Problem: $\min_{x \in \mathbb{R}^n} f(x)$

- $f: \mathbb{R}^n \to \mathbb{R}$ differentiable, may be non-convex
- Assume $f^* = \inf_{x \in \mathbb{R}^n} f(x) > -\infty$.
- Find stationary point: $\|f'(\bar{x})\| \le \varepsilon$

E.g. $\boxed{B = I}$

Let fix $B = B^T \succ 0$, we use:

$$\|h\| = \langle Bh, h \rangle^{1/2}, \quad h \in \mathbb{R}^n \qquad \|h\| = \langle \overbrace{B^{1/2}h}^{u}, B^{1/2}h \rangle^{1/2} = \|B^{1/2}h\|_2$$

$$\|g\|_* = \langle g, B^{-1} g \rangle^{1/2}, \quad g \in \mathbb{R}^n \qquad \|g\|_* = \|B^{-1/2} g\|_2$$

$$A = A^T: \quad \|A\| = \max_{s \in \mathbb{R}^n : \|s\| \le 1} \|As\|_* = \max_{\substack{h \in \mathbb{R}^n \\ \|h\| \le 1}} \langle Ah, h \rangle =$$

$$= \max_{\substack{u \in \mathbb{R}^n \\ \|u\|_2 \le 1}} \langle B^{-1/2} A B^{-1/2} u, u \rangle = \max\{ \lambda_{max}(B^{-1/2} A B^{-1/2}), -\lambda_{min}(B^{-1/2} A B^{-1/2}) \}$$

- We assume the Hessian of f is Lipschitz continuous with constant $L > 0$:

$$\|f''(x) - f''(y)\| \le L\|x - y\| \qquad \forall x, y \in \mathbb{R}^n$$

$$\Updownarrow$$

$$\langle (f''(x) - f''(y))h, h \rangle \le L\|x-y\|, \quad \forall h \in \mathbb{R}^n : \|h\| \le 1.$$

$\Leftrightarrow$

$$-L\|x-y\|B + f''(x) \preceq f''(y) \preceq f''(x) + L\|x-y\|B$$

Taylor Approximation Bound

Lemma $\forall x, y \in \mathbb{R}^n$

(1) $$\|f'(y) - f'(x) - f''(x)(y-x)\|_* \le \frac{L}{2}\|y-x\|^2$$

(2) $$|f(y) - Q(x;y)| \le \frac{L}{6}\|y-x\|^3$$

$$Q(x;y) = f(x) + \langle f'(x), y-x\rangle + \frac{1}{2}\langle f''(x)(y-x), y-x\rangle.$$

