CONVEX FUNCTIONS

Lecture 8 - CMSC764

Tom Goldstein

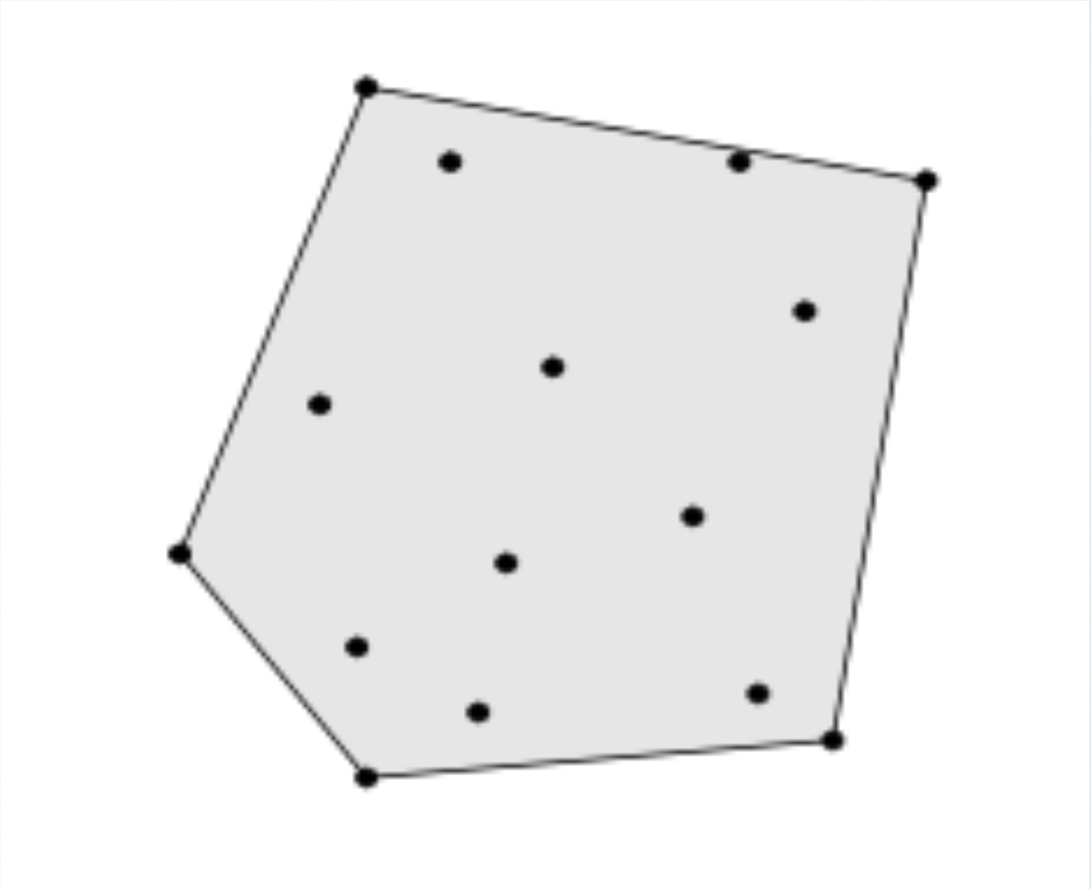
UNIVERSITY OF MARYLAND

CONVEX SETS

Def: A set is convex is every line between two points stays in the set?

$$\theta x_1 + (1-\theta)x_2, \quad 0 \leq \theta \leq 1$$

More General:
All **convex combinations**
lie in set

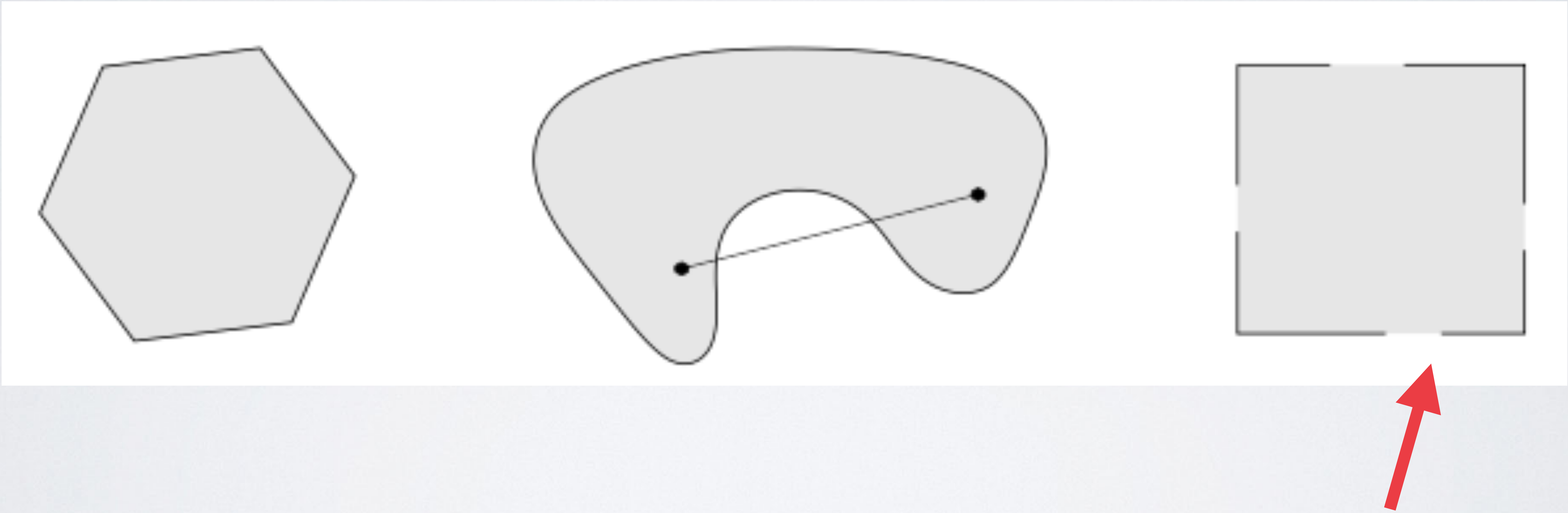


$$\sum \theta_i x_i = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

$$\sum \theta_i = \theta_1 + \theta_2 + \theta_3 = 1, \quad \theta_i \geq 0 \, \forall i$$

These definitions are the same!

