Basic interior point method theory:

Self-concordancy &

barrier functions

$(\mathbb{E}, \langle \cdot, \cdot \rangle)$: finite dim inner prod. sp. ($\Rightarrow$ induced norm $\|\cdot\|$ & operator norm)

* $f : \mathbb{E} \to \mathbb{R}$ is (Fréchet) diff at $x \in \operatorname{dom} f$:

$$\exists \underset{\text{Gradient}}{g(x)} \in \mathbb{E} \text{ st. } \quad \lim_{\|\Delta x\| \to 0} \frac{f(x+\Delta x) - f(x) - \langle g(x), \Delta x \rangle}{\|\Delta x\|} = 0.$$

* f is twice diff at $x \in \operatorname{dom} f$: $f \in \mathcal{C}^1(\mathbb{E})$, &

$$\exists \underset{\text{Hessian}}{H(x)} \in \mathcal{L}(\mathbb{E}, \mathbb{E}) \text{ s.t. } \quad \lim_{\|\Delta x\| \to 0} \frac{\| g(x+\Delta x) - g(x) - H(x)\Delta x \|}{\|\Delta x\|} = 0.$$

* Gradient & Hessian are dependent on the underlying inner prod.

* Changing to another inner prod $\langle \cdot, \cdot \rangle_S = \langle \cdot, S \cdot \rangle$ ($S \in \mathcal{L}(\mathbb{E}, \mathbb{E})$ is P.D.)

gives $\begin{cases} \text{new gradient} : & S^{-1} g(x) \\ \text{new Hessian} : & S^{-1} H(x) \end{cases}$.

Newton's method on $\min f(x)$

I. If ① $f \in \mathcal{C}^2(\mathbb{E})$ is strongly convex ($\exists m > 0$ s.t. $h^T H(x) h \geq m \|h\|^2$ $\forall x, h \in \mathbb{E}$),

② $H : \mathbb{E} \to \mathcal{L}(\mathbb{E}, \mathbb{E})$ is Lipschitz continuous (near minimizer z),

i.e. $\| H(x) - H(y) \| \leq L \|x - y\| \quad \forall x, y \in B(z, \delta)$

then ① $\|\nabla f(x_{i+1})\| \leq \frac{L}{2m} \|\nabla f(x_i)\|$

② If $\frac{L}{2m} \|\nabla f(x_i)\| < 1$, we have quad. conv.

II. Newton's method is AFFINELY INVARIANT.

BUT not the convergence analysis — L & m change when we use another coor. system!

Local inner prod.

Assuming $f: \mathbb{E} \to \mathbb{R}$ satisfies

(A1) $\quad \operatorname{dom} f$ being open convex,

(A2) $\quad f \in \mathcal{C}^2(\mathbb{E})$,

(A3) $\quad H(x) \succ 0 \qquad \forall x \in \operatorname{dom} f$ (w.r.t. the underlying inner prod. $\langle \cdot, \cdot \rangle$),

define a family of local inner prod. on $\mathbb{E}$:
$\forall x \in \operatorname{dom} f$,

$$\langle u, v \rangle_x := \langle u, H(x) v \rangle \qquad \forall u, v \in \mathbb{E}$$

* (A2) + (A3) $\Rightarrow$ $\forall \varepsilon > 0$, $\exists \delta > 0$ s.t.

$$1 - \varepsilon \le \frac{\|v\|_y}{\|v\|_x} \le 1 + \varepsilon \qquad \forall y \in B(x, \delta)$$

open ball.

Local inner prod.

* $\langle \cdot, \cdot \rangle_x$ is independent of the reference inner prod.

Given another inner prod $\langle \cdot, \cdot \rangle_S = \langle \cdot, S \cdot \rangle$ on $\mathbb{E}$ (where $S \succ 0$),

- Hess. of f becomes $H_S(x) = S^{-1} H(x)$
- local inner prod. becomes

$$\begin{aligned} \langle u, v \rangle_{S,x} &:= \langle u, H_S(x) v \rangle_S \\ &= \langle u, S H_S(x) v \rangle \\ &= \langle u, H(x) v \rangle = \langle u, v \rangle_x . \end{aligned}$$

* $\langle \cdot, \cdot \rangle_x$ induces a local norm

$$\|v\|_x := \sqrt{\langle v, H(x) v \rangle} .$$

Local inner prod.