Proof. By Taylor theorem, $h \in \mathbb{R}^n$ $\|h\| \le 1$

$$\langle f'(y) - f'(x) - f''(x)(y-x), h\rangle =$$

$$= \int_0^1 \langle (f''(x+\tau(y-x)) - f''(x))(y-x), h\rangle \, d\tau \le$$

$$\le \int_0^1 \|f''(x+\tau(y-x)) - f''(x)\| \cdot \|y-x\| \, d\tau \le$$

$$\le L\|y-x\|^2 \cdot \int_0^1 \tau \, d\tau = \frac{L}{2}\|y-x\|^2.$$

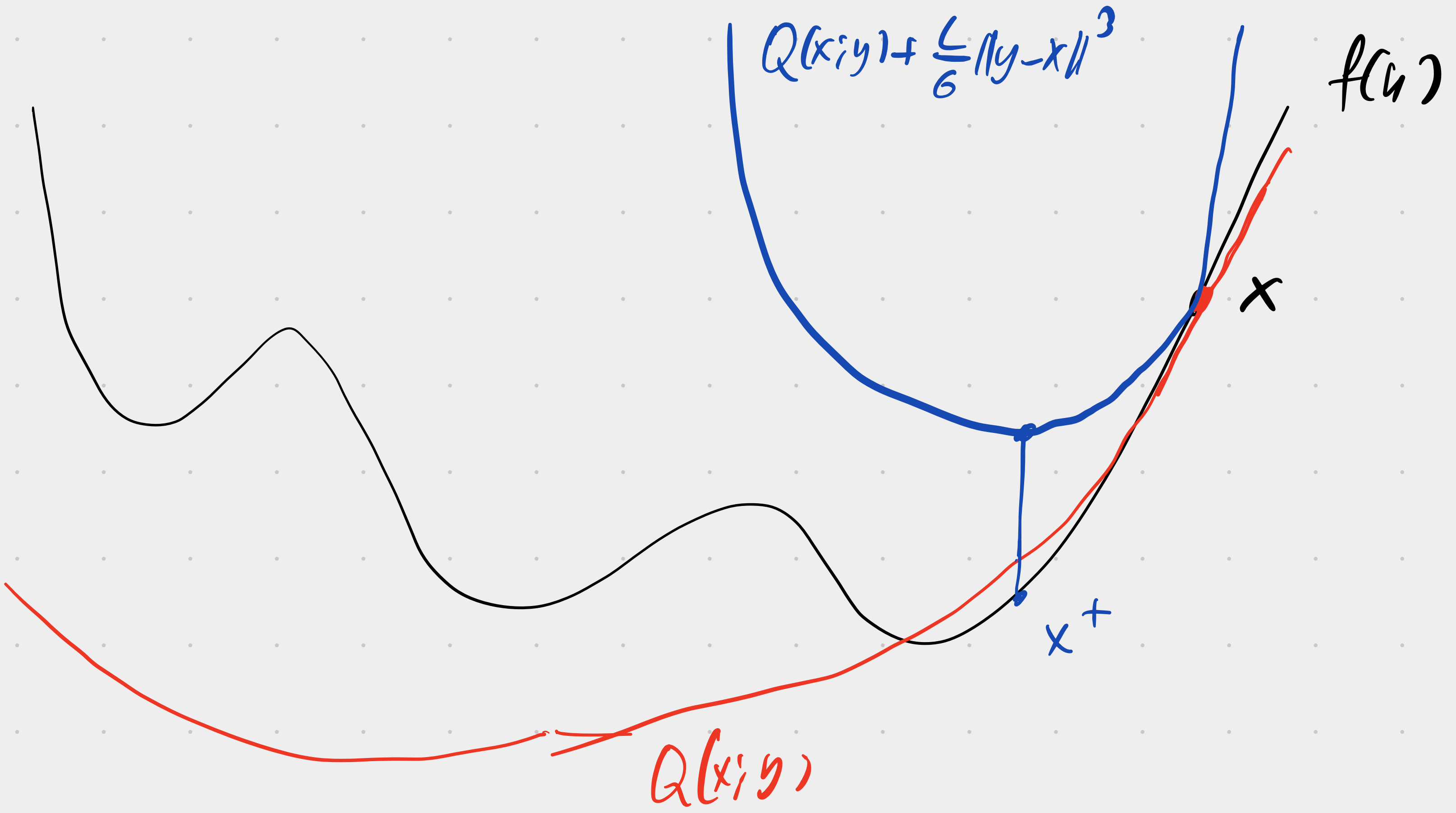
$$|f(y) - Q(x;y)| = |\underbrace{f(y) - f(x) - \langle f'(x), y-x\rangle}_{=} - \langle \tfrac{1}{2} f''(x)(y-x), y-x\rangle|$$

$$= \left| \int_0^1 (1-\tau)\left[\langle f''(x+\tau(y-x))(y-x), y-x\rangle - \langle f''(x)(y-x), y-x\rangle\right] d\tau \right| \le$$

$$\le L\|y-x\|^3 \int_0^1 (1-\tau)\tau \, d\tau = \frac{L}{6}\|y-x\|^3. \qquad \square$$

Global Upper Model:

$$f(y) \le Q(x;y) + \frac{L}{6}\|y-x\|^3 \qquad \forall x,y \in \mathbb{R}^n$$



Idea:

$$x^+ = \operatorname*{argmin}_y \Big[\underbrace{Q(x;y) + \frac{M}{6}\|y-x\|^3}_{\Omega(x;y)} \Big]$$

— non-convex function in y

- $M=0$: classic Newton's step
- $M \geq L$: Global Progress $f(x) - f(x^+) \geq \dots$ progress?

