Summary

1. Global Opt. $\mathcal{F} = \{ f : \mathbb{R}^n \to \mathbb{R} : f \text{ is continuous} \}$
diff

Global solution? $f(\bar{x}) - f^* \le \varepsilon$

2. $\mathcal{F} = \{ -//- \}$

The goal: $\bar{x}$: $\|f'(\bar{x})\| \le \varepsilon$

Option 1 : something in between?

Option 2 : $\mathcal{F}' \subset \mathcal{F}$: for $f'(\bar{x}) = 0 \Rightarrow \bar{x}$ - global sol.

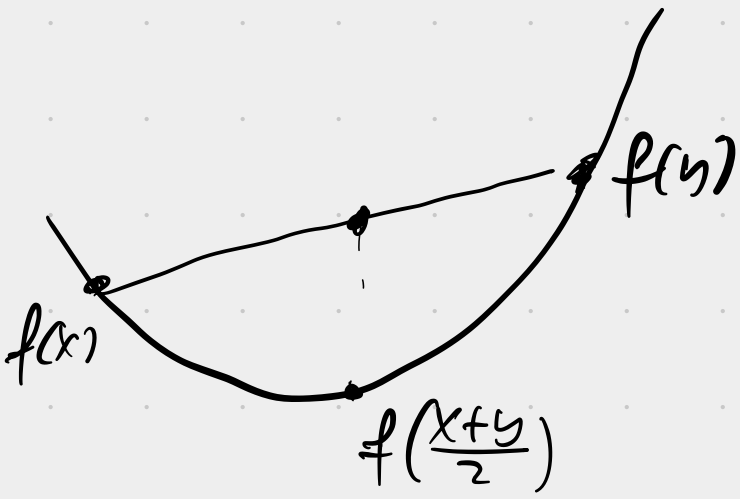
$\Rightarrow$ convexity.

Convex Functions

$f : \mathbb{R} \to \mathbb{R}$ is continuous.

f is convex if $x, y \in \mathbb{R}$:

$$f\left(\frac{x+y}{2}\right) \leq \frac{1}{2}\left(f(x) + f(y)\right)$$



Prop. 1 $x, y \in \mathbb{R}$ $\quad \lambda \in [0,1]$:

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$$

Prop. 2 (Jensen's inequality)

$x_1, \dots, x_N$ $\quad \lambda \in \Delta^N = \{\lambda \in \mathbb{R}_+^N : \langle e, \lambda \rangle = 1\}$

$$f\left(\sum_{i=1}^{N} \lambda_i x_i\right) \leq \sum_{i=1}^{N} \lambda_i f(x_i)$$

Prop. 3 ξ – random variable. Then

$$f(\mathbb{E}\xi) \leq \mathbb{E} f(\xi).$$

Maximizing Convex Function

Theorem The following conditions are equivalent:

- f is convex.
- For any $[x,y]$ and for any affine function $t \mapsto at+b$ we have

$$\max_{t\in[x,y]} f(t)-at-b = \max\{f(x)-ax-b,\ f(y)-ay-b\}.$$

Proof. ($\Rightarrow$) $t\in[x,y] \Rightarrow t=\lambda x+(1-\lambda)y,\ 0\leq\lambda\leq 1$

$$f(t)-at-b \leq \lambda[f(x)-ax-b]+(1-\lambda)[f(y)-ay-b]$$
$$\leq \max\{f(x)-ax-b,\ f(y)-ay-b\}.$$

($\Leftarrow$)

$$f(\underbrace{\lambda x+(1-\lambda)y}_{t})-\lambda f(x)-(1-\lambda)f(y) \overset{?}{\leq} 0$$

