Convex Optimization : Lecture 3

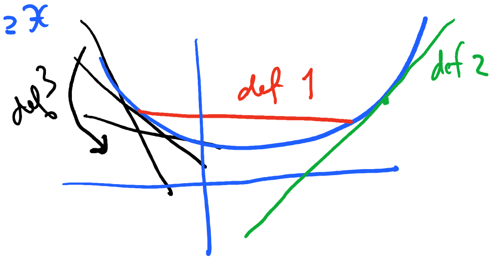
Wednesday, August 31, 2016 1:27 PM

Last time: Defined convex set in two ways

1. $\mathcal{X}$ convex if $\mathcal{X}$ contains all linesegments
   i.e., $x_1, x_2 \in \mathcal{X} \Rightarrow \lambda x_1 + (1-\lambda) x_2 \in \mathcal{X}$ $\forall \lambda \in [0,1]$

2. $\mathcal{X}$ convex, closed if
   $$\mathcal{X} = \bigcap_{H \supseteq \mathcal{X}} H .$$

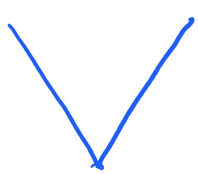
Def: $f : \mathbb{R}^n \to \mathbb{R}$ convex

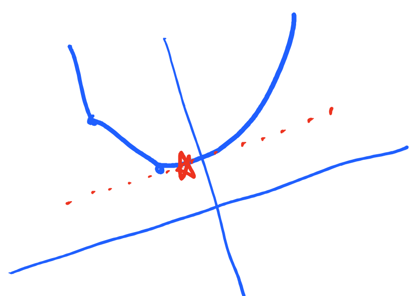


def 1: no assumptions on f
2: diff'ble
3: 2x diff'ble

$\boxed{\text{Thm: min: } f(x) \text{ if } f \text{ is convex then } \hat{x} \text{ is an optimal sol'n iff } \nabla f(\hat{x}) = 0 \text{ (requires } \nabla f \text{ to exist).}}$

This class: ① Extend this idea to non-diff'ble fns, ex:



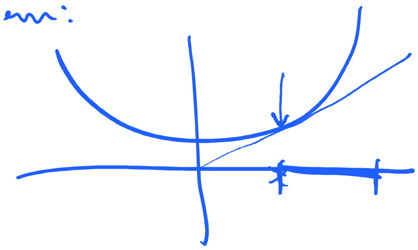


~~gradients~~ sub-differentials
sub-gradients

② Optimality conditions for the general convex constrained problem:

convex constrained problem:

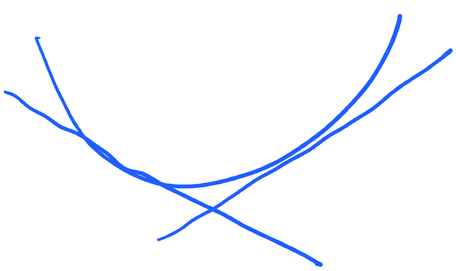
$$\min: f(x)$$
$$\text{st. } x \in \mathcal{X}$$

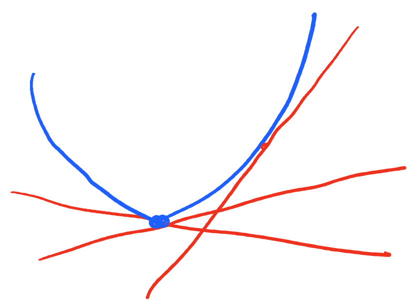


<u>Caution</u>: 2-dim'l pictures are very simple concepts in fact are deep & useful.

Sub-gradients and sub-differentials of f, a convex fn, at a point x.

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



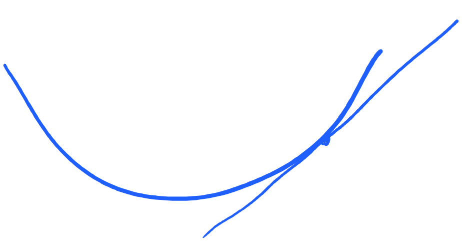


Re-define gradient as follows:

If $f: \mathbb{R}^n \to \mathbb{R}$ is a diff'ble convex fn, $\forall x$

$\exists!$ vector g s.t. $f(y) \geq f(x) + \langle g_x, y - x \rangle \quad \forall y$

<u>Def'n</u>: g_x is called the gradient of f at x.



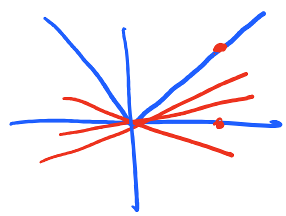
To define the subgradient & sub-differential, we use exactly the same definition, but without the requirement for uniqueness.