IS IT CONVEX?

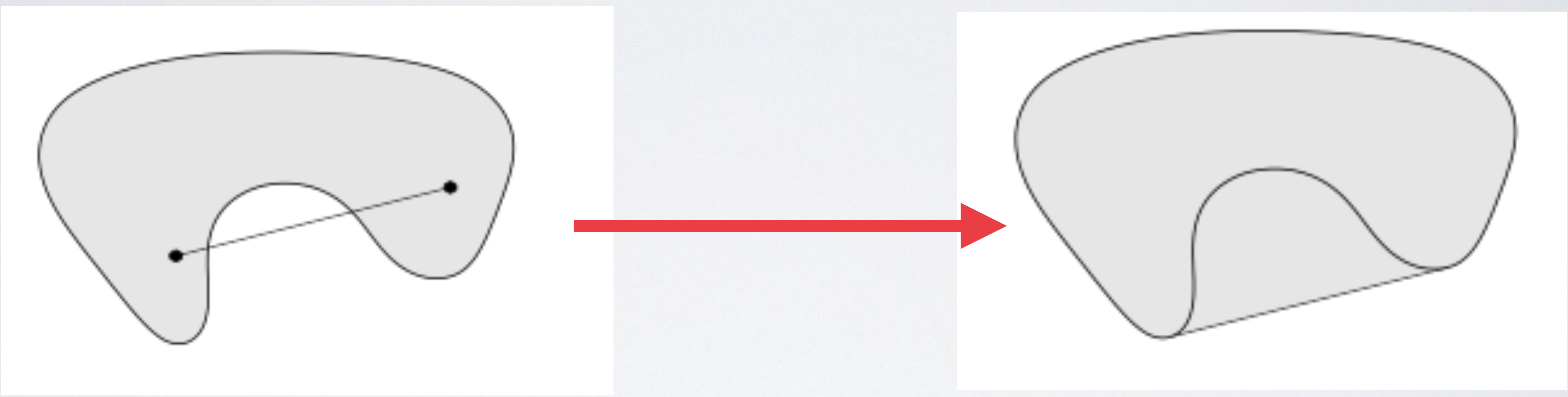


What if the square was round?

figures form Boyd and Vandenberghe

CONVEX HULL

All convex combinations of points an a set



…it’s always the **smallest** convex superset

IMPORTANT EXAMPLES?

Are these convex?

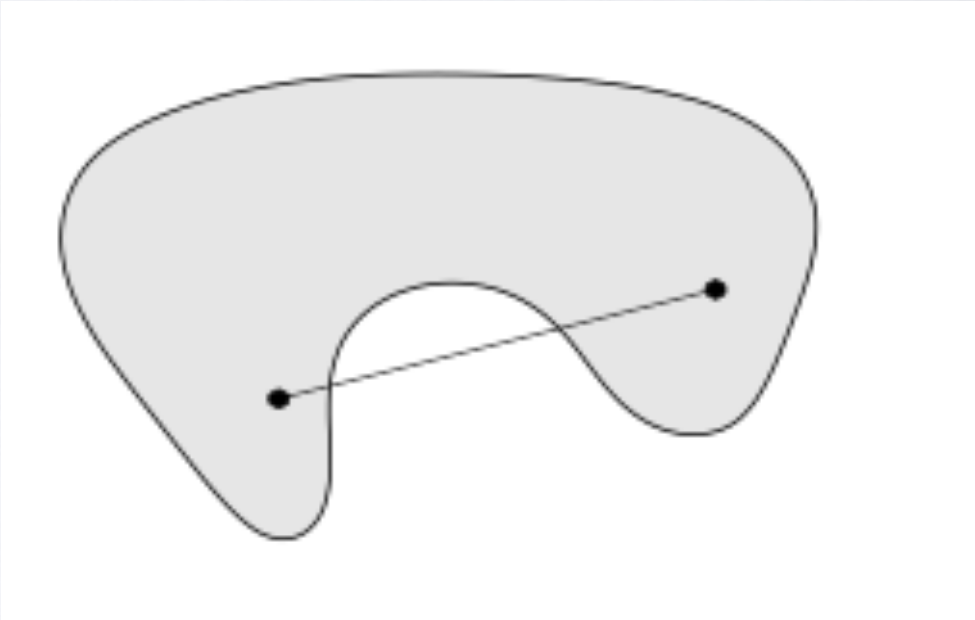
Hyperplane	$\{x \mid a^T x = b\}$
Half-space	$\{x \mid a^T x \geq b\}$
Sphere	$\{x \mid \quad \|x - x_0\| = b\}$
Ball	$\{x \mid \quad \|x - x_0\| \leq b\}$
Polynomials	$\{f \mid f = \sum_i a_i x_i^i\}$

FUNCTIONS OVER NON-CONVEX DOMAIN

Ω

Convex??

