Lecture 4

Scribe?

Last time

- 2nd-order optimality cond.
- Basic convexity

Agenda

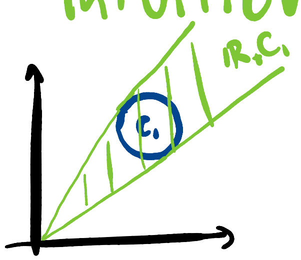
- More on convexity
- Characterization smooth convex.
- Subgradients.

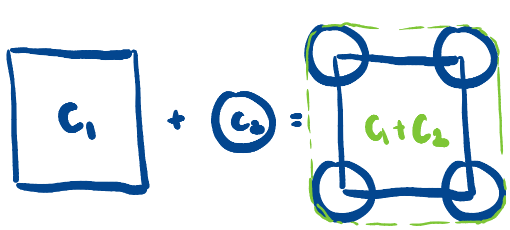
More on convexity

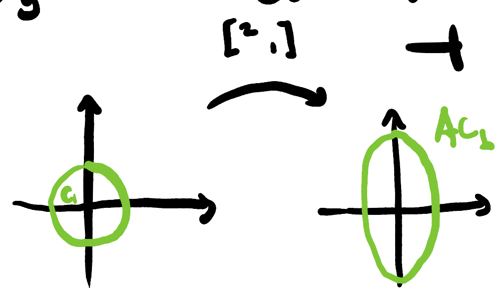
Lemma: Assume that $C_1, C_2 \in \mathbb{R}^d$ convex sets, $C_3 \in \mathbb{R}^n$.
Then, the following are convex

1. (Scaling) $\mathbb{R}_+ C_1 = \{\lambda \bar{x} \mid \lambda \geq 0 \text{ and } x \in C_1\}$
2. (Sums) $C_1 + C_2 = \{\bar{x}_1 + \bar{x}_2 \mid x_1 \in C_1, x_2 \in C_2\}$
3. (Intersections) $C_1 \cap C_2$.
4. (Linear images and preimages)
   Let $A: \mathbb{R}^d \to \mathbb{R}^n$ is linear,
   $A C_1$ and $A^{-1} C_3$ are convex.

Intuition







Proof: Exercise $\square$

Equivalence of operations

Function	Epigraph
$\lambda f(x/\lambda)$	$\lambda \operatorname{epi} f$
$\max_i f_i$	$\bigcap_i \operatorname{epi} f_i$
$f(Ax)$	$[A \times I]^{-1} \operatorname{epi} f$

Fill in as exercise

What are sum?

Lemma (First-order characterization of convexity)

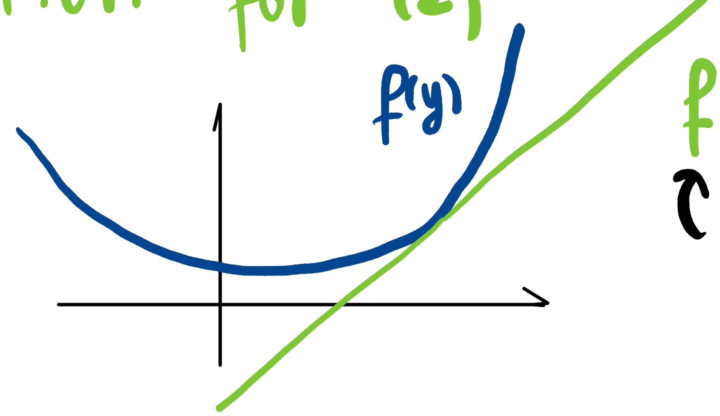
Suppose that $f: \mathbb{R}^d \to \mathbb{R}$ is differentiable. Then, the following are equivalent:

(1) f is convex.

(2) $\forall x, y \in \mathbb{R}^d \quad f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$

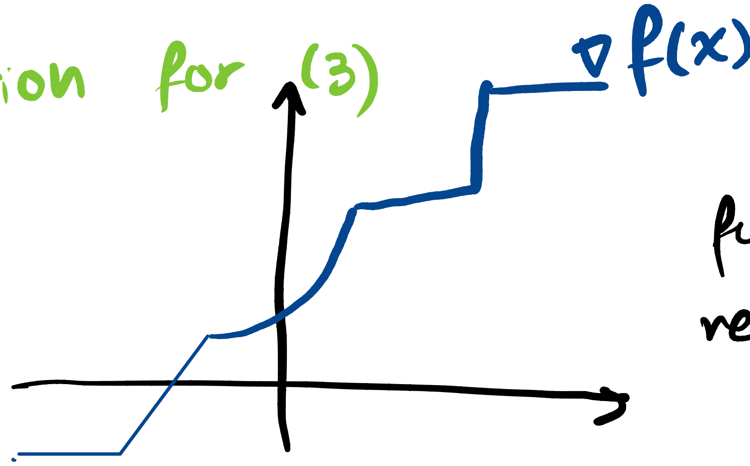
(3) $\forall x, y \quad \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0.$ $\dashv$

Intuition for (2)



$f(x) + \langle \nabla f(x), y - x \rangle$ supports the epigraph

Intuition for (3)



In 1D the function is monotone.

