Lecture 4

Sept/05/2024

Last time

- 2nd-order optimality cond.
- Basic convexity

Agenda

- Characterization smooth convex.
- Subgradients.

Lemma (First-order characterization of convexity)

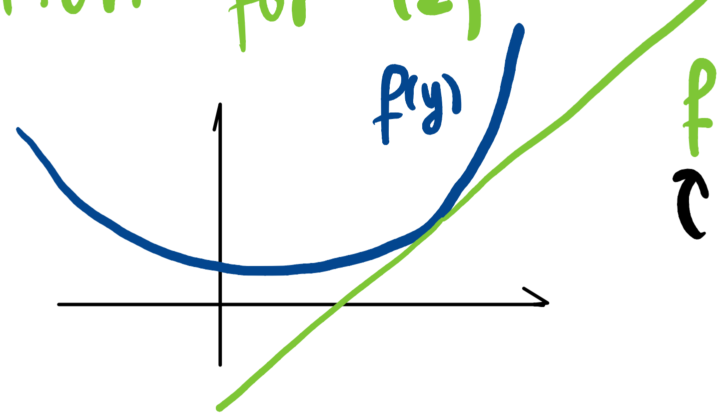
Suppose that $f: \mathbb{R}^d \to \mathbb{R}$ is differentiable. Then, the following are equivalent:

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

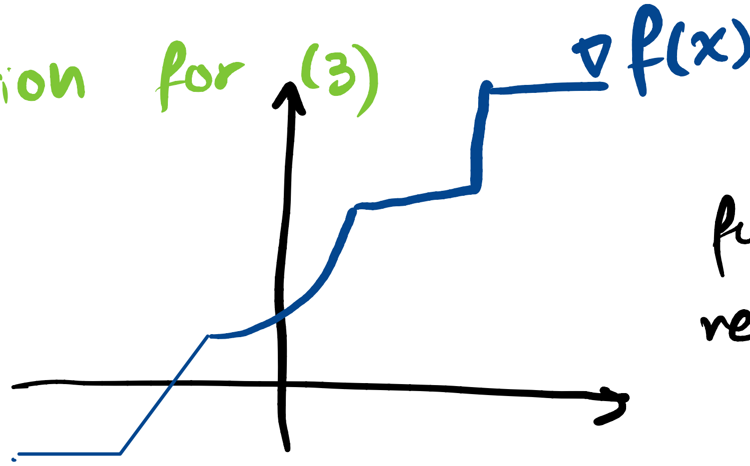
Intuition for (2)



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

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

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

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

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

$\Rightarrow \quad (1-t)$ 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)

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

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

Then
$$
\begin{aligned}
f(y) &= \varphi(1) = \varphi(0) + \int_0^1 \varphi'(t)\,dt \\
&= \varphi(0) + \varphi'(0) + \int_0^1 [\varphi'(t) - \varphi'(0)]\,dt \\
&= f(x) + \nabla f(x)(y-x) \\
&= f(x) + \nabla f(x)(y-x) + \overbrace{\int_0^1 \nabla f'(x + t(y-x))^T(y-x)\,t\,dt}^{\geq 0} \\
&\geq f(x) + \nabla f(x)(y-x)
\end{aligned}
$$

$\square$

Lemma 2nd-order characterization

Assume f twice differentiable. Then,

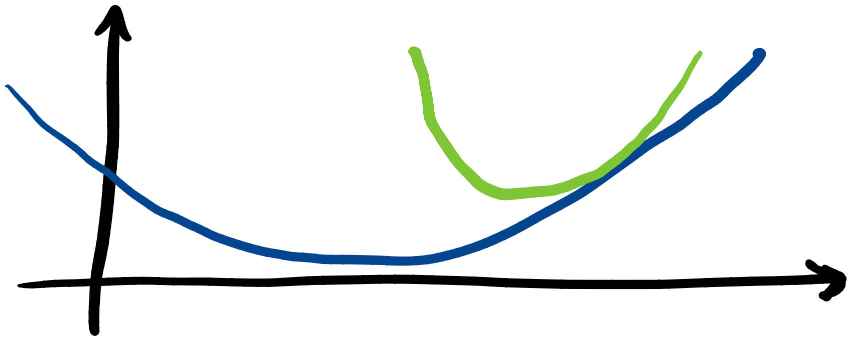
f is convex $\Leftrightarrow$ $\nabla^2 f(x) \succeq 0 \quad \forall x.$

$s^T \nabla^2 f(x) s \geq 0 \ \forall s,$

$\lambda_{min}(\nabla^2 f(x)) \geq 0.$

Intuition

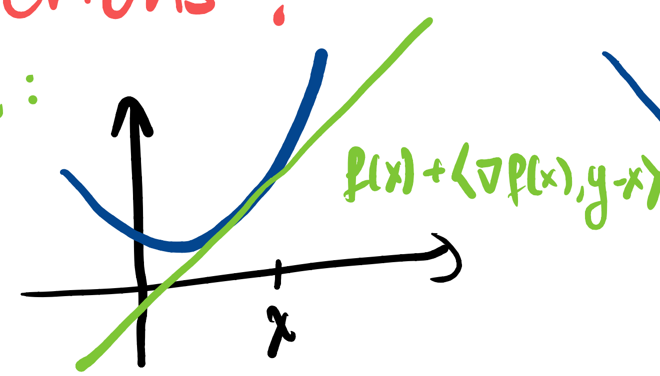
Second order model never curves down!