* Gradient & Hessian wrt local inner prod. $\langle \cdot, \cdot \rangle_x$

$$g_x(y) \quad , \quad H_x(y)$$

* For any reference inner prod,

$$g_x(y) = H(x)^{-1} g(y)$$

$$H_x(y) = H(x)^{-1/2} H(y) H(x)^{-1/2}$$

* $-g_x(x) = -H(x)^{-1} g(x)$ is the Newton step of f at x.

Self-concordant fcn.

* $f: \mathbb{E} \to \mathbb{R}$ is a (strongly non-degen) s.c. fcn. :
  $\forall x \in \operatorname{dom} f$,

  (1) $B_x(x,1) := \{ y \in \mathbb{E} : \|y-x\|_x < 1 \} \subseteq \operatorname{dom} f$

  Dikin ellipsoid

  (2) $\forall y \in B_x(x,1)$,

$$1 - \|y-x\|_x \le \frac{\|v\|_y}{\|v\|_x} \le \frac{1}{1-\|y-x\|_x} \quad \forall v \neq 0$$

* [Nesterov et Nemirovskii]
  - $f: \mathbb{R} \to \mathbb{R}$ is s.c. if ① f is $\mathcal{C}^3$
    ② $|f'''(x)| \le 2 |f''(x)|^{3/2}$
  - $f: \mathbb{R}^n \to \mathbb{R}$ is s.c. if $\phi(t) := f(x+ty)$ is s.c. $\forall x, y \in \mathbb{R}^n$.

Why

$$1 - \|y-x\|_x \leq \frac{\|v\|_y}{\|v\|_x} \leq \frac{1}{1-\|y-x\|_x} \quad ?$$

Thm 2.2.1 (Renegar)

$\forall x \in \operatorname{dom} f$, $y \in B_x(x,1) \subseteq \operatorname{dom} f$, TFAE:

(1) $$1 - \|y-x\|_x \leq \frac{\|v\|_y}{\|v\|_x} \leq \frac{1}{1-\|y-x\|_x} \quad \forall v \neq 0$$

(2) $$\|H_x(y)\|_x, \ \|H_x(y)^{-1}\|_x \leq \frac{1}{(1-\|y-x\|_x)^2}$$

(3) $$\|I - H_x(y)\|_x, \ \|I - H_x(y)^{-1}\|_x \leq \frac{1}{(1-\|y-x\|_x)^2} - 1 .$$

Before the proof, note that $H_x(x) = I$, and

$$\| I - H_x(y) \|_x \leq \frac{1}{(1 - \|y-x\|_x)^2} - 1 \qquad \forall y \in B_x(x,1) \subseteq \operatorname{dom} f$$

implies

(A) $\forall y \in B_x(x, 1/2)$

$$\| I - H_x(y) \|_x \leq 8 \|y - x\|_x$$

$\leftarrow$ (local) Lipschitz continuity of Hessian

(B)
$$\frac{\| D^3 f(x) (y-x) \|_x}{\|y-x\|_x} \leq \frac{\| H_x(y) - H_x(x) - D^3 f(x)(y-x) \|_x + \| H_x(y) - I \|_x}{\|y-x\|_x}$$

$$\therefore \overline{\lim_{y \to x}} \frac{\| D^3 f(x) (y-x) \|_x}{\|y-x\|_x} \leq \overline{\lim_{y \to x}} \frac{2 - \|y-x\|_x}{(1-\|y-x\|_x)^2} \leq 2$$

$$\therefore \| D^3 f(x) \|_x \leq 2.$$

Proof. Obs $$\frac{\|v\|_y^2}{\|v\|_x^2} = \frac{\langle v, H_x(y) v\rangle_x}{\|v\|_x^2}$$

Letting $0 < \lambda_1 \leq \cdots \leq \lambda_n$ be the eigvals of $H_x(y)$,

$$\max_{v \neq 0} \frac{\|v\|_y^2}{\|v\|_x^2} = \lambda_n \quad \text{and} \quad \min_{v \neq 0} \frac{\|v\|_y^2}{\|v\|_x^2} = \lambda_1$$

$$= \|H_x(y)\|_x \qquad\qquad = \frac{1}{\|H_x(y)^{-1}\|_x}$$

$$\therefore \quad 1 - \|x-y\|_x \leq \frac{\|v\|_y}{\|v\|_x} \leq \frac{1}{1-\|y-x\|_x}$$

$$\Leftrightarrow \quad (1 - \|y-x\|_x)^2 \leq \frac{1}{\|H_x(y)^{-1}\|_x} \quad \text{and} \quad \|H_x(y)\| \leq \frac{1}{(1-\|y-x\|_x)^2}$$