<u>Def'n</u>: g_x is called a <u>sub-gradient</u> of f at the point x, if

$$f(y) \geq f(x) + \langle g_x, y-x \rangle \quad \forall y.$$

<u>Def'n</u>: $\partial f(x) = \{ g_x : g_x \text{ is a subgradient of } f \text{ at } x \}$ is called <u>sub-differential</u> of f at x.

<u>Ex</u>: $f(x) = |x|$



$$\partial f(0) = \partial |0|$$
$$= \{ g : -1 \leq g \leq 1 \}$$

Can check directly from the definition.

$$\partial f(x) = \begin{cases} \operatorname{sgn} x & \text{if } x \neq 0 \\ [-1,1] & \text{if } x = 0 \end{cases}$$

<u>Thm</u>: $\min: f(x)$
$\quad$ st: $x \in \mathbb{R}^n$.

$\hat{x}$ is optimal iff $0 \in \partial f(\hat{x})$

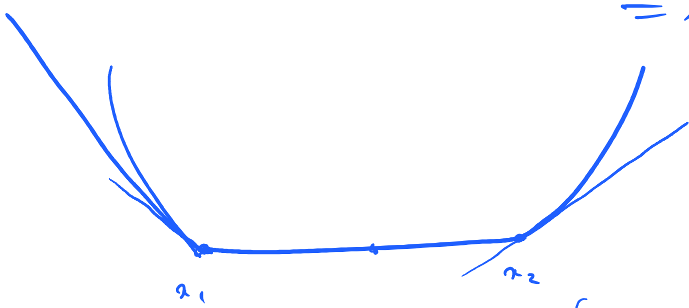
prove one direction "$\Leftarrow$"

By def'n: $f(y) \geq f(\hat{x}) + \langle g, y - \hat{x} \rangle \ \forall y$

true $\forall g \in \partial f(\hat{x})$.

If $0 \in \partial f(\hat{x}) \Rightarrow f(y) \geq f(\hat{x}) + 0 \ \forall y$

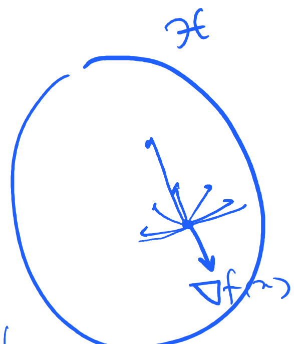
$\Rightarrow \hat{x}$ is optimal.



$\partial f(x_1) = [-\alpha, 0]$ $\qquad$ $\partial f(x_2) = [0, \beta]$

<u>Constraints</u>: $\quad \min: f(x)$
$\quad$ st: $x \in \mathcal{X}$

For starters - consider only a function f where $\nabla f(x)$ defined everywhere.



[$\hat{x} \in \mathcal{X}$ is an optimal solution if we cannot improve it (locally).]

All feasible directions of motion, i.e., all directions that (infinitessimally) keep us in the feasible set, do not improve the solution.

"All directions that improve the solution from x"

$\{v : f(x + \varepsilon v) < f(x)$ for ε small enough$\}$

$= \{v : \langle \nabla f(x), v \rangle < 0\}$

Intuitive explanation: $f(x+\varepsilon v) = f(x) + \langle \nabla f(x), \varepsilon v \rangle + O(\varepsilon^2)$

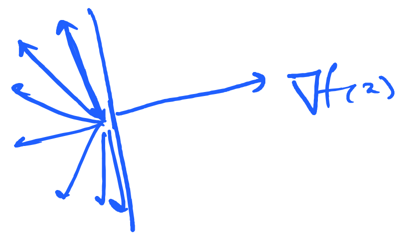
$= f(x) + \varepsilon \underline{\langle \nabla f(x), v \rangle}$ $\cancel{+ O(\varepsilon^2)}$

ignore.

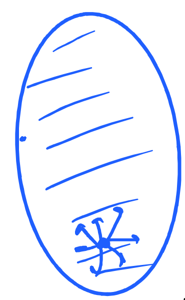
For a step in direction v to be a good idea, i.e., $f(x+\varepsilon v) < f(x)$

we need: $\langle \nabla f(x), v \rangle < 0$.

