**Motivation:** First-order vs. Second-order methods.

$$\min_{x \in Q} f(x)$$

- Unconstrained min. $Q = \mathbb{R}^n$, $Q = \{x : Ax = b\}$ – Affine constraint
- $Q$ - complicated $\{Ax \le b\}$, $f(x) = \langle c, x \rangle$

## Gradient Methods $f(x), f'(x)$

- $x^+ = x - \eta f'(x)$
- $d'(x^+) = d'(x) - \eta f'(x)$ Mirror descent
- Composite problems

$$x^+ = \operatorname{argmin}_y \{ \langle f'(x), y \rangle + \beta_d(x; y) + \psi(y) \}$$

## Second-order methods

Main assumption: we can solve

$Ax = b$ $\longrightarrow$ linear algebra (LU, QR, Cholesky, ...)

$\longrightarrow$ first-order methods matrix-vector products $Ah$

Oracle: $f(x), f'(x), f''(x) \in \mathbb{R}^{n \times n}$

# Newton's Method

Idea:

$$f(y) = \underbrace{f(x) + \langle f'(x), y-x\rangle + \frac{1}{2}\langle f''(x)(y-x), y-x\rangle} + \bar{\bar{o}}(\|y-x\|^2)$$

Next point:

$$x^+ = \underset{y \in \mathbb{R}^n}{\operatorname{argmin}} \; \langle f'(x), y\rangle + \frac{1}{2}\langle f''(x)(y-x), y-x\rangle$$

Assume: $f''(x) \succ 0 \iff f$ is strictly convex.

$$f'(x) + f''(x)(x^+ - x) = 0 \qquad \longleftarrow \text{linear system}$$

$$\iff \quad x^+ = x - f''(x)^{-1} f'(x) \qquad \text{Newton's step.}$$

N.B.: Gradient Methods w.r.t. $\|y\| = \langle By, y\rangle^{1/2}$

$$x^+ = x - \gamma B^{-1} f'(x)$$

In practice:

- Damped Newton step:

$$x^+ = x - \gamma f''(x)^{-1} f'(x), \quad \gamma > 0$$

- Regularization:

$$x^+ = x - (f''(x) + \delta I)^{-1} f'(x), \quad \delta \geq 0$$

- Inexact Newton-type methods:

$$f''(x) \mapsto H \approx f''(x)$$

Local Quadratic Convergence: $\gamma \to 1$

$\delta \to 0$

# Affine-invariance

$f(x)$

$$x^+ = x - f''(x)^{-1} f'(x)$$

Consider $x = Ay + b$, $A$ is invertable

Define $F(y) = f(Ay + b) = f(x)$

$$y^+ = y - F''(y)^{-1} F'(y)$$

**Proposition** Newton's step is affine-invariant:

$$x = Ay + b$$

Then $\quad x^+ = Ay^+ + b.$

**Proof**

$$F'(y) = A^T f'(Ay + b) = A^T f'(x)$$
$$F''(y) = A^T f''(Ay + b) A = A^T f''(x) A$$

$$y^+ = y - F''(y)^{-1} F'(y) = y - A^{-1} f''(x)^{-1} A^{-T} A^T f'(x)$$
$$= y - A^{-1} f''(x)^{-1} f'(x)$$

Therefore,

$$Ay^+ + b = Ay + b - f''(x)^{-1} f'(x) = x - f''(x)^{-1} f'(x) = x^+$$

$\square$

**Remark.** Note for the gradient method (w.r.t. Euclidean norm)

$$x^+ = x - \eta f'(x),$$

it's invariant under $x = Uy + b$, $UU^T = I$ — orthogonal,

But not for an arbitrary $A$.

# Self-concordant Functions.

Consider $f: Q \to \mathbb{R}$ , $Q \subseteq \mathbb{R}^n$ - open convex set;
$f$ is strictly convex, three times differentiable

## Euclidean Structure

$$x \mapsto f''(x) = f''(x)^T \succ 0$$

$$\langle u, v \rangle_x := \underbrace{\langle f''(x) u, v \rangle}_{\text{"standard Euclidean norm"}} = D^2 f(x)[u, v] \quad , \quad \forall u, v \in \mathbb{R}^n$$

it doesn't depend on a coordinate system.