2nd equiv: obs $I - H_x(y)$ has eigvals $1-\lambda_n \leq \cdots \leq 1-\lambda_1$

$$\|I - H_x(y)\|_x \leq \max\{\lambda_n - 1, 1-\lambda_1\}$$
$$\leq \max\{\lambda_n - 1, \tfrac{1}{\lambda_1} - 1\}$$

Proof. Obs $\dfrac{\|v\|_y^2}{\|v\|_x^2} = \dfrac{\langle v, H_x(y) v\rangle_x}{\|v\|_x^2}$

Letting $0 < \lambda_1 \leq \cdots \leq \lambda_n$ be the eigvals of $H_x(y)$,

$\forall v \neq 0$

$1 - \|y-x\|_x \leq \dfrac{\|v\|_y}{\|v\|_x} \leq \dfrac{1}{1-\|y-x\|_x}$

iff

$\|H_x(y)\|_x,\ \|H_x(y)^{-1}\|_x \leq \dfrac{1}{(1-\|y-x\|_x)^2}$

iff

$\|I - H_x(y)\|_x,\ \|I - H_x(y)^{-1}\|_x \leq \dfrac{1}{(1-\|y-x\|_x)^2} - 1$

$$\max_{v \neq 0} \frac{\|v\|_y^2}{\|v\|_x^2} = \lambda_n \quad \text{and} \quad \min_{v\neq 0} \frac{\|v\|_y^2}{\|v\|_x^2} = \lambda_1$$

$$= \|H_x(y)\|_x \qquad\qquad = \frac{1}{\|H_x(y)^{-1}\|_x}$$

$$\therefore \quad 1 - \|x-y\|_x \leq \frac{\|v\|_y}{\|v\|_x} \leq \frac{1}{1-\|y-x\|_x}$$

$$\Leftrightarrow \quad (1-\|y-x\|_x)^2 \leq \frac{1}{\|H_x(y)^{-1}\|_x} \quad \text{and} \quad \|H_x(y)\| \leq \frac{1}{(1-\|y-x\|_x)^2}$$

2nd equiv: obs $I - H_x(y)$ has eigvals $1-\lambda_n \leq \cdots \leq 1-\lambda_1$

$$\|I - H_x(y)\|_x \leq \max\{\lambda_n - 1, 1-\lambda_1\}$$
$$\leq \max\left\{\lambda_n - 1, \frac{1}{\lambda_1} - 1\right\}$$

Examples of self-concordant functions:

* linear, quadratic fcns.

* "log barriers".

$$x \in \mathbb{R}^n_{++} \mapsto -\sum_{i=1}^n \log x_i$$

$$X \in S^n_{++} \mapsto -\log\det X.$$

Calculus of self-concordant fcns.

* $f \in SC(\mathbb{E})$, $\mathcal{L} : \mathbb{E} \to \mathbb{E}$ bijective linear

  $\Rightarrow \hat{f}(x) := f(\mathcal{L}x + y_0)$ (where $y_0 \in \mathbb{E}$) is s.c.

* $\alpha \geq 1$, $f \in SC(\mathbb{E}) \Rightarrow \alpha f \in SC(\mathbb{E})$.

* $f_1, f_2 \in SC(\mathbb{E}) \Rightarrow f_1 + f_2 \in SC(\mathbb{E})$

* $f \in SC(\mathbb{E})$, $c \in \mathbb{E} \Rightarrow f + \langle c, \cdot \rangle \in SC(\mathbb{E})$

Self concordancy & Newton's method

* Quad models approximate s.c. fcns well (locally).

[Thm 2.2.2, Renegar]

If $f \in SC(\mathbb{E})$, $x \in \operatorname{dom} f$,

then $\forall y \in B_x(x,1)$,

$$|f(y) - q_x(y)| \leq \frac{\|y-x\|_x^3}{3(1-\|y-x\|_x)},$$

where

$$q_x(y) := f(x) + \langle g(x), y-x \rangle + \frac{1}{2}\langle y-x, H(x)(y-x)\rangle$$

$$= f(x) + \langle g_x(x), y-x\rangle_x + \frac{1}{2}\|y-x\|_x^2.$$

Self-concordancy & Newton's method.