Picture:



All feasible directions



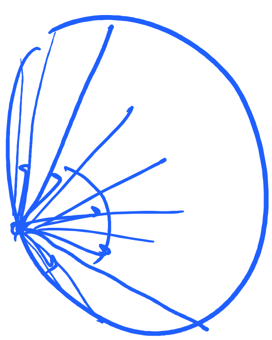
$\subseteq \mathbb{R}^2$

If $x \in \operatorname{int}(\mathcal{X})$ then

$\{\text{all feasible directions}\} = \mathbb{R}^2$.

Def'n: $C \subseteq \mathbb{R}^n$

We call C open if $\forall x \in C$, $\exists \varepsilon > 0$ s.t. $B_\varepsilon(x) \subseteq C$.



V_ε = all directions v s.t. $x + \varepsilon v \in \mathcal{X}$

$$V = \bigcup_{\varepsilon > 0} V_\varepsilon$$

$$= \{ v \text{ s.t. } \exists \varepsilon > 0 \text{ with } x + \varepsilon v \in \mathcal{X} \}$$

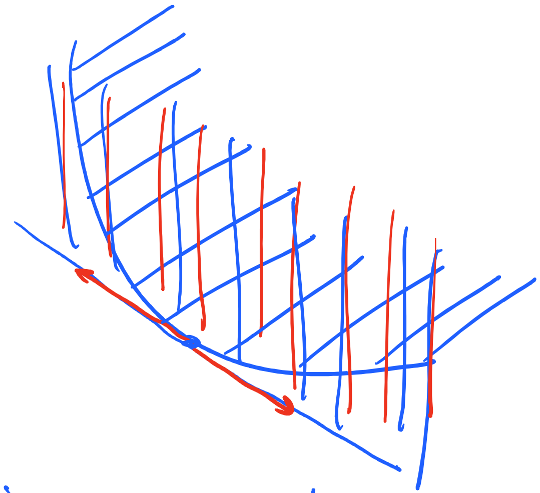
Def: For $C \subseteq \mathbb{R}^n$, a point $x \in C$ is called an interior pt if $\exists \varepsilon > 0$ s.t. $B_\varepsilon(x) \subseteq C$

Def'n: $C \subseteq \mathbb{R}^n$, $\underline{\operatorname{int} C}$ (aka C°) is the collection of interior points of C.

Def'n: $C \subseteq \mathbb{R}^n$, C is closed iff C^c is open.

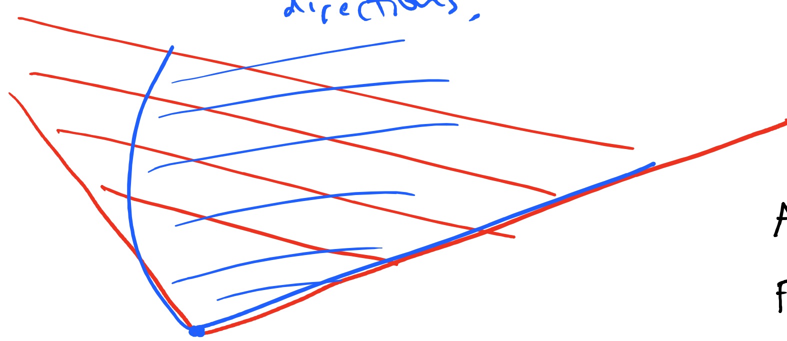
Def'n: $C \in \mathbb{R}^n$, C is [illegible] if $\forall \{x_n\}$ s.t.

$= \{ v \text{ s.t. } \ldots \quad x + \varepsilon v \in \mathcal{X} \}$



Def'n: Tangent cone of a set $\mathcal{X}$ at x

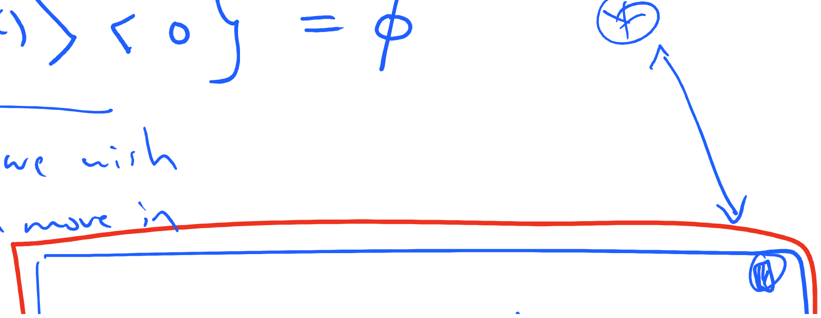
$T_{\mathcal{X}}(x) =$ closure of the set of all feasible directions.