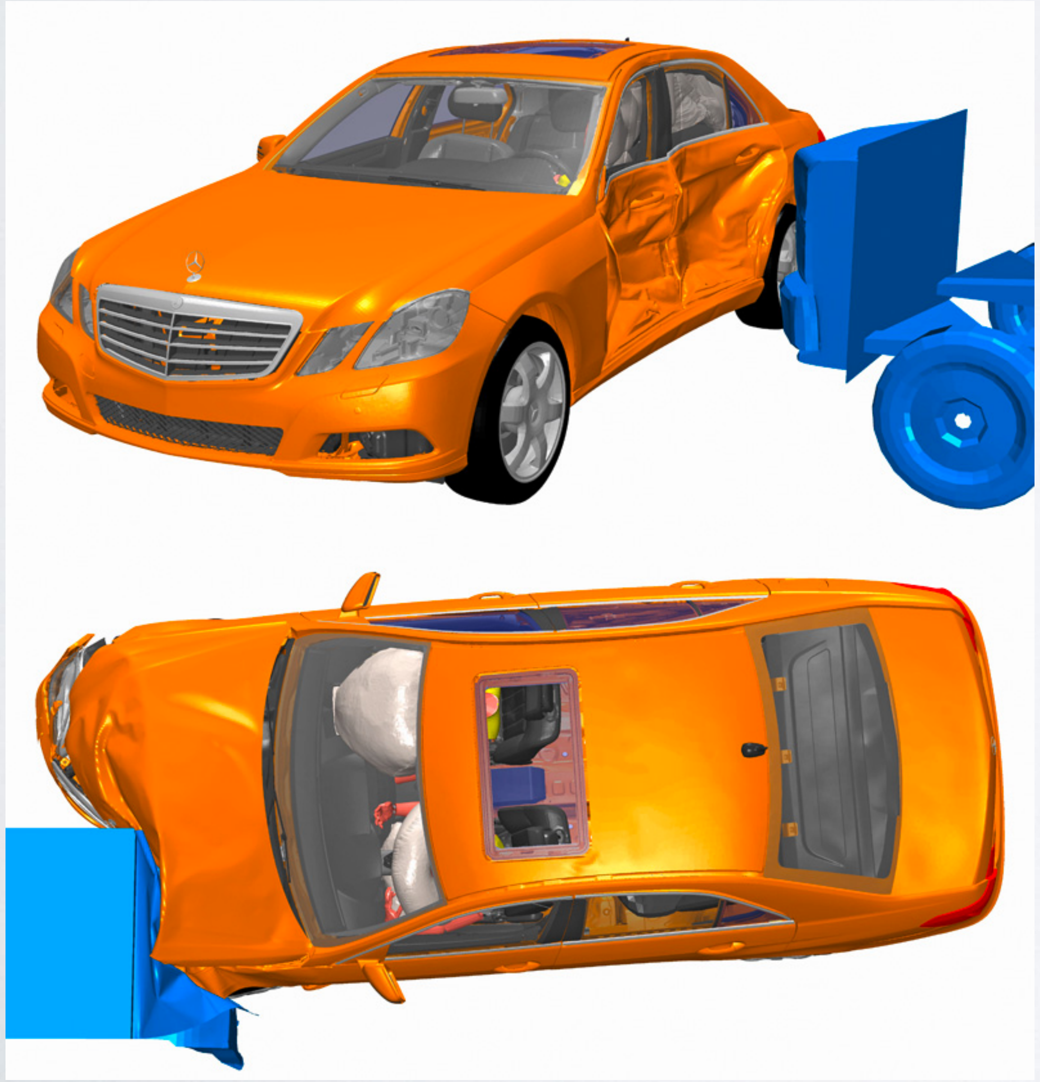
$$f : \Omega \to \mathbb{R}$$

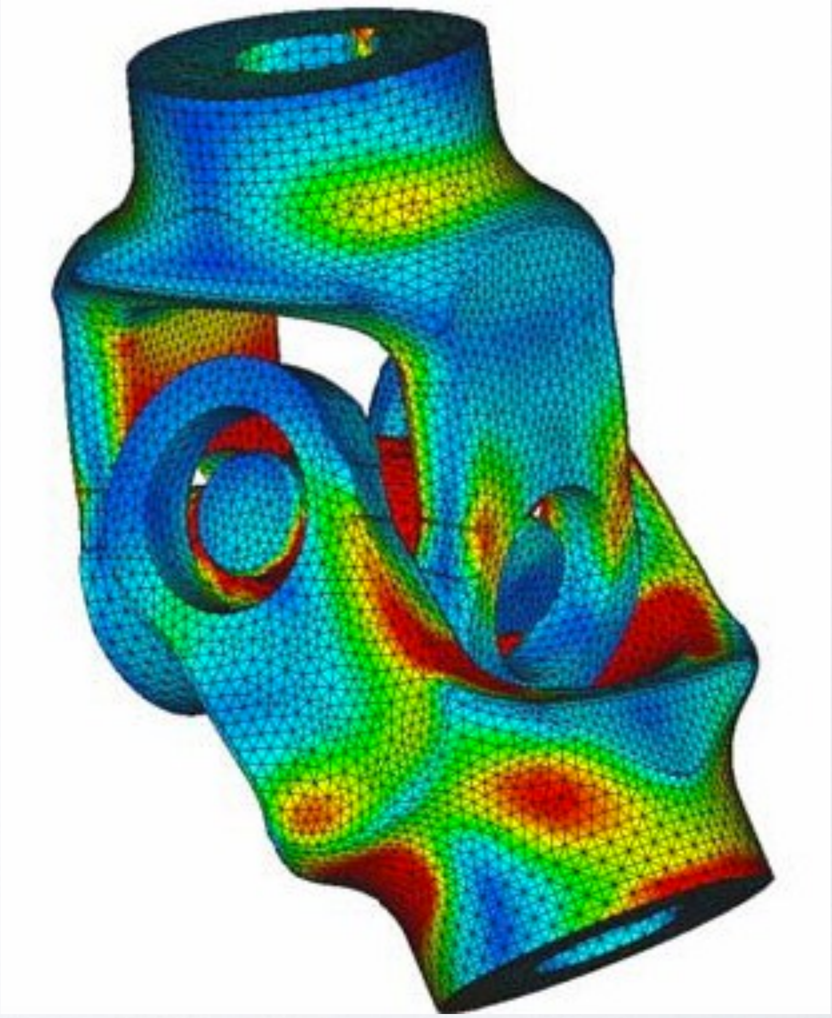


EXAMPLE: LINEAR STRESS IN FINITE ELEMENTS

Functions space over non-convex domain is **convex**

Biharmonic equation $$\min_{u:\Omega \to R} \|\Delta u\|^2 - \langle u, r \rangle$$





UNIT BALL

$$\{x| \quad \|x\| \leq b\}$$

Convex??

Does it depend on which norm??