* Newton's method works well with self-concordant fcns:

  [Thm 2.2.3, Renegar]

  If $f \in SC(E)$, $x \in \mathrm{Dom} f$, z minimizes f and $z \in B_x(x,1)$,

  then Newton iterate $x_+ = x - H(x)^{-1} g(x)$ satisfies

$$\|x_+ - z\|_x \leq \frac{\|x-z\|_x^2}{1 - \|x-z\|_x} .$$

* In fact, if $x_1 = x_+$ & $\{x_i\}$ is the Newton sequence, we have

$$\|x_i - z\|_z \leq \frac{1}{4}\left(4\|x-z\|_z\right)^{2^i} .$$

Self-concordancy & Newton's method.

But we don't know z, most of the time, in advance...

[Thm 2.2.5, Renegar]

If $f \in SC(\mathbb{E})$ & $\|n(x)\|_x \leq 1/4$, where $n(x) := -g_x(x)$

$= -H(x)^{-1} g(x)$

for some $x \in \mathbb{E}$, then

① f has a minimizer $z \in \mathbb{E}$, and

②
$$\|z - x_+\|_x \leq \frac{3\|n(x)\|_x^2}{(1-\|n(x)\|_x)^3} .$$

Some more properties of s.c. fcns:

* [Thm 2.2.8, Renegar]

  If $f \in SC(\mathbb{E})$ & $\inf_{\mathbb{E}} f > -\infty$

  then f has a minimizer.

* Coercivity : [Thm 2.2.9, Renegar].

  If $f \in SC(\mathbb{E})$ and $\tilde{x} \in \partial \mathrm{Dom} f$,

  then $\forall$ seq $\{x_i\}_i \subseteq \mathrm{dom} f$ that conv to $\tilde{x}$,

  $$f(x_i) \to +\infty \quad \& \quad \|g(x_i)\| \to +\infty.$$

Barrier functions

$f : \mathbb{E} \to \mathbb{R}$ is a barrier fcn :

(1) $f \in SC(\mathbb{E})$

(2) $$\vartheta_f := \sup_{x \in \mathrm{dom} f} \| g_x(x) \|_x^2 < +\infty .$$

$*$ $-g_x(x) = -H(x)^{-1} g(x)$ is the Newton step of f at x.

ϑ_f is the complexity value of f.

$*$ [Nesterov et Nemirovskii] $\vartheta_f \geq 1 \quad \forall f \in SCB(\mathbb{E})$.

Also : If $f \in SCB(\mathbb{R}^n)$ is s.t. $\mathrm{dom} f = \mathbb{R}^n_{++}$, then $\vartheta_f \geq n$.

Common barrier fcns.

$f : x \in \mathbb{R}^n_{++} \mapsto -\sum_i \log x_i$ $\quad$ $\|g_x(x)\|_x^2 = n$

$\vartheta_f = n$

$f : X \in \mathbb{S}^n_{++} \mapsto -\ln\det X$ $\quad$ $\|g_X(X)\|_X^2 = n$

$\vartheta_f = n$

$f : x \in B(0,1) \subseteq \mathbb{R}^n$
$\mapsto -\ln(1 - \|x\|^2)$ $\quad$ $\|g_x(x)\|^2 = \dfrac{2\|x\|^2}{1 + \|x\|^2}$

$\vartheta_f = 1$

$f : (x,t) \in SOC \subseteq \mathbb{R}^{n+1}$
$\mapsto -\ln(t - \|x\|^2)$ $\quad$ $\vartheta_f = 2$

Basic calculus of barrier fcns.

* [Thm 2.3.1, Renegar]

  If $f_1, f_2 \in SCB(\mathbb{E})$, $\operatorname{dom} f_1 \cap \operatorname{dom} f_2 \neq \emptyset$

  then (1) $f_1 + f_2 \in SCB(\mathbb{E})$

  (2) $\vartheta_{f_1+f_2} \leq \vartheta_{f_1} + \vartheta_{f_2}$

* $f \in SCB(\mathbb{E})$ & $\alpha \geq 1$

  $\Rightarrow \alpha f \in SCB(\mathbb{E})$

* But $f \in SCB(\mathbb{E}) \not\Rightarrow f + \langle c, \cdot \rangle \in SCB(\mathbb{E})$.