Back to defining optimality:

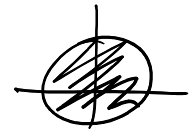
A point $\hat{x}$ is optimal if

$$\underbrace{T_{\mathcal{X}}(\hat{x})}_{\text{directions we can move in}} \cap \{ v : \underbrace{\langle v, \nabla f(\hat{x}) \rangle < 0}_{\text{directions we wish we could move in}} \} = \phi$$



Def'n: C ... closed if $\forall \{x_n\}$ s.t. $x_n \in C$, $x_n \to \bar{x}$, then $\bar{x} \in C$.

Ex: $\{x_1^2 + x_2^2 \le 1\}$ closed.



$\{x_1^2 + x_2^2 < 1\}$ open.

ϕ, $\mathbb{R}^n$ both open and closed.

Check: arbitrary int. of closed sets are closed.

Finite int. of open sets is open

Arb. union of open sets is open

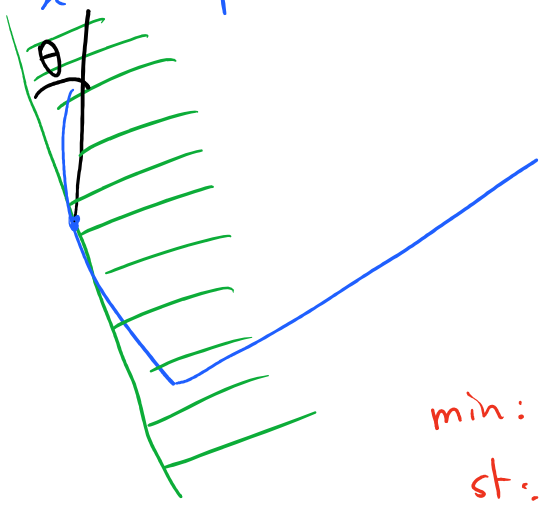
Finite union of closed sets is closed.

$$\bigcup_{n=1}^{\infty} \left[\tfrac{1}{n}, 1\right] = (0, 1]$$

we can move in

we could move in

$\hat{x}$ is optimal if $\boxed{\langle v, \nabla f(\hat{x}) \rangle \geq 0 \quad \forall v \in T_{\mathcal{X}}(\hat{x})}$



Aside: Last time we saw

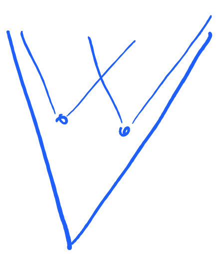
$$\min: \|X\beta - y\|_2^2$$
$$\text{st}: \beta \text{ has at most } k \text{ non-zeros.}$$

$$\min: \|X\beta - y\|_2^2$$
$$\text{st}: \beta \in \mathcal{X}_{NC}$$

$$\min: \|X\beta - y\|_2^2$$
$$\text{st}: \beta \in \mathcal{X}_C \supseteq \mathcal{X}_{NC}$$

Conditions for optimality.

Def'n: A set C is called a convex cone if $\forall x_1, x_2 \in C$, $\lambda_1 x_1 + \lambda_2 x_2 \in C \quad \forall \lambda_1, \lambda_2 \geq 0$



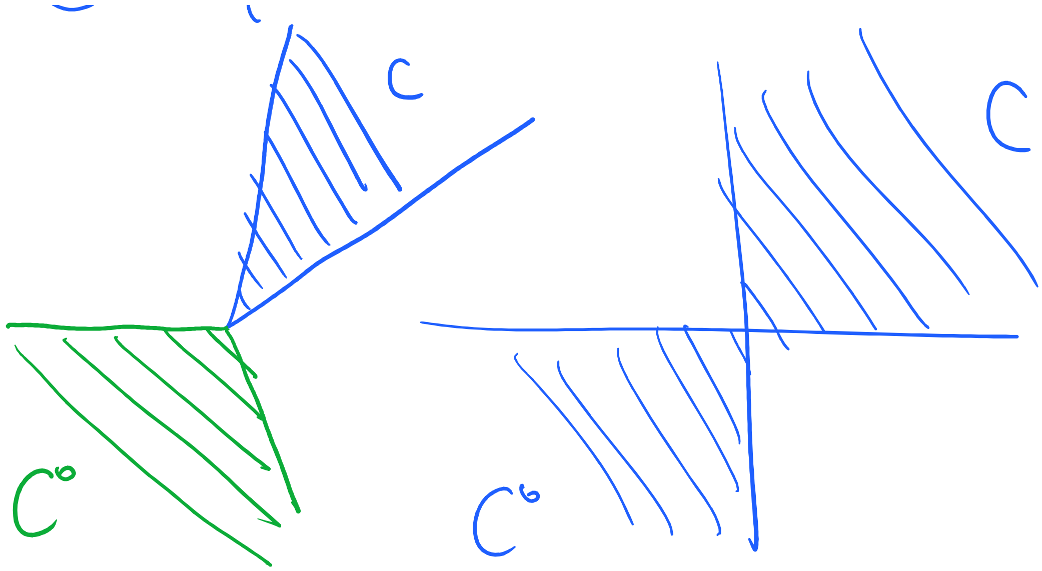
Ex: $S_n^+ =$ $n \times n$ symmetric matrices with ≥ 0 e-values

S_n^+ is a convex cone.