Main question: how to compute x^+?

(1) If f is convex $f''(x) \succeq 0 \Rightarrow \Omega(x;y)$ – convex function w.r.t. y ✓

x^+ is unique (apply gradient method; IPM; ....)

(2) In fact, it's enough $\nabla_y \Omega(x;x^+) \approx 0$

(3) Actually, we can always compute x^+ – global minimizer of $\Omega(x;y)$ in y.

$$\min_y \left[\langle f'(x), y-x\rangle + \frac{1}{2}\langle f''(x)(y-x), y-x\rangle + \frac{M}{6}\|y-x\|^3 \right]$$

let x^+ be stationary point:

$$\underbrace{f'(x) + f''(x)(x^+-x)} + \frac{M}{2} r B(x^+-x) = 0$$

$$r = \|x^+-x\|$$

Use that

$$\|f'(x^+) - f'(x) - f''(x)(x^+-x)\|_* \leq \frac{L}{2} r^2$$

We get:

$$\|f'(x^+) + \frac{M}{2} r B(x^+-x)\|_* \leq \frac{L}{2} r^2$$

Triangle inequality: $\|f'(x^+)\|_* \leq \frac{L}{2} r^2 + \frac{M}{2} r^2 = \frac{L+M}{2} r^2.$

<u>Lemma</u>

$$\|f'(x^+)\|_* \leq \frac{L+M}{2} r^2 = \frac{L+M}{2}\|x^+-x\|^2, \quad \forall M \geq 0$$

GM: $x^+ - x = -\gamma f'(x)$

$\|x^+-x\| = \gamma \|f'(x)\|$

Strongly Convex Functions

$$f''(x) \succeq \mu B \quad \forall x \in \mathbb{R}^n, \quad \mu > 0.$$

$$f'(x) + f''(x)(x^+-x) + \frac{M}{2} r B(x^+-x) = 0 \qquad \langle \cdot, x^+-x\rangle$$

$$\underbrace{\langle f''(x)(x^+-x), x^+-x\rangle}_{\substack{\vee\!\shortparallel \\ \mu r^2}} + \frac{M}{2} r \underbrace{\langle B(x^+-x), x^+-x\rangle}_{\substack{r^3 \\ \vee\!\shortparallel \\ 0}} = \underbrace{\langle f'(x), x-x^+\rangle}_{\leq \|f'(x)\|_* \cdot r}$$

$\Rightarrow$ For strongly functions: $r \leq \frac{\|f'(x)\|_*}{\mu}$

Theorem

$$\|f'(x^+)\|_* \leq \frac{L+M}{2\mu^2}\|f'(x)\|_*^2 = \underbrace{\frac{L+M}{2\mu^2}\|f'(x)\|_*}_{<1}\cdot\|f'(x)\|_*$$

$\Rightarrow$ local quadratic convergence!

Region of quadratic convergence:

$$Q = \left\{ x : \|f'(x)\|_* < \frac{2\mu^2}{L+M} \right\}.$$

- $M=0$: Classic Newton: $Q = \left\{ x : \underset{\Uparrow}{\|f'(x)\|_*} < \frac{2\mu^2}{L} \right\}$

Self-concordant function: $\lambda(x) = \langle f'(x), f''(x)^{-1} f'(x)\rangle^{1/2} < \frac{2}{M}$

constant of self-concord.