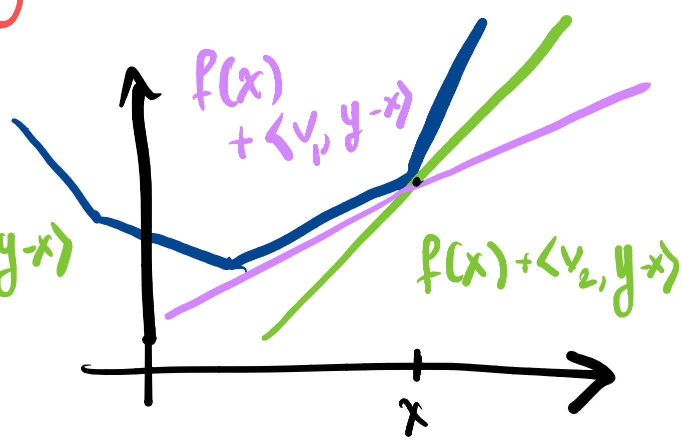


Proof Exercise. $\square$

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

Idea:





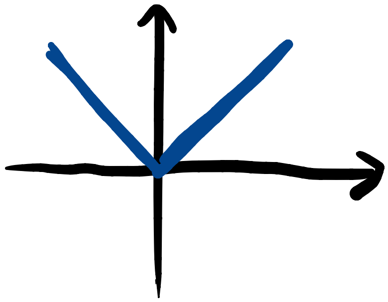
Def: Consider a convex function $f: \mathbb{R}^d \to \mathbb{R}$. The subdifferential of f at x is

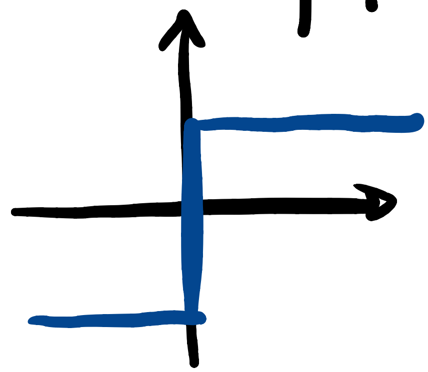
$$\partial f(x) = \{ v \mid \forall y \in \mathbb{R}^d \ f(y) \geq f(x) + \langle v, y-x \rangle \}$$

Examples

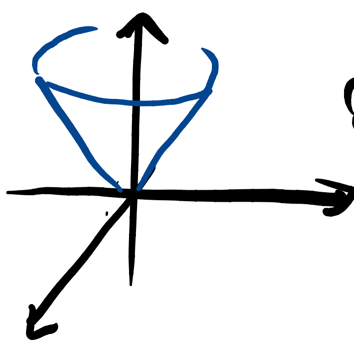
1. $f(x) = |x|$

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





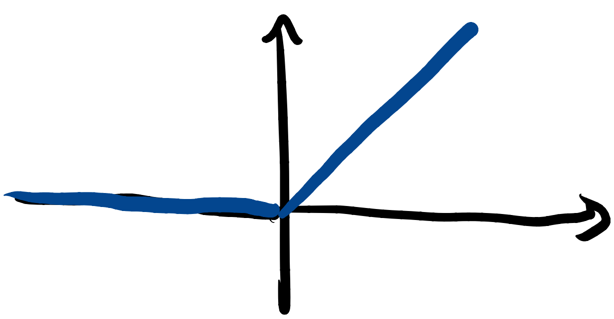
2. $f(x) = \|x\|$

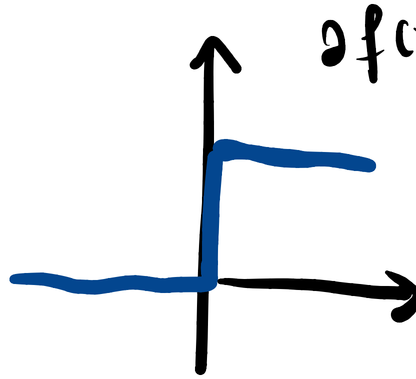


$$\partial f(x) = \begin{cases} \dfrac{x}{\|x\|} \\ \{ y \mid \|y\| \leq 1 \} \end{cases}$$

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

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



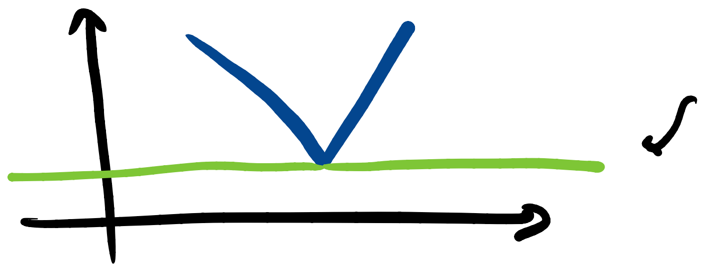


What do we just gained?

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



Proof: Assume x^* is a minimizer.

$$\Updownarrow$$

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

$$\Updownarrow$$

Assume that $0 \in \partial f(x)$. $\square$

Proposition: Subdifferential calculus

Suppose that $f_1, f_2: \mathbb{R}^d \to \mathbb{R}$ are convex functions. Then the following holds

1. (Sums) $\partial (f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x)$.
2. (Chain rule) If $A: \mathbb{R}^n \to \mathbb{R}^d$ linear

$$\partial (f_1 \circ A)(x) = A' \, \partial f_1(Ax).$$

3. (Scalings)

$$\partial (\alpha f_1)(x) = \alpha \, \partial f_1(x).$$

4. (Max) For all x, define $M(x) = \{ i \mid f_i(x) = \max\{f_1(x), f_2(x$

$$\partial \max\{f_1, f_2\}(x) = \mathrm{conv}\{ g \in \partial f_i \mid i \in M(x) \}.$$

convex hull

5. (Smooth functions) Assume that f_i is diff at x.

$$\partial f_i(x) = \{ \nabla f(x) \}.$$

This one you should prove

What did we gain?

A way to compute subdiff. for complicated functions!