Proof: (1) $\Rightarrow$ (2) Let $x, y \in \mathbb{R}^d$ and $t \in (0, 1]$

Convexity ensures that

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

$$\Updownarrow$$

$$\frac{f(x + \lambda(y-x)) - f(x)}{\lambda} \leq f(y) - f(x)$$

Taking $\lambda \to 0$ $\Rightarrow$ $\langle f(x), x-y \rangle + f(x) \leq f(y)$.

(2) $\Leftarrow$ (1) Let $x, y \in \mathbb{R}$, $\lambda \in [0, 1]$ and

$z_t = (1-t)x + t y$

$\Rightarrow$ $f(x) \geq f(z_t) + \langle \nabla f(z_t), x - z_t \rangle$ ①

$f(y) \geq f(z_t) + \langle \nabla f(z_t), y - z_t \rangle$ ②

$\Rightarrow$ $(1-t)$ ① $+ t$ ② gives

$(1-t) f(x) + t f(y) \geq f(z_t) + \langle \nabla f(z_t), \underbrace{(1-t)x + t y - z_t}_{=0} \rangle$

$\geq f(z_t)$.

(2) $\Rightarrow$ (3)

$$
\begin{array}{rl}
 & f(x) \geq f(y) + \nabla f(y)^T(x-y) \\
+ & f(y) \geq f(x) + \nabla f(x)^T(y-x) \\
\hline
 & 0 \geq (\nabla f(x) - \nabla f(y))^T \cdot (y-x)
\end{array}
$$

(3) $\Rightarrow$ (2) Define $\varphi(t) = f(x + t(y-x))$

Then $f(y) = \varphi(1) = \varphi(0) + \int_0^1 \varphi'(t)\,dt$

$$= \varphi(0) + \varphi'(0) + \int_0^1 [\varphi'(t) - \varphi'(0)]\,dt$$

$$= f(x) + \nabla f(x)(y-x)$$

$$= f(x) + \nabla f(x)(y-x) + \overbrace{\int_0^1 \nabla f(x + t(y-x))^T (y-x)\, t\, dt}^{\geq 0}$$

$$\geq f(x) + \nabla f(x)(y-x)$$

$\square$

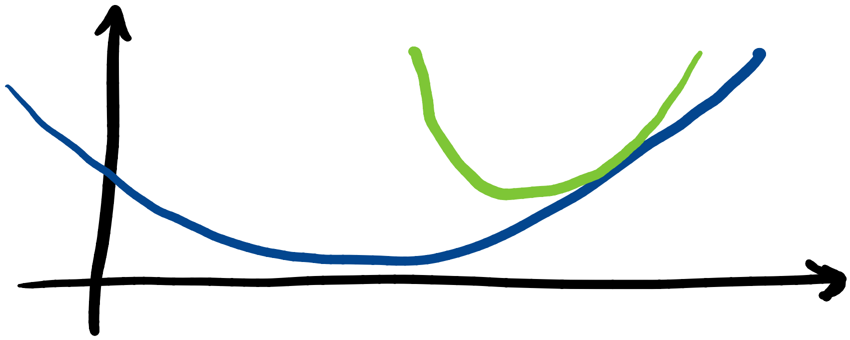
Lemma 2nd-order characterization

Assume f twice differentiable. Then,

f is convex $\Leftrightarrow$ $\nabla^2 f(x) \succeq 0 \quad \forall x.$

Intuition

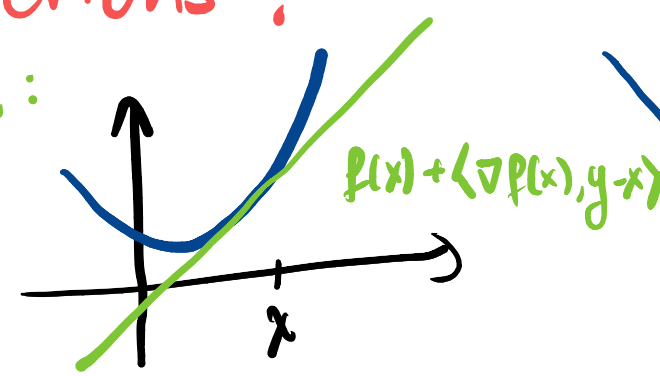
Second order model never curves down!

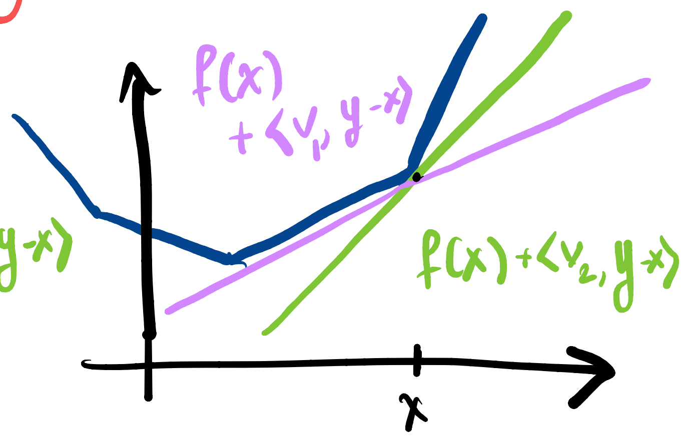


Proof Exercise. $\square$

Question: How can we assess optimality for general convex functions?