## Local Norm:

$$\|h\|_x := \sqrt{\langle h, h \rangle_x} = \sqrt{\langle f''(x) h, h \rangle} \quad , \quad h \in \mathbb{R}^n, \; x \in Q$$

## Third Derivative

$$\|h\|_x \approx \|h\|_y \quad \text{when } x \approx y.$$

Let $h \in \mathbb{R}^n$ be fixed.

Define $\varphi(t) = \|h\|^2_{x+tu}$ , $u \in \mathbb{R}^n$ - fixed direction,
$t > 0$ - small parameter
$x + tu \in Q$

$\varphi(t) = \langle f''(x+tu) h, h \rangle$

$$\varphi'(0) =: D^3 f(x)[h, h, u] \in \mathbb{R}$$

$D^3 f(x)$ - symmetric three-linear form.

**Def** We say that $f$ is <u>self-concordant</u> with $M \geq 0$:

$$|D^3 f(x)[h, h, u]| \leq M \cdot \|h\|_x^2 \cdot \|u\|_x \quad \forall h, u \in \mathbb{R}^n, \; x \in Q$$

Note, $u := h$ : $\Updownarrow$

$$|D^3f(x)[h,h,h]| \le M \cdot \|h\|_x^3 = M \cdot \langle f''(x)h, h\rangle^{3/2} \quad \forall h \in \mathbb{R}^n,\ x \in Q$$

$$D^3f(x)[h,h,u] = D^3f(x)[u,h,h]$$

Let $u$ be fixed, then $D^3f(x)[u,h,h] = \langle A_u h, h\rangle$.

$\exists$ a symmetric matrix $A_u \in \mathbb{R}^{n\times n}$

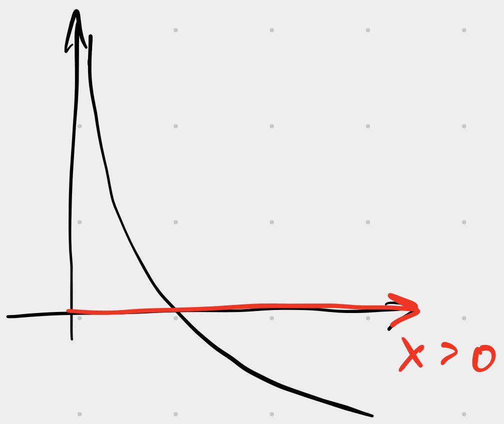
$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}\ ; \quad \frac{\partial^3 f(x)}{\partial x_i \partial x_j \partial x_k} = \frac{\partial^3 f(x)}{\partial x_i \partial x_k \partial x_j} = \dots$$

## Examples

1. $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$

$f''(x) = A \qquad f'''(x) = 0 \qquad \Rightarrow \boxed{M = 0}$

2. $f(x) = -\ln x$ Negative Logarithm




"barrier" for $\mathbb{R}_{>0}$

$f'(x) = -\frac{1}{x}$

$f''(x) = \frac{1}{x^2} > 0 \qquad |f''(x)| \overset{?}{\le} L \to +\infty$

$f'''(x) = -\frac{2}{x^3}$

$$|D^3f(x)[h,h,h]| \overset{?}{\le} M \langle f''(x)h, h\rangle^{3/2} \qquad h \in \mathbb{R}$$

$$|f'''(x) \cdot h^3| \le M \cdot (f''(x) \cdot h^2)^{3/2} = M f''(x)^{3/2} \cdot h^3$$

$$\Rightarrow \quad |f'''(x)| \le M \cdot f''(x)^{3/2}$$

$$\left|-\frac{2}{x^3}\right| = \frac{2}{x^3} \overset{?}{\leq} M \cdot \frac{1}{x^3} \Rightarrow \boxed{M = 2}$$

"Standard self-concordant function"

3. let $f$ have Lipschitz Hessian:

$$\|f''(x) - f''(y)\| \leq L\|x-y\| \Rightarrow \|D^3 f(x)\| \leq L$$

$f$ is strongly convex:

$$f''(x) \succeq \mu I \Rightarrow \mu\|h\|^2 \leq \|h\|_x^2 \quad \forall h.$$

$$D^3 f(x)[h,h,h] \leq L\|h\|^3 \leq \frac{L}{\mu^{3/2}} \|h\|_x^3$$

$$\Rightarrow \boxed{M = \frac{L}{\mu^{3/2}}} \qquad M = \inf_{\substack{\text{choice} \\ \text{of norm } \|\cdot\|}} \frac{L_{\|\cdot\|}}{\mu_{\|\cdot\|}^{3/2}}.$$