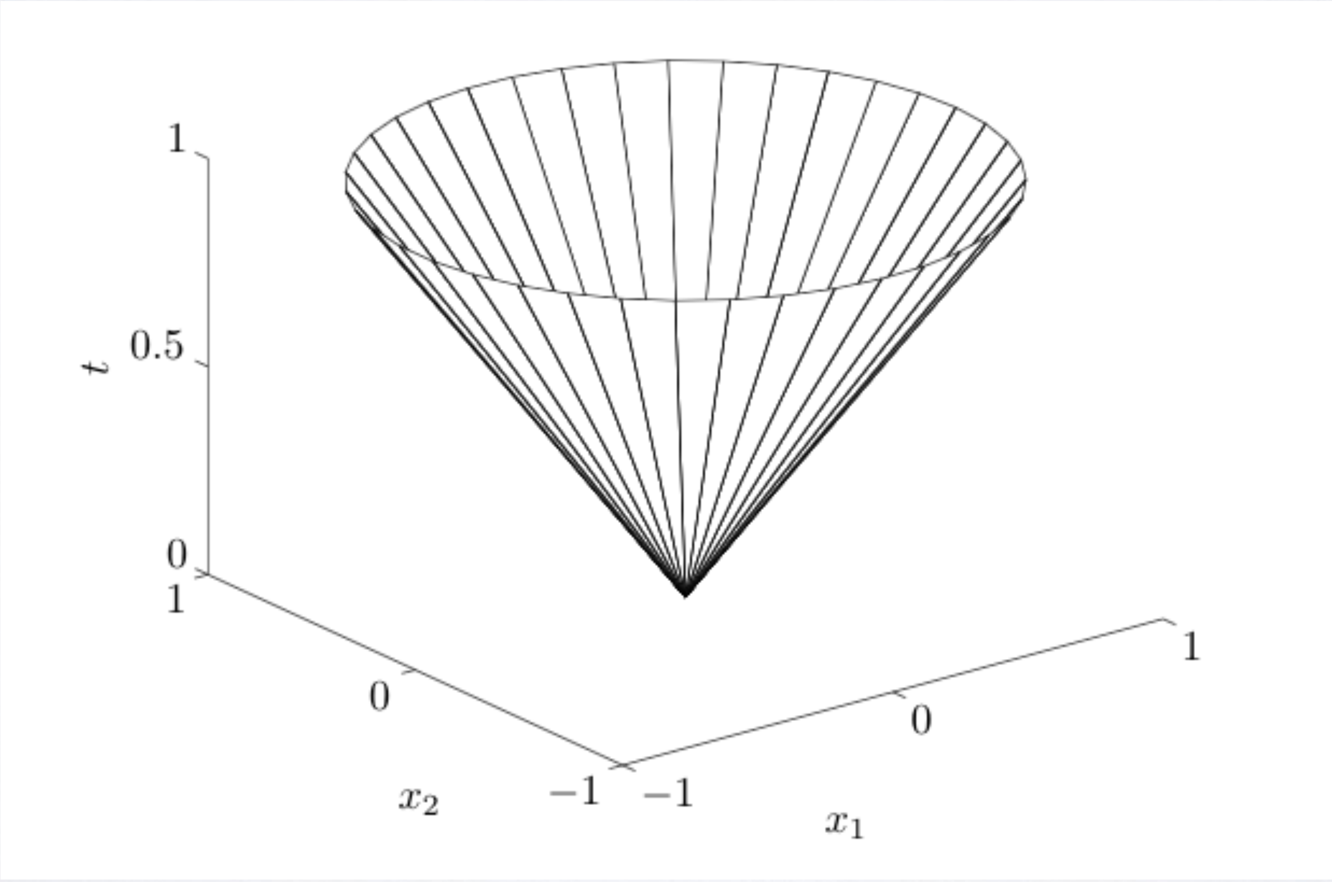
No! Because of triangle inequality.

CONES

$$x \in C \implies ax \in C, \quad \forall a \geq 0$$

Second-order cone:

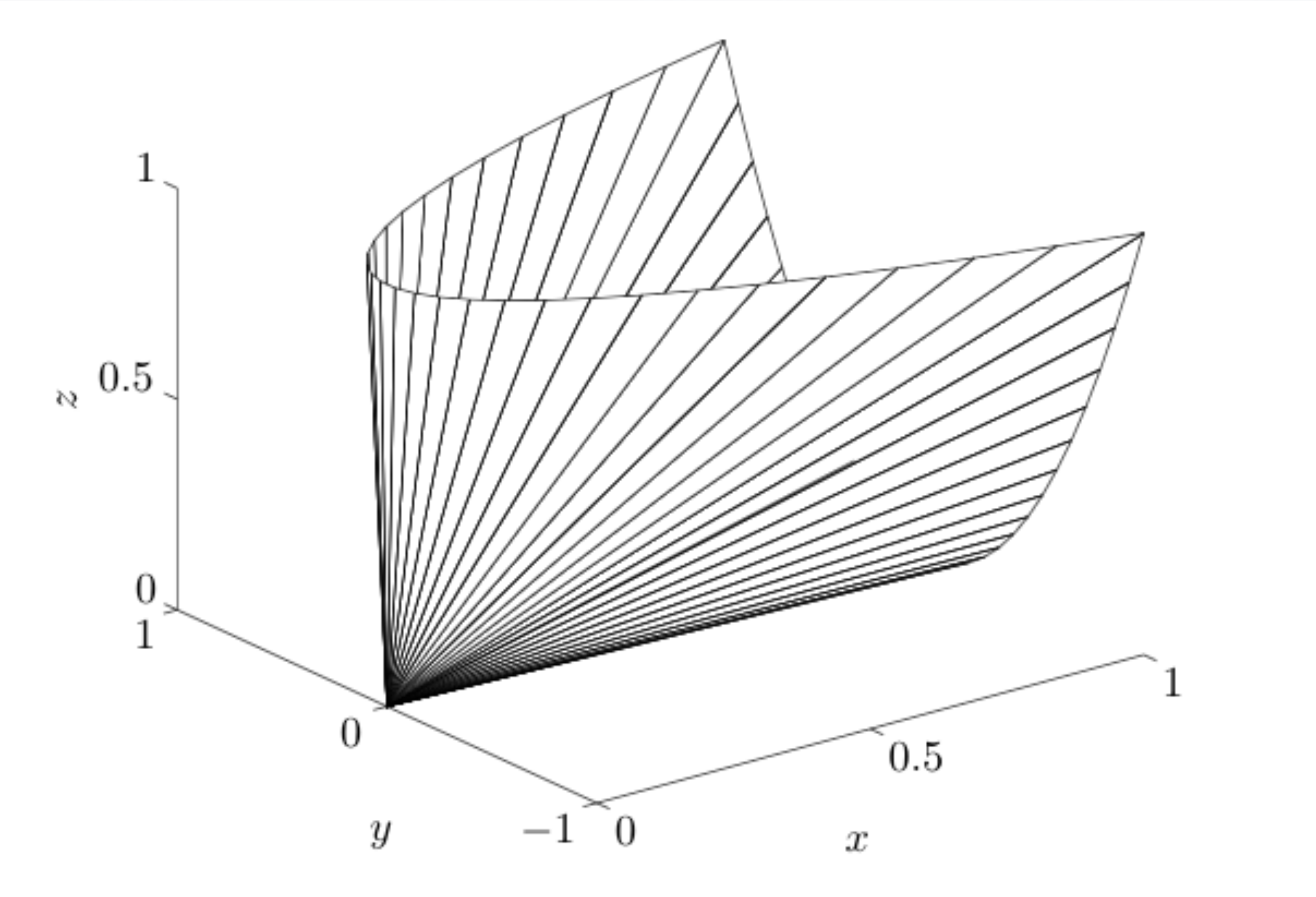
$$C_2 = \{(x,t) | \quad \|x\| \leq t\} \in \mathbb{R}^{n+1}$$