$$t=y+\lambda(x-y) \iff \lambda=\frac{t-y}{x-y}$$

$$f(t)-\frac{t-y}{x-y}\cdot[f(x)-f(y)]-f(y)\leq 0$$

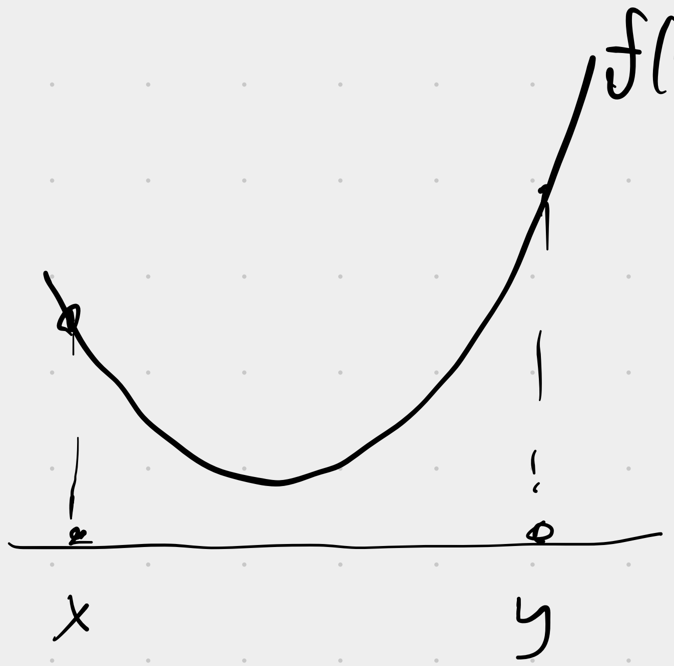
$$a=\frac{f(x)-f(y)}{x-y} \qquad b=f(y)-y\cdot a$$

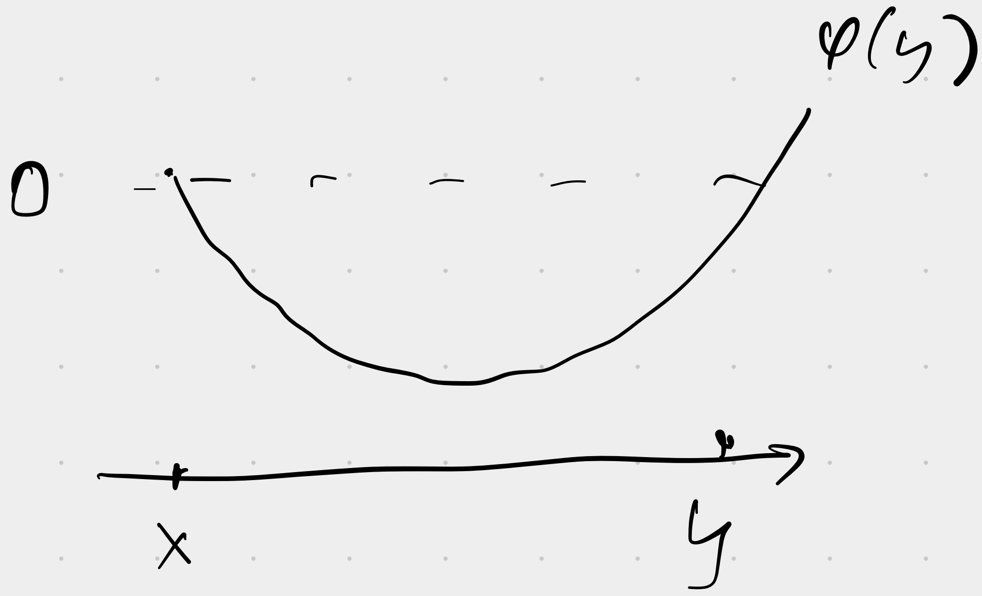
$$\varphi(t) = f(t) - at - b \le 0 \qquad \forall t \in [x,y]$$

$$\varphi(x) = f(x) - f(y) - \frac{x-y}{x-y}\left[f(x) - f(y)\right] = 0$$

$$\varphi(y) = 0$$

$$\max_{t\in[x,y]} \varphi(t) \le \max\{\varphi(x), \varphi(y)\} = 0.$$





$\square$

Multivariate Funct.

$f : \mathbb{R}^n \to \mathbb{R} \qquad \forall x, y:$

$$f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y), \quad 0 \le \lambda \le 1.$$

Prop. $K \subset \mathbb{R}^n$ – compact.

$$\max_{x\in K} f(x) = \max_{x \in \partial K} f(x).$$

f – concave if $-f$ is convex.