A very useful property of barrier fcns

[Thm 2.3.3, Renegar]

If $f \in SCB(\mathbb{E})$ & $x, y \in \operatorname{dom} f$,

then $$\langle g(x), y - x \rangle < \vartheta_f .$$

* The inner prod is arbitrary; in fact,

$$\langle g(x), y-x \rangle = \langle H(x)^{-1} g(x), H(x)(y-x) \rangle$$

$$= \langle g_x(x), y-x \rangle_x .$$

Proving $\langle g_x(x), y-x \rangle_x < \theta_f$ for fixed $x, y \in \operatorname{dom} f$:

Let $v(t) = x + t(y-x)$,

$\phi(t) = f(x+t(y-x)) = f(v(t))$.

Note that $0, 1 \in \operatorname{dom} \phi$. Wlog assume $\phi'(0) > 0$.

$\forall\, t \in \operatorname{dom}\phi \cap \mathbb{R}_+$, $\quad \theta_f \geq \| g_{v(t)}(v(t)) \|^2_{v(t)}$

$$\geq \frac{\left[\langle g_{v(t)}(v(t)), y-x \rangle_{v(t)}\right]^2}{\| y-x \|^2_{v(t)}}$$

$$= \frac{\phi'(t)^2}{\phi''(t)}$$

$\therefore \forall\, s \in \operatorname{dom}\phi \cap \mathbb{R}_+$, $\quad \int_0^s \frac{\phi''(t)}{\phi'(t)^2}\, dt \geq \int_0^s \frac{1}{\theta_f}\, dt = \frac{s}{\theta_f}$

$$\therefore \quad \frac{1}{\phi'(0)} - \frac{1}{\phi'(s)} \geq \frac{s}{\theta_f}$$

$$\Rightarrow \quad \phi'(s) \geq \frac{\theta_f \cdot \phi'(0)}{\theta_f - \phi'(0) \cdot s}$$

$$\therefore \quad \frac{\theta_f}{\phi'(0)} \notin \operatorname{dom} \phi \quad , \quad \text{but} \quad (0,1) \subseteq \operatorname{dom} \phi$$

$$\Rightarrow \quad \frac{\theta_f}{\phi'(0)} > 1$$

$$\therefore \quad \theta_f > \phi'(0) = \langle g_x(x) \, , \, y-x \rangle_x \, .$$

$\blacksquare$

Barrier fcns and convex sets.

* [Nesterov et Nemirovskii]

  Every open convex set $S \subseteq \mathbb{E}$ containing no lines is the domain of some $f \in SCB(\mathbb{E})$.

  Moreover, $\exists$ universal constant C (indep. of n) s.t. $\theta_f \leq C \cdot n$.

Complexity value & the geometry of dom f.

* $z \in \mathbb{E}$ is the analytic center of $f \in SCB(\mathbb{E})$:

  z solves $\min_{x \in \mathrm{dom} f} f(x)$.

* [Corollary 2.3.5, Renegar]

  If $f \in SCB(\mathbb{E})$ & z is the analytic center of f,

  then $B_z(z, 1) \subseteq \mathrm{dom} f \subseteq B_z(z, 4\vartheta + 1)$.

Optimization over dom f ($f \in SCB$)

$$\text{val} = \min \langle c, x \rangle \quad \text{s.t.} \quad x \in \overline{\text{dom} f}.$$

(for example, $\text{dom} f = \{x : Ax = b, x > 0\}$ where $f : \{x : Ax = b\} \to \mathbb{R}$.)

* Central path : minimizers $z(\eta)$ of

$$f_\eta(x) = \eta \langle c, x \rangle + f(x)$$

1st order cond :

$$\eta c + g(z(\eta)) = 0.$$

* Thm 2.3.3 $\Longrightarrow$ $\forall y \in \text{dom} f$,

$$\langle c, z(\eta) \rangle - \langle c, y \rangle = \frac{1}{\eta} \langle g(z(\eta)), y - z(\eta) \rangle \leq \frac{1}{\eta} \vartheta_f.$$

$$\therefore \langle c, z(\eta) \rangle \leq \text{val} + \frac{1}{\eta} \vartheta_f$$