SEMIDEFINITE CONE

$$S^n = \{A \in \mathbb{R}^{n\times n} |\, A = A^T,\, A \succeq 0\}$$

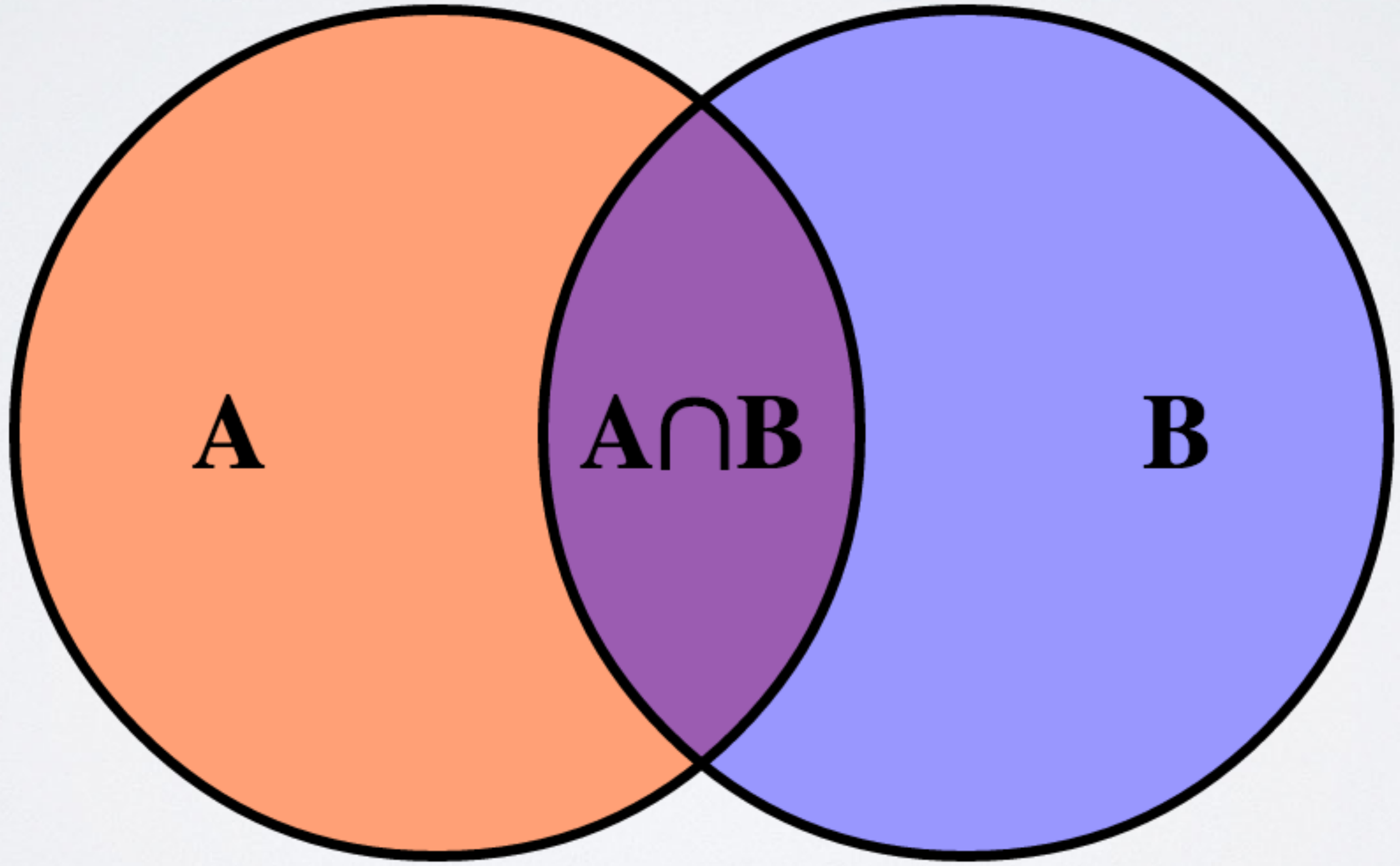
Set of PSD matrices?



Why is this a cone?

