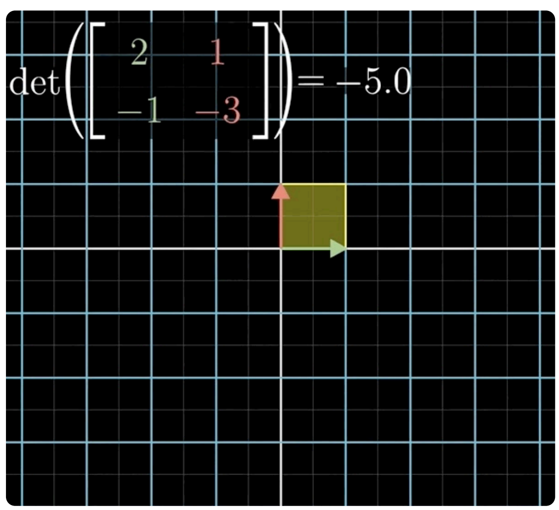
Waitlist and section changes : will be announced later this week
HW1 due Friday

Outline
Recap PSD matrices
Unconstrained optimization
Convexity
Constrained optimization

PSD matrices

$\underline{\underline{A}}$ is PSD iff $x^T A x \geq 0 \quad \forall x \in \mathbb{R}^d$

iff all eigenvalues are nonnegative



$f(x)$: gradient at x $\nabla f(x) \in \mathbb{R}^d$

Hessian at x $\nabla^2 f(x) \in \mathbb{R}^{d \times d}$

x^* is a local min if $\exists r > 0$ s.t. $f(x^*) \leq f(x)$

$\forall x$ s.t. $\|x - x^*\| \leq r$

global min $(r \to \infty)$ $\quad f(x^*) \le f(x) \;\; \forall x$

<u>Property 1:</u> If $f(x)$ is diff'ble and x^* is local min then $\nabla f(x^*) = 0$ (stationary point)

<u>Proof:</u> We have $f(\underline{x}^*) \le f(\underline{x}^* + t\underline{y}) \;\; \forall y$ and small enough t.

<u>Define</u> $g(t) \triangleq f(\underbrace{\underline{x}^* + t y}_{z})$ (should increase with t)
$(t \ge 0)$

(1) $$g'(0) = \sum_{i=1}^{d} \frac{\partial f(z)}{\partial z_i} \cdot \frac{\partial z_i}{\partial t} \Bigg|_{t=0} = \sum_{i=1}^{d} \frac{\partial f(x)}{\partial x_i} \Bigg|_{x=x^*} \cdot y_i$$

$$= \langle \nabla f(x^*), y \rangle$$

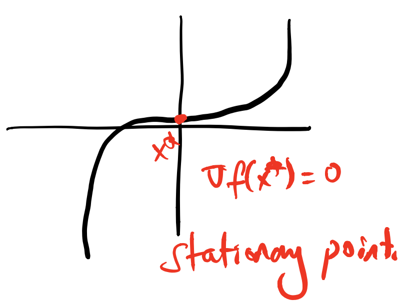
(2) $$g'(0) = \lim_{t \to 0^+} \frac{f(x^* + ty) - f(x^*)}{t} \ge 0 \quad \text{(by local min)}$$

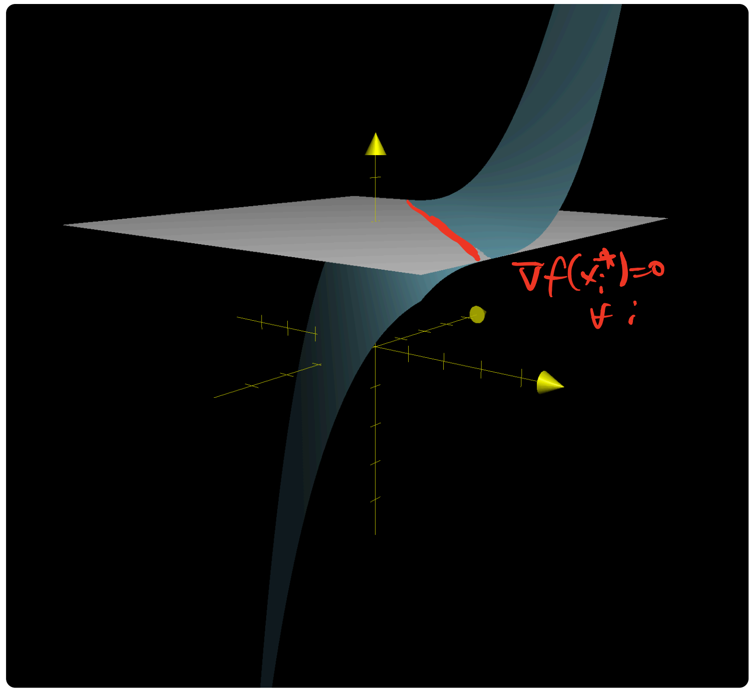
(1) & (2) $\Rightarrow$ $\langle \nabla f(x^*), y \rangle \ge 0 \quad \forall y.$

Pick $y = -\nabla f(x^*)$ $\Rightarrow$ $-\langle \nabla f(x^*), \nabla f(x^*) \rangle \ge 0$