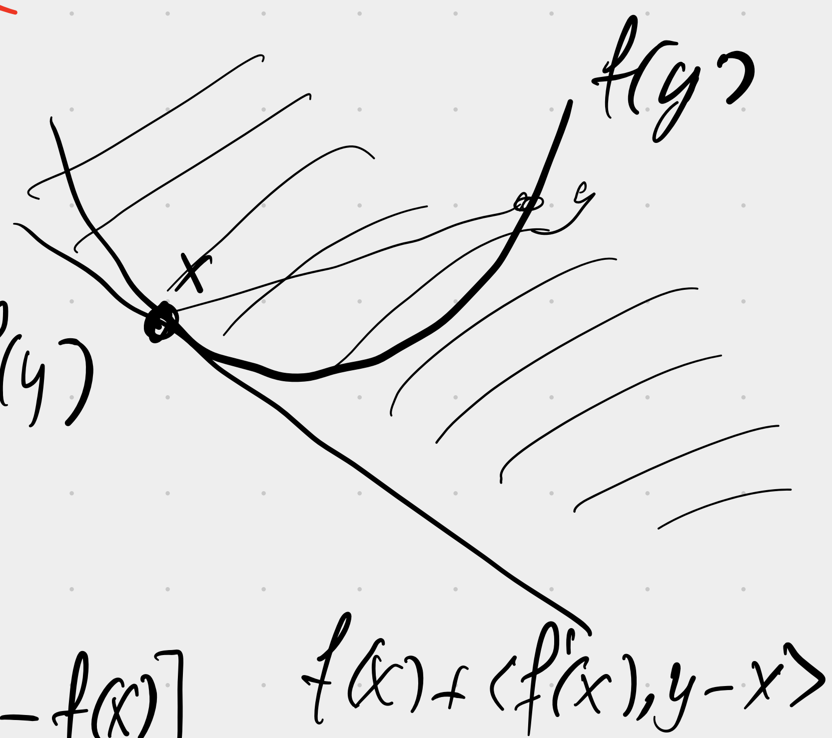
Differentiable Convex Functions.

Theorem let $f: \mathbb{R}^n \to \mathbb{R}$ be twice continuous diff.

Then, the following conditions are equivalent:

1. f is convex.
2. $\forall x, y: \quad f(y) \geq f(x) + \langle f'(x), y-x \rangle$
3. $\forall x, y: \quad \langle f'(y) - f'(x), y-x \rangle \geq 0$
4. $\forall x: \quad \nabla^2 f(x) \succeq 0$.

Proof



$$f((1-\alpha)x + \alpha y) \leq (1-\alpha) f(x) + \alpha f(y)$$

Rearrange:

$$f(y) \geq f(x) + \underbrace{\frac{1}{\alpha}\left[f(x+\alpha(y-x)) - f(x)\right]}_{\alpha \to 0 \quad \to \langle f'(x), y-x \rangle}$$

Exercise: To finish the proof.

Ex. $f(x) = e^x, \; -\ln x, \; x\ln x, \; |x|^p, \; p \geq 1, \ldots$

Global Optimality

Corollary Let $f: \mathbb{R}^n \to \mathbb{R}$ diff., convex. Then

$$f'(\bar{x}) = 0 \iff \bar{x} \text{ is a global minimum.}$$

Proof ($\Leftarrow$) proved.

($\Rightarrow$) $f(y) \geq f(\bar{x}) + \langle f'(\bar{x}), y - \bar{x} \rangle =$
$= f(\bar{x}) \qquad \forall y \in \mathbb{R}^n.$ $\square$

Theorem Let $\mathcal{F} \subset \{ f: \mathbb{R}^n \to \mathbb{R} \text{ diff.} \}$ be a maximal class s.t.

1. For any $f \in \mathcal{F}$: $f'(\bar{x}) = 0 \Rightarrow \bar{x}$ is a global minimum.
2. For any $f_1, f_2 \in \mathcal{F}$: $\alpha f_1 + \beta f_2 \in \mathcal{F}$, $\alpha \geq 0, \beta \geq 0$.
3. All affine functions $\langle a, x \rangle + b \in \mathcal{F}$.

Then, $\mathcal{F}$ = convex and diff. functions.

Proof $f \in \mathcal{F}$. Let $x \in \mathbb{R}^n$:

$$\varphi(y) = f(y) - \langle f'(x), y \rangle \in \mathcal{F}$$

$\varphi'(y) = f'(y) - f'(x) \Rightarrow \varphi'(x) = 0 \Rightarrow x$ is a global min of φ.

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

$$\Rightarrow \quad f(y) \geq f(x) + \langle f'(x), y-x\rangle. \qquad \square$$

Convex and Smooth Functions

- $f : \mathbb{R}^n \to \mathbb{R}$ is convex
- f is smooth: $\|f'(y) - f'(x)\|_* \leq L\|y-x\|.$

<u>Theorem</u>. The following conditions are equivalent:

1. f is convex and smooth
2. $0 \leq f(y) - f(x) - \langle f'(x), y-x\rangle \leq \frac{L}{2}\|y-x\|^2.$
3. $0 \leq \langle f'(y) - f'(x), y-x\rangle \leq L\|y-x\|^2$
4. $0 \leq \langle f''(x)h, h\rangle \leq L\|h\|^2 \ \forall x, h$ (if f is twice cont. diff.)
5. $f(y) \geq f(x) + \langle f'(x), y-x\rangle + \frac{1}{2L}\|f'(y) - f'(x)\|_*^2.$
6. $\langle f'(y) - f'(x), y-x\rangle \geq \frac{1}{L}\|f'(y) - f'(x)\|_*^2.$

<u>Proof</u>. $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4$ are trivial.