ALLOWED OPERATIONS

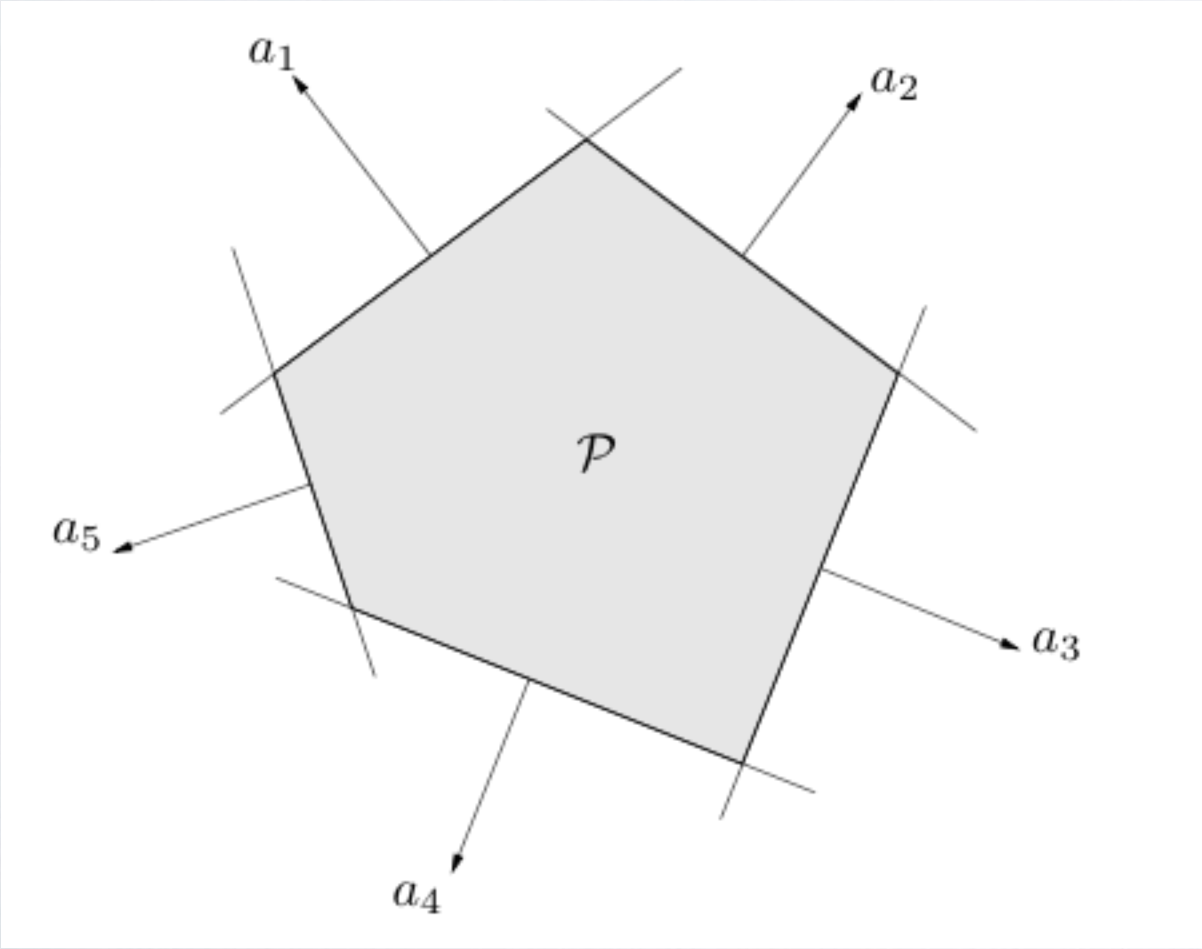
The **intersection** of convex sets is convex

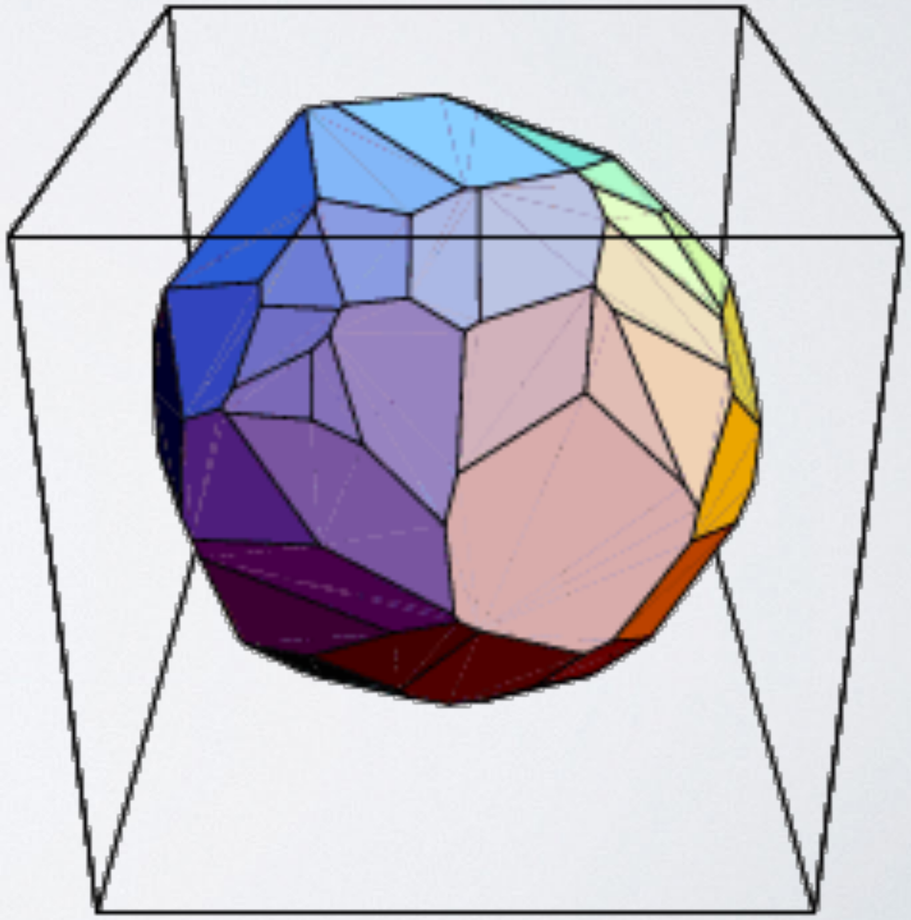


SIMPLEX

$$\{x: \ Ax \le b\}$$

Is it convex? Why?



