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
\[
\begin{align*}
\min & \quad c^T x \\
\text{subject to} & \quad A x \geq b, \\
& \quad x \in \mathbb{R}^n
\end{align*}
\]

unconstrained problem
\[
\min f_c(x) = tc^T x - Q(x)
\]

central path

- $c^T \log(x, x_T - b)$
What properties were crucial?

1. The Hessian is smooth. That property is needed for the Newton step to make sense (i.e. $x_0 \xrightarrow{\text{comp}} x^*$ closer to $x^*$).

$$\forall x, y, s.t. \| x - y \|_{H^{-1}} \leq \delta \Rightarrow \exists H, \exists C, \exists \varphi, \exists \psi$$

$\forall \| \varphi(x) \|_{H^{-1}} \leq \beta$ norm is bounded

$$\Rightarrow \tilde{t} \geq t + (1 - \frac{\delta}{\frac{1}{C}})$$

$$\Rightarrow 0 (\frac{\sqrt{\beta}}{\delta}) \text{ iterations}$$

$\Rightarrow C^T x^* \omega_{H} - C^T x^* < \varepsilon$

(Not a lot about $dP$, was used)
**self-concordant functions**

Informal definition: a class of functions for which Newton's method works.

In the quadratic convergence close to the optimal 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 go to infinity in the boundary
2. It is not affine invariant
**Definition**

\[ f: \mathbb{R} \to \mathbb{R} \text{ is self-concordant if} \]

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

Observations

1. **Constant \( K \) is not important**

\[ \text{if } |f''(x)| \leq K f''(x)^{3/2} \Rightarrow \]

\[ \tilde{f}(x) = \frac{K^2}{4} f(x) \text{ is self-concordant} \]

2. **Affine invariant**

\[ g(y) = f(ay + b) = f(x) \]

\[ g'''(y) = a^3 f'''(ay+b) = a^3 f'''(x) \]

\[ g''(y) = a^2 f''(x) \]

\[ |a^2 f''(x)| \leq 2 (a^2 f''(x)^{3/2})^{3/2} = 2 \cdot a^3 f''(x)^{3/2} \]
Definition of self-concordant function
in higher dimensions
(one dimension $\rightarrow$ higher dimension)

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

\[ \Rightarrow \forall v, x \in \mathbb{R}^n \text{ } \mathbb{E}_{v \in \mathbb{S}^{n-1}} \left[ \mathbb{E}_{x \in \mathbb{S}^{n-1}} \right] 2 \left( \nabla^T \nabla f(x) v \right)^{3/2} \leq \left( \nabla^T \nabla^2 f(x) v \right)^{3/2} \]

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

Examples

\[ f(x) = -\log x \text{ self-concordant} \]

\[ f(x) = -\sum_{i=1}^n \log (1 - a_i^T x) \text{ self-concordant} \]

\[ f(x) = -\log \det x \text{ self-concordant} \]
If \( y \) are close changing its quadratic norm does not affect its distance

\[ \| y - x \| \sqrt{f(x)} \ll 1 \] then also

\[ \| y - x \| \sqrt{f(y)} \] is small

**Proof**

\[ \varphi(c+1) = \frac{1}{\sqrt{u^T \nabla^2 f(x+tu)v}} = \frac{1}{\| u \| \sqrt{f(x+tu)}} \]

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

\( \nabla \)

Take \( u = y - x \). Then

\[ \varphi(1) = \frac{1}{\| y - x \| \sqrt{f(x)}} \quad \text{and} \quad \varphi(1) = \frac{1}{\| y - x \| \sqrt{f(y)}} \]

\[ \| y - x \| \sqrt{f(x)} \ll 1 \Rightarrow \varphi(1) \gg 1 \]

\[ \| y - x \| \sqrt{f(y)} \ll 1 \Rightarrow \varphi(1) \gg \varphi(0) - 1 \]

\[ \| y - x \| \sqrt{f(y)} \ll 1 \Rightarrow \frac{1}{1 - \| y - x \| \sqrt{f(y)}} \]
Theorem

If $\|y-x\| \leq 1$ then

$$(1 - \|y-x\| \Delta f(x)) \Delta^2 f(x) \leq \Delta^2 f(y) \leq (1 - \|y-x\| \Delta f(x)) \Delta^2 f(x)$$

→

a "smoothness" of the Hessian property

that permits to prove quadratic convergence in a region around

out of the Newton's method.

2) $\|\nabla^2 f(x)\| \leq 6 \times \times$

Definition

$\det f$ be a self-concaveout function. We say it is $\theta$-self concaveout if

$$\sup_{\text{sup } \nabla f(x), u} [2 \langle \nabla f(x), u \rangle - u^T \nabla^2 f(x) u] \leq 6$$

$\Rightarrow$
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**

$$\nabla_u (2 <\nabla f(x), u> - u^T \nabla^2 f(x) u) = 0$$

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

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

so

$$2 <\nabla f(x), u> - u^T \nabla^2 f(x) u =$$

$$= \|\nabla f(x)\|_2 (\nabla^2 f(x))^{-1}$$

so

$$\|\nabla f(x)\|_2 (\nabla^2 f(x))^{-1} \leq \delta$$
change \( u \) to \( \tilde{u} \) and take the gradient w.r.t. \( \xi \)

\[
\mathcal{N}_g \left( \varphi, u \right) = -\kappa^2 u^2 \nabla^2 \varphi(x) u = 0
\]

\[
\Rightarrow \left< \nabla \varphi(x), u \right> - \kappa u^2 \nabla^2 \varphi(x) u = 0 \Rightarrow
\]

\[
\Rightarrow \kappa = \frac{\left< \nabla \varphi(x), u \right>}{u^2 \nabla^2 \varphi(x) u}
\]

Plugging back \( \kappa \) we get:

\[
\nabla^2 \varphi(x) > \frac{1}{6} \nabla \varphi(x) \nabla \varphi(x)
\]
min \{ f(x) : \text{convex function} \}
\text{s.t. } x \in \mathbb{Q}

\Rightarrow \min \{ a : \text{convex set} \} = \min \{ a : \text{convex set} \} 
\Rightarrow \mathbb{Q} \cap \{ \text{for } a \leq b \}

\text{When turning the optimization problem into an unconstrained one we introduce an appropriate barrier function}
\text{s.t. } f(x) \to \infty
\text{when } f(x) \to -\infty

\textbf{Theorem}
\text{let } f_t(x) = t < c_1 x^2 + \phi(x)
\text{then } <c_1 x^*_t - c_1 x^* > \leq \frac{b}{t}
$x^*_t$ is optimal for $f_t(x)$ so

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

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

It is enough to prove that

$\langle \nabla \phi(x), y - x \rangle \leq \beta \quad \forall x, y$

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

$g(0) \geq 0 \quad \text{forwards}$

holds.

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

$\geq \frac{1}{\beta} \langle \nabla \phi(x + s(y - x)), y - x \rangle^2 = \frac{1}{\beta} g(s)^2$
\[ g(0) > 0 \]

\[ g'(s) > \frac{1}{b} g^2(s) \]

If \( s \) is close to 0, \( g(s) \) is large, then it decreases super fast

\[ \frac{g'(s)}{g^2(s)} > \frac{1}{b} \Rightarrow \int_{s=0}^{1} \frac{g'(s)}{g^2(s)} \, ds > \frac{1}{b} \Rightarrow \]

\[ -\frac{1}{g(r)} + \frac{1}{g(0)} > \frac{1}{b} \Rightarrow \]

\[ \Rightarrow g(0) < b \]

(If \( g(0) < 0 \), then we would be fine since \( g(0) < 0 \))