|y-x\|_{\nabla^2 f(x)}} \quad \text{and} \quad \varphi(1) = \frac{1}{\|y-x\|_{\nabla^2 f(y)}}$$

$$\left.\begin{array}{l} \|y-x\|_{\nabla^2 f(x)} \ll 1 \Rightarrow \varphi(0) > 1 \\ |\varphi'(t)| \leq 1 \end{array}\right\} \varphi(1) > \varphi(0) - 1$$

$$\|y-x\|_{\nabla^2 f(y)} \leq \frac{\|y-x\|_{\nabla^2 f(x)}}{1 - \|y-x\|_{\nabla^2 f(x)}}$$

Theorem

If $\|y-x\|_{\nabla^2 f(x)} \leq 1$ then

$$(1-\|y-x\|_{\nabla^2 f(x)})^2 \nabla^2 f(x) \preceq \nabla^2 f(y) \preceq (1-\|y-x\|_{\nabla^2 f(x)})^{-2} \nabla^2 f(x)$$

a "smoothness" of the Hessian property that permits to prove quadratic convergence in a region around OPT of the Newton's method.

② $\|\nabla f(x)\|_{(\nabla^2 f(x))^{-1}} \leq 6 \quad \forall x$

Definition

Let f be a self-concordant function. We say it is 6-self concordant if

$$\sup_{u \in \mathbb{R}^n} \left[2\langle \nabla f(x), u \rangle - u^T \nabla^2 f(x) u \right] \leq 6$$

$\Longrightarrow$

some intuition behind the definition:

The definition, assuming that the hessian changes slowly and thus the second order approximation of $f(x+u)$ is good, bounds the increase of the function value if a full Newton step is taken.

observations

$$\nabla_u \left(2 \langle \nabla f(x), u \rangle - u^T \nabla^2 f(x) u \right) = 0$$

$$\Rightarrow 2 \nabla f(x) - 2 \nabla^2 f(x) \cdot u = 0$$

$$\Rightarrow u = (\nabla^2 f(x))^{-1} \nabla f(x)$$

so $2 \langle \nabla f(x), u \rangle - u^T \nabla^2 f(x) u =$

$$= \| \nabla f(x) \|_{(\nabla^2 f(x))^{-1}}$$

so $\| \nabla f(x) \|_{(\nabla^2 f(x))^{-1}} \leq 6$

change u to λu and take the gradient w.r.t λ

$$\nabla_\lambda \left(2\lambda \langle \nabla f(x), u \rangle - \lambda^2 u^T \nabla^2 f(x)\, u \right] = 0$$

$$\Rightarrow \langle \nabla f(x), u \rangle - \lambda u^T \nabla^2 f(x)\, u = 0 \Rightarrow$$

$$\Rightarrow \lambda = \frac{\langle \nabla f(x), u \rangle}{u^T \nabla^2 f(x)\, u}$$

plugging back λ we get

$$\nabla^2 f(x) \succcurlyeq \frac{1}{6} \nabla f(x) \nabla f(x)$$

$\min f_0(x)$ ← convex function

s.t $x \in Q$ ← convex set

$\Rightarrow$ $\min a$

s.t $x \in Q' =$

$= Q \cap \{f_0(x) \leq a\}$

when turning the optimization problem into an unconstrained one we introduce an appropriate barrier function

s.t $\varphi(x) \to +\infty$ when $f_0(x) \to a$

we will focus on linear optimization

$f_0(x) = \langle c, x \rangle$

Theorem

Let $f_t(x) = t \langle c, x \rangle + \varphi(x)$ ← 6-self concordant barrier

then $\langle c, x_t^* \rangle - \langle c, x^* \rangle \leq \frac{6}{t}$

proof

x_t^* is optimal for $f_t(x)$ so

$$\nabla f_t(x_t^*) = 0 \Rightarrow t \cdot c + \nabla \varphi(x) = 0 \Rightarrow$$

$$\Rightarrow \langle c, x_t^* - x^* \rangle = \frac{1}{t} \langle \nabla \varphi(x_t^*), x^* - x_t^* \rangle$$

It is enough to prove that

$$\langle \nabla \varphi(x), y - x \rangle \le \vartheta \quad \forall x, y$$

$$g(s) = \langle \nabla \varphi(x + s(y-x)), y - x \rangle$$

$g(0) > 0$ o.w holds.

$$g'(s) = (y-x)^T \nabla^2 \varphi(x + s(y-x))(y-x) \ge$$

$$\ge \frac{1}{\vartheta} \langle \nabla \varphi(x + s(x-y)), y - x \rangle^2 = \frac{1}{\vartheta} g^2(s)$$

$g(0) > 0$

$g'(s) > \frac{1}{6} g^2(s)$

If s close to 0 $g(s)$ is large

then it decreases super fast

$\frac{g'(s)}{g^2(s)} > \frac{1}{6} \Rightarrow \int_{s=0}^{1} \frac{g'(s)}{g^2(s)} ds > \frac{1}{6} \Rightarrow$

$\Rightarrow -\frac{1}{g(1)} + \frac{1}{g(0)} > \frac{1}{6} \Rightarrow$

$\Rightarrow g(0) < 6$ (If $g(1) < 0$ then we would be fine since $g(0) < 0$)