$\Rightarrow \lambda(x^+) \leq M \cdot \lambda(x)^2$

Computing Cubic Step $\quad A = A^T \not\succeq 0.$

$$\min_{y \in \mathbb{R}^n} \left\{ P(y) = \langle g, y\rangle + \frac{1}{2}\langle Ay, y\rangle + \varphi\big(\underbrace{\langle By, y\rangle}_{\|y\|^2}\big)\right.$$

$$\varphi(s) = \frac{M}{6} s^{3/2}, \quad s \geq 0.$$

Trust-region methods $\quad \varphi(s) = \begin{cases} 0, & s \leq R^2 \\ +\infty, & \text{otherwise} \end{cases}$

Conjugate Function:

$$\varphi^*(\tau) = \max_{s>0} [\tau s - \varphi(s)]$$

$\varphi^{**}(s) = \varphi(s)$ when φ closed convex proper

$$\varphi(s) = \max_{\tau>0} [\tau s - \varphi^*(\tau)]$$

$\varphi(s) = \frac{M}{6} s^{3/2}$, $\varphi^*(\tau) = \frac{2^4}{3M^2}\tau^3$

$$\min_y P(y) = \min_{y\in\mathbb{R}^n} \max_{\tau>0} \Big[\langle g,y\rangle + \tfrac{1}{2}\langle Ay,y\rangle + \underbrace{\tfrac{\tau}{2}\langle By,y\rangle - \varphi^*(\tfrac{\tau}{2})}_{\varphi(\langle By,y\rangle)}\Big]$$

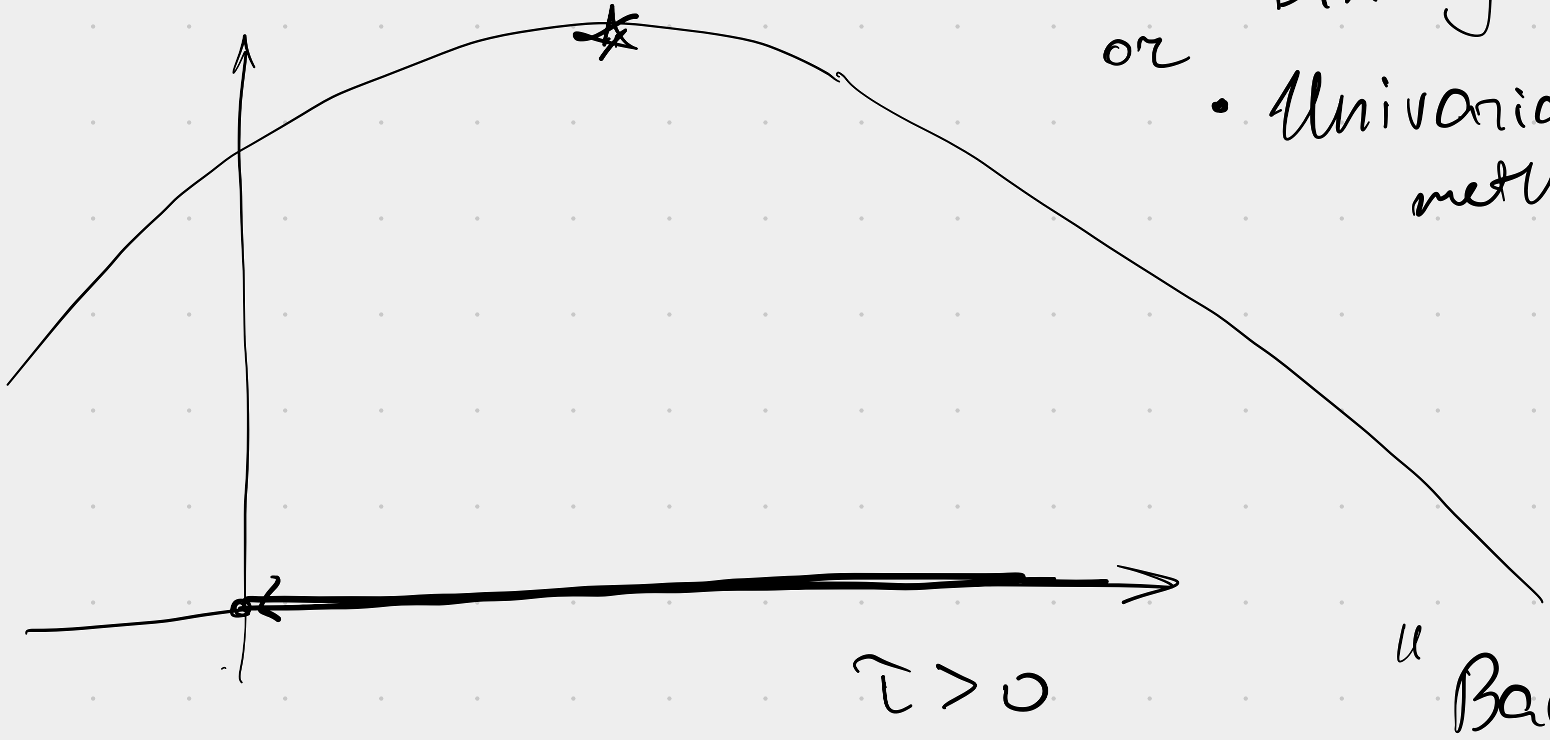
$\boxed{=} \geq \max_{\tau\in W} \Phi(\tau)$ $\quad \boxed{W = \{\tau > 0 : A+\tau B \succ 0\}} \subset \mathbb{R}$