Idea:





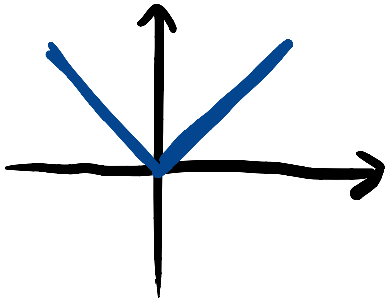
Def: Consider a convex function $f : \mathbb{R}^d \to \mathbb{R}$. The subdifferential of f at x is

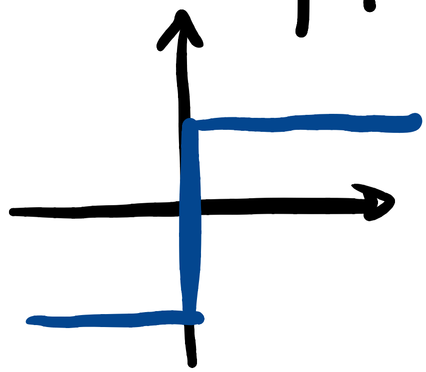
$$\partial f(x) = \{ v \mid \forall y \in \mathbb{R}^d \; f(y) \geq f(x) + \langle v, y-x \rangle \}$$

Examples

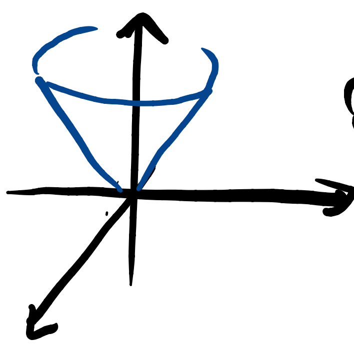
1. $f(x) = |x|$

$$\partial f(x) = \begin{cases} 1 & x \geq 0 \\ [-1,1] & x = 0 \\ -1 & x < 0 \end{cases}$$





2. $f(x) = \|x\|$

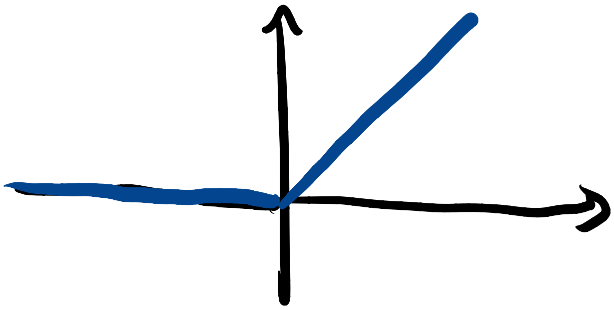


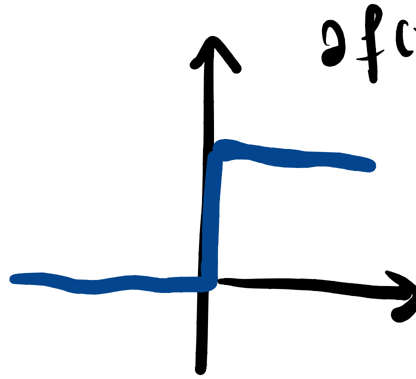
$$\partial f(x) = \begin{cases} \dfrac{x}{\|x\|} \\ \{ y \mid \|y\| \leq 1 \} \end{cases}$$

3. $f(x) = \max\{0, x\}$

ReLU

$$\partial f(x) = \begin{cases} 1 & x > 0 \\ [0,1] & x = 0 \\ 0 & x < 0 \end{cases}$$



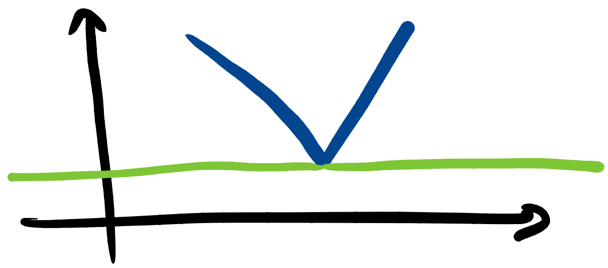


What do we just gained?

Theorem: Optimality cond for general convex func.

Suppose $f:\mathbb{R}^d \to \mathbb{R}$ is convex. Then x^* is a minimizer iff $0 \in \partial f(x^*)$.

Intuition



Nothing goes under.

Proof: Assume x^* is a minimizer.

$$\Updownarrow$$

$$f(x^*) + \langle 0, y - x \rangle \leq f(y) \quad \forall y.$$

$$\Updownarrow$$

Assume that $0 \in \partial f(x)$. □

Proposition: Subdifferential calculus

Suppose that $f, h: \mathbb{R}^d \to \mathbb{R}$ are convex functions. Then the following holds

1 (Sums) $\partial(f+h)(x) = \partial f(x) + \partial h(x)$.