$-\| \nabla f(x^*) \|_2^2 \ge 0$

$\nabla f(x^*) = 0 \quad \square.$





Property 2 If f is twice cont. diff'ble and x^* is a local minimum, then $\nabla^2 f(x^*)$ is positive semidefinite

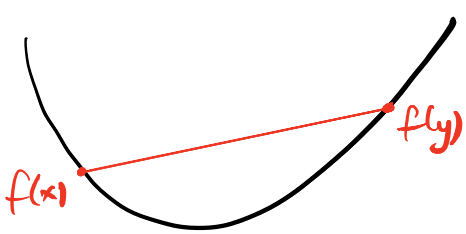
$$z^T \nabla^2 f(x^*) z \geq 0 \quad \forall z \in \mathbb{R}^d.$$

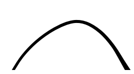
Intuition: By Taylor's exp.

$$f(x) \approx f(0) + \langle \nabla f(0), x \rangle + \frac{1}{2} \underbrace{x^T \nabla^2 f(0) x}_{\geq 0}$$

Convexity: We say f is convex if

$$f(tx + (1-t)y) \leq t f(x) + (1-t) f(y) \quad \forall x, y \in \mathbb{R}^d, \; t \in [0,1]$$



- f is concave if $-f$ convex 

- convex $\Rightarrow$ easy to understand the problem of min. f

<u>Property 3</u> : If f is convex, then every local min. is a global min.

<u>Proof:</u> Suppose x^* is a local min but not global. Then $\exists y^*$ s.t $\boxed{f(y^*) < f(x^*)}$ By convexity

$$\begin{aligned} f(tx^* + (1-t)y^*) &\le t f(x^*) + (1-t) f(y^*) \\ &< t f(x^*) + (1-t) f(x^*) \\ &= f(x^*) \end{aligned}$$

Let $t \to 1$ $\Rightarrow$ contradiction to local min. of x^*

$\therefore$ x^* is globally min.

(Does not mean unique global minimum.)

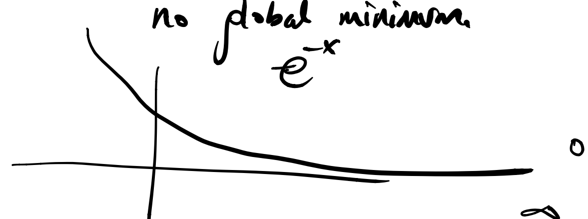
<u>Property 4</u> If f is strictly convex, then f has at most one global min.

Proof: exercise.

<u>Exercise:</u> Give an example $f(x)$ s.t. (1) Convex and more than one global opt.

(2) Strictly convex and no global minimum.

$$\inf_x e^{-x} = 0$$



Second order characterization of convexity

Let f be a twice cont. diff'ble function. Then

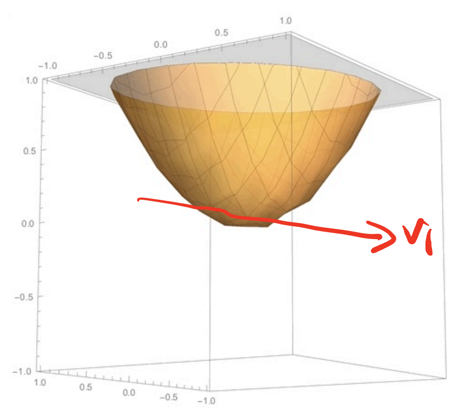
(a) f is convex $\Longleftrightarrow$ $\nabla^2 f(x)$ is PSD $\forall x \in \mathbb{R}^d$.

(b) f is strictly convex $\Longleftarrow$ $\nabla^2 f(x)$ is PD $\forall x \in \mathbb{R}^d$.

ex. $f(x) = x^4$

$f'(x) = 4x^3$

$f''(x) = 12 \cdot x^2 \Rightarrow \nabla^2 f(x)\big|_{x=0} = 0$



$H = \begin{pmatrix} 3. & 1 \\ 1 & 3. \end{pmatrix}$

{4., 2.}

$$f(x) \approx f(0) + \langle \nabla f(0), x \rangle + x^T \nabla^2 f(0)\, x$$

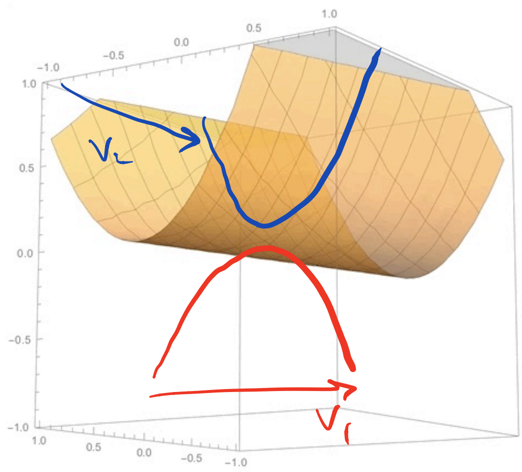
v_1 is an eigvec with eigvalue 0 $\qquad \underbrace{v_1^T \lambda \cdot v_1}_{0}$

$$f(v_1 t) \approx f(0) + \langle \nabla f(0), x \rangle.$$

$$f(v_2 t) \approx f(0) + \langle \nabla f(0), x \rangle + v_2^T \nabla^2 f(0) \cdot v_2 \cdot t^2$$