$(A+\tau B)y^* = -g$

$$\Phi(\tau) = \min_{y\in\mathbb{R}^n} \Big[\langle g,y\rangle + \tfrac{1}{2}\langle (A+\tau B)y,y\rangle\Big] - \varphi^*(\tfrac{\tau}{2})$$

$$= -\tfrac{1}{2}\langle g, (A+\tau B)^{-1} g\rangle - \varphi^*(\tfrac{\tau}{2})$$

$$= \underline{-\tfrac{1}{2}\langle g, (A+\tau B)^{-1} g\rangle} - \underline{\frac{2^4}{3M^2}\left(\frac{\tau}{2}\right)^3} \text{ - concave}$$

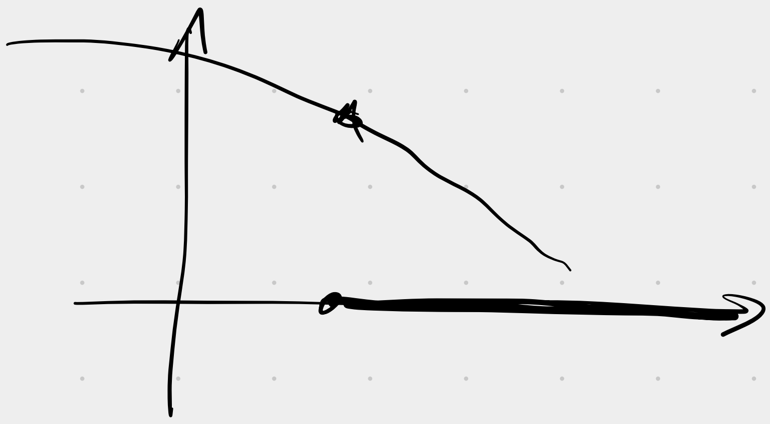


- Binary search in τ

or

- Univariate Newton's method

"Bad case"



Computing Cubic Step:

(1) Solve the Dual Problem:

$$\max \Phi(\tau) \longrightarrow \tau^* = \frac{M}{2} r^*$$

$$f'(x) + f''(x)(x^+ - x) + \frac{M}{2} r B(x^+ - x) = 0$$

$$\Updownarrow$$

(2) $$x^+ - x = -\left(f''(x) + \frac{M}{2} r^* B\right)^{-1} f'(x)$$

$$r = \|x^+ - x\| = \left\|\left(f''(x) + \frac{M}{2} r B\right)^{-1} f'(x)\right\|_x$$