$4 \Rightarrow 2.$

Taylor's Formula:

$$f(y) - f(x) - \langle f'(x), y-x\rangle = \int_0^1 (1-\tau)\underbrace{\langle f''(x+\tau(y-x))(y-x), y-x\rangle}\, d\tau$$

$$0 \leq \cdot \leq L\|y-x\|^2.$$

5. $x = x^*$.

$$\boxed{f(y) - f^* \geq \frac{1}{2L}\|f'(y)\|_*^2.} \quad \forall y$$

$y \mapsto y^+(L)$ - gradient step.

$$f(y) - f^* \geq f(y) - f(y^+(L)) \geq \frac{1}{2L}\|f'(y)\|_*^2.$$

Fix x. **Define:** $\varphi(y) = f(y) - \langle f'(x), y\rangle$

$$\varphi(y) - \varphi^* \geq \frac{1}{2L}\|\varphi'(y)\|_*^2$$

$\varphi'(y) = f'(y) - f'(x)$. $\quad \varphi'(x) = 0 \quad \varphi^* = \varphi(x) = f(x) - \langle f'(x), x\rangle$

$$\cancel{f(y)} - \cancel{f(x)} - \langle f'(x), y - x\rangle \geq \frac{1}{2L}\|f'(y) - f'(x)\|_*^2$$

+

(5) ⇒ (6). $\quad \cancel{f(x)} - \cancel{f(y)} - \langle f'(y), x - y\rangle \geq \frac{1}{2L}\|f'(y) - f'(x)\|_*^2$

$$\langle f'(y) - f'(x), y - x\rangle \geq \frac{1}{L}\|f'(y) - f'(x)\|_*^2 . \geq 0.$$

$\Downarrow$

I. $\langle f'(y) - f'(x), y - x\rangle \geq 0 \Rightarrow f$ is convex.

II. Cauchy-Shwarz:

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

$\Rightarrow \|f'(y) - f'(x)\|_* \leq L\|y - x\|. \Rightarrow$ smooth. ■

Convergence of Gradient Method.

For one step:

$$x_{k+1} = \arg\min_y \left\{ f(x_k) + \langle f'(x_k), y - x_k \rangle + \frac{L}{2}\|y - x_k\|^2 \right\}$$

The progress:

$$f(x_k) - f(x_{k+1}) \geq \frac{1}{2L}\|f'(x_k)\|^2.$$

Convex

$$f^* \geq f(x_k) + \langle f'(x_k), x^* - x_k \rangle$$

$$\Leftrightarrow \quad F_k = f(x_k) - f^* \leq \langle f'(x_k), x_k - x^* \rangle$$

$$\leq \|f'(x_k)\|_* \cdot \|x_k - x^*\|$$

$$\leq \|f'(x_k)\|_* \cdot R$$

$\forall k \geq 0: \quad \|x_k - x^*\| \leq R.$

Euclidean: $\|x_{k+1} - x^*\| \leq \|x_k - x^*\| \leq \dots \leq \|x_0 - x^*\|$

$$R = \|x_0 - x^*\|.$$

We get:

$$\boxed{F_k - F_{k+1} \geq \frac{1}{2LR^2} \cdot F_k^2}$$

$F_k \to 0$

$$F_k \simeq \frac{1}{k}$$

(continuous time $\dot{\varphi}_t = -\text{const}\cdot\varphi_t^2$

$$\frac{d}{dt}\left[-\frac{1}{\varphi_t}\right] = \frac{\varphi_t'}{\varphi_t^2} = -\text{const}$$

Therefore: $\frac{1}{\varphi_t} \approx O(t)$. $\varphi_t \approx O(\frac{1}{t})$).

$$\frac{1}{F_{k+1}} - \frac{1}{F_k} = \frac{F_k - F_{k+1}}{F_{k+1}\cdot F_k} \geq \frac{1}{2LR^2}\cdot \frac{F_k^2}{F_{k+1}\cdot F_k} \geq \frac{1}{2LR^2}$$

$$F_k \geq F_{k+1}$$

Telescoping:

$$\frac{1}{F_k} \geq \frac{1}{F_0} + \frac{k}{2LR^2}.$$

<u>Theorem</u> Let f be convex and smooth.

Assume: $\|x_k - x^*\| \leq R \quad \forall k \geq 0$,

where $\{x_k\}$ – the sequence generated by the gradient method. Then:

$$f(x_k) - f^* \leq O\left(\frac{LR^2}{k}\right).$$