Def'n: If C is a convex cone, its polar cone is defined as:

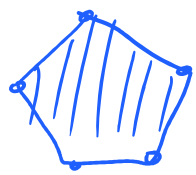
$$C^\circ = \{ v : \langle v, w \rangle \leq 0 \quad \forall w \in C \}$$

Picture:



Def'n: $v_1, \dots, v_k$, Convex Hull $(v_1, \dots, v_k)$ = smallest convex set containing $v_1, \dots, v_k$

$$= \{ v = \lambda_1 v_1 + \dots + \lambda_k v_k, \text{ s.t. } \lambda_i \geq 0, \ \Sigma \lambda_i \leq 1 \}$$

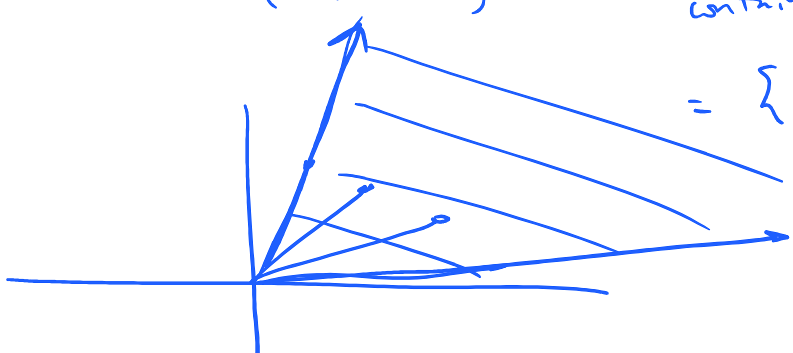


Check!

Def'n: $v_1, v_2, \dots, v_k$

Cone $\{v_1, \dots, v_k\}$ = smallest convex cone that contains $v_1, \dots, v_k$

$$= \{ v = \Sigma \lambda_i v_i \text{ s.t. } ?? \}$$



Ex: Let $v_1, v_2 \neq 0$, $v_1 \perp v_2$

check

$S = \text{cone}\{v_1 v_1^T, v_2 v_2^T\} \subseteq S_n$ = $n \times n$ symmetric matrices

Find S°

$\min: f(x)$
$\text{st}: x \in \mathcal{X}$

$\hat{x}$ is an optimal solution if $\langle v, \nabla f(\hat{x}) \rangle \geq 0 \quad \forall v \in T_{\mathcal{X}}(\hat{x})$

$$\iff \langle v, \nabla f(\hat{x}) \rangle \leq 0 \quad \forall v \in -T_{\mathcal{X}}(\hat{x})$$

$$\iff \nabla f(\hat{x}) \in \left(-T_{\mathcal{X}}(\hat{x})\right)^\circ$$

Defn: $\left(T_{\mathcal{X}}(\hat{x})\right)^\circ = N_{\mathcal{X}}(\hat{x})$

$\hat{x}$ is optimal iff $\nabla f(\hat{x}) \in -N_{\mathcal{X}}(\hat{x})$

$$0 \in \nabla f(\hat{x}) + N_{\mathcal{X}}(\hat{x})$$

$\min: f(x)$ $\quad$ f convex
$\text{st}: x \in \mathcal{X}$ $\quad$ $\mathcal{X}$ convex

Thm: $\hat{x}$ is an optimal solution iff

$$0 \in \partial f(\hat{x}) + N_{\mathcal{X}}(\hat{x})$$

KKT conditions

Q: If $\mathcal{X} = \mathbb{R}^n$, $N_{\mathcal{X}}(\hat{x})$

Q: If $\mathcal{X} = \mathbb{R}^n$, $N_{\mathcal{X}}(\hat{x})$

$$T_{\mathcal{X}}(\hat{x}) = \mathbb{R}^n$$

$$N_{\mathcal{X}}(\hat{x}) = \left(T_{\mathcal{X}}(\hat{x})\right)^{\circ} = \left(\mathbb{R}^n\right)^{\circ}$$

$$= \{0\}$$