$v_2^T \lambda_2 v_2 \cdot t^2$

$-t^2$



$$\begin{pmatrix} 0.333333 & 1 \\ 1 & 3. \end{pmatrix}$$

$\{3.33333, 0.\}$

First order characterization of convexity

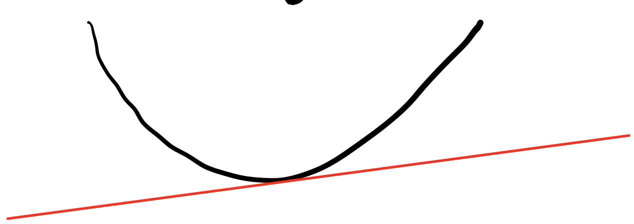
Property 5: Suppose $f: \mathbb{R}^d \to \mathbb{R}$ is diff'ble

Then $f(x)$ is convex iff $\forall x, y \in \mathbb{R}^d$

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

Similarly strictly convex iff $\forall x \neq y$

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



Pf: (first direction) Assume $f(x)$ convex

$$f(ty + (1-t)x) \leq t f(y) + (1-t) f(x)$$
$$= f(x) + t\big(f(y) - f(x)\big)$$

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

t

Let $t \to 0$

$$= \langle \nabla f(x), y-x \rangle \le f(y) - f(x). \qquad \square$$

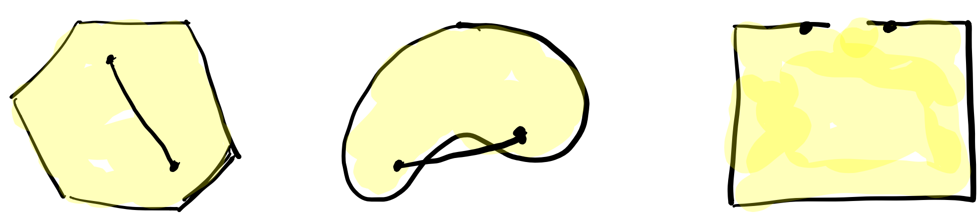
<u>Property 6</u>: f convex, cont. diff'ble then x^* is a global min <u>iff</u> $\nabla f(x^*) = 0$

<u>Pf:</u> Forward implication is property 1.
Reverse implication follows from property 5.

$$f(y) \ge f(x^*) + \langle \underbrace{\nabla f(x^*)}_{0}, y - x \rangle$$

<u>Convex Sets</u>: A set $C \subseteq \mathbb{R}^d$ is convex if for any two points $x, y \in C$, the segment $[x, y]$ lies in C

$$\{ t\underline{x} + (1-t)\underline{y} : 0 \le t \le 1 \} \subseteq C$$

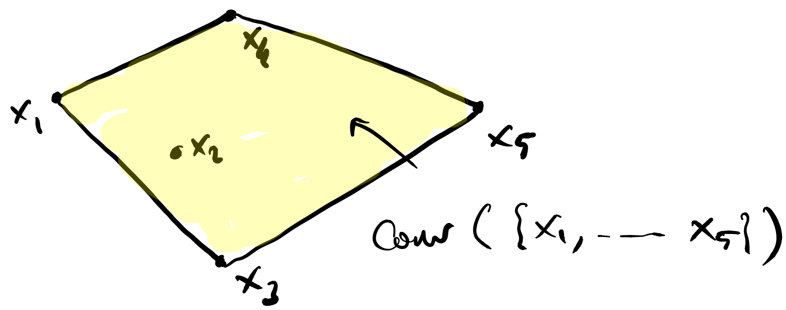


convex   not convex   not convex

<u>Convex hull</u>: of a set of points $\mathcal{X} \subseteq \mathbb{R}^d$ denoted $\mathrm{conv}(\mathcal{X})$ is the set of all convex combinations of points in $\mathcal{X}$

$$\mathrm{Conv}(\mathcal{X}) = \left\{ \sum_{i=1}^{m} \alpha_i \underline{x_i} : x_i \in \mathcal{X}, \ \alpha_i \ge 0, \ \sum_{i=1}^{m} \alpha_i = 1 \right.$$

- $\mathrm{Conv}(\mathcal{X})$ is the smallest convex set that contains $\mathcal{X}$.



Constrained Optimization

$$\min_x \quad f(x)$$

subject to $g_i(x) \le 0 \quad \forall i \{1, -, m\}$

x (opt variable) is feasible if it satisfy all the constraints

$\Sigma x_i = 1$ Equality const. can be written as

$h(x) = 0$, $g_1(x) = h(x) \le 0$

$g_2(x) = -h(x) \le 0$

Lagrangian: associated to const opt problem has the form

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) = f(x) + \lambda^T g(x)$$

where $\lambda_i \ge 0 \quad \forall i$ $\lambda = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix}$