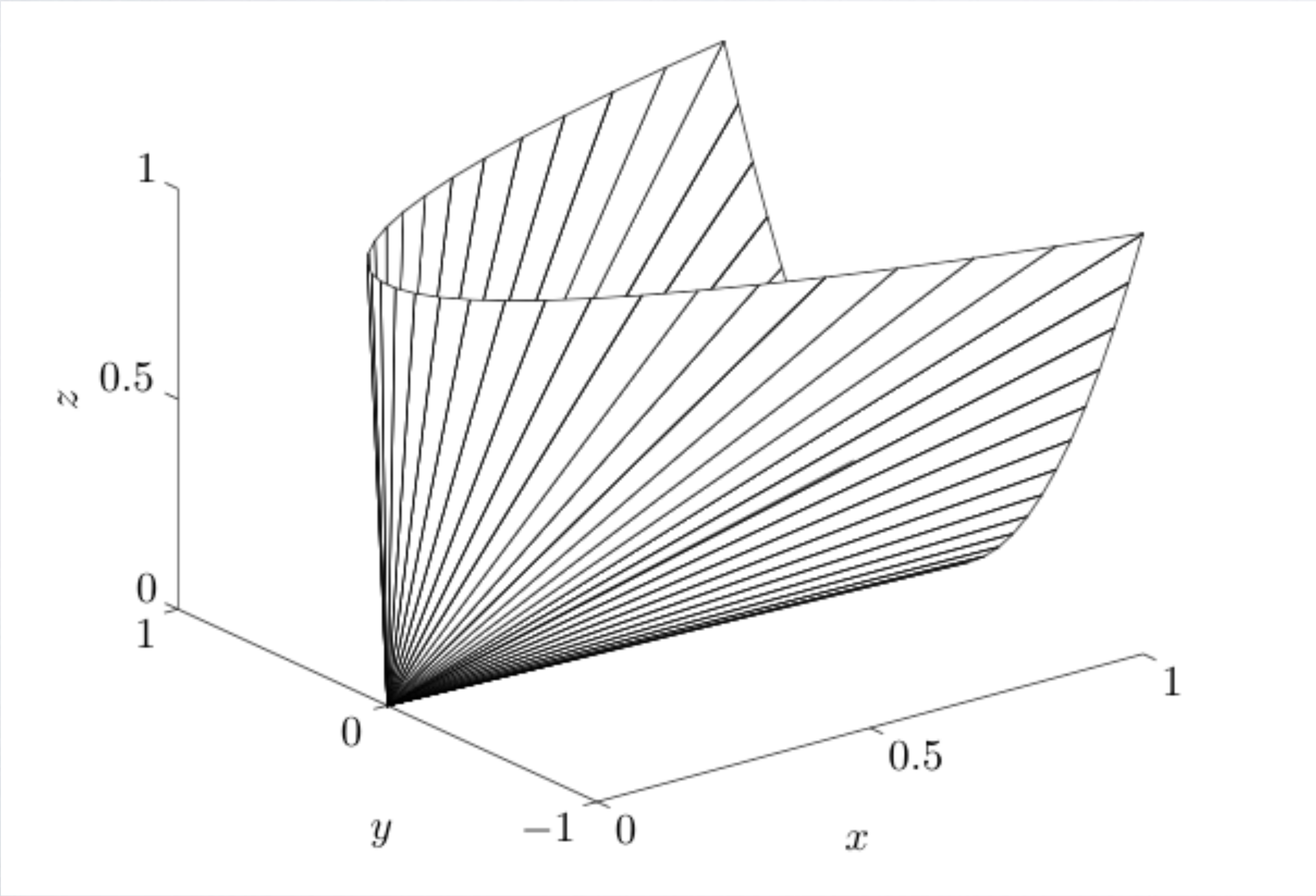


SEMIDEFINITE CONE

$$x^T A x \geq 0, \forall x$$

$$\mathbb{S}_+ = \bigcap_{x \in \mathbb{R}^N} \{A \mid \quad x^T A x \geq 0\}$$

Convex???



OTHER ALLOWED OPERATIONS

Set sum $A + B = \{x + y | x \in A, y \in Y\}$

Set Product $A \times B = \{(x, y) | x \in A, y \in Y\}$

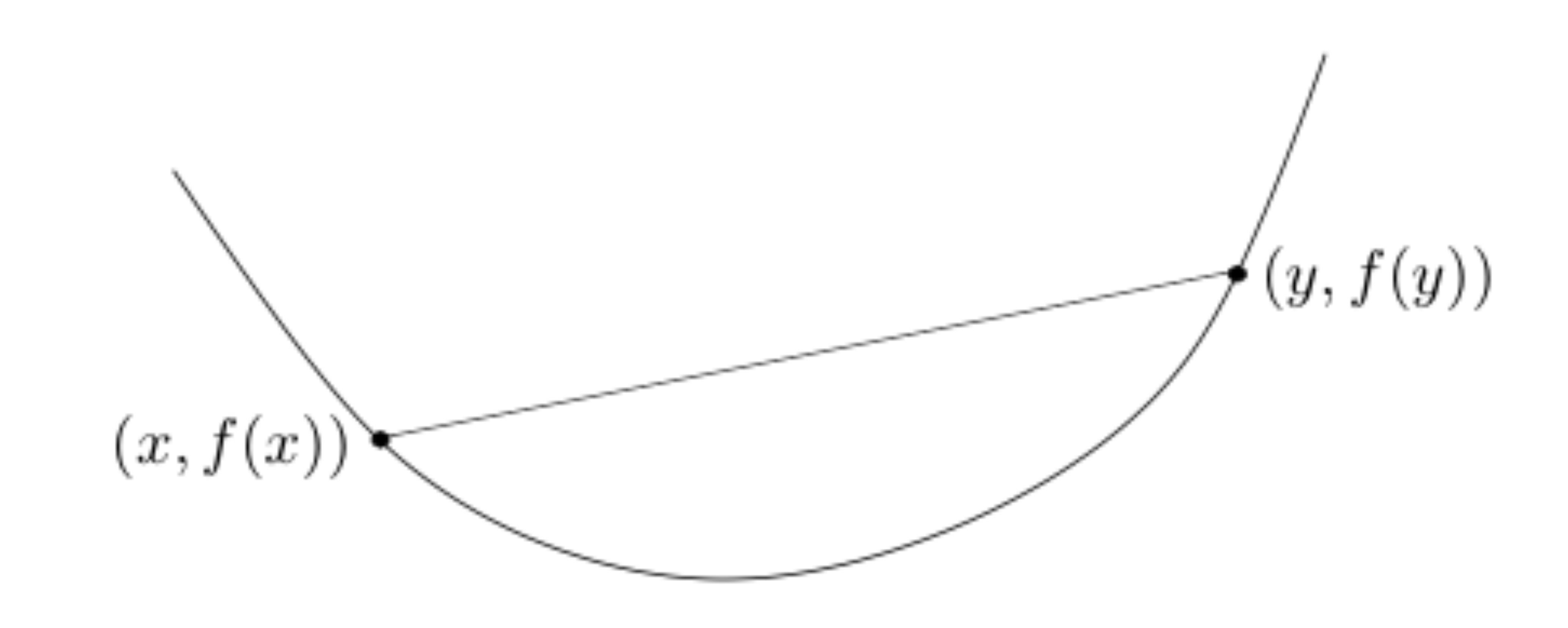
What about union?

$$A \cup B$$

CONVEX FUNCTIONS

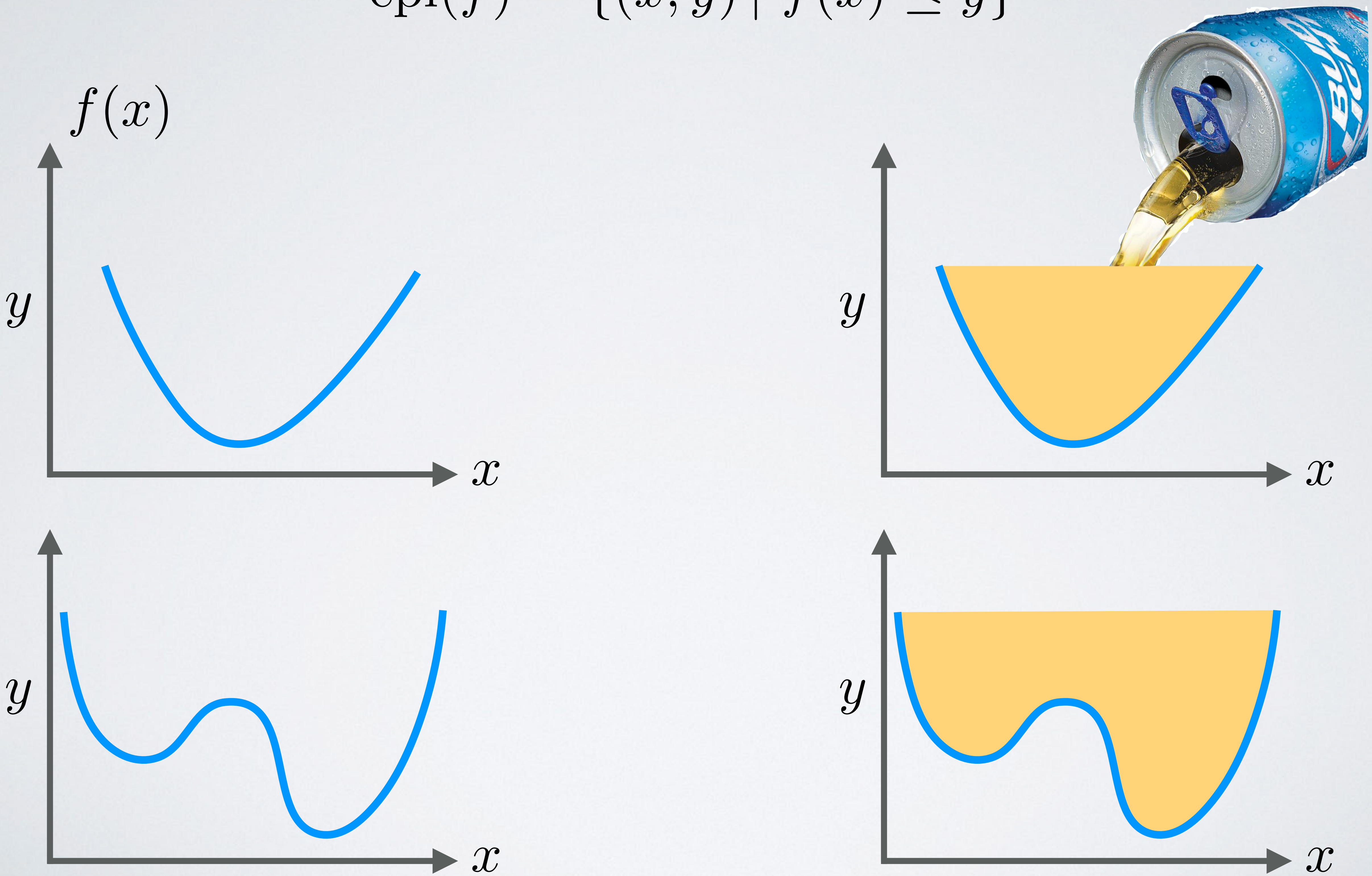
Jensen's Inequality

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



EPIGRAPH

$$\mathrm{epi}(f) = \{(x, y) \mid f(x) \leq y\}$$



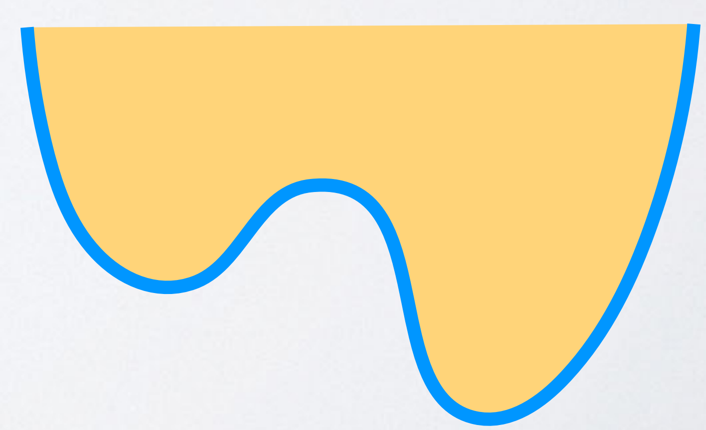
EPIGRAPH

$$\mathrm{epi}(f) = \{(x, y) \mid f(x) \leq y\}$$

convex function = convex epigraph

convex

non-convex



SOME DEFINITIONS

Theorem

"Any **proper**, **closed**, function with **bounded level sets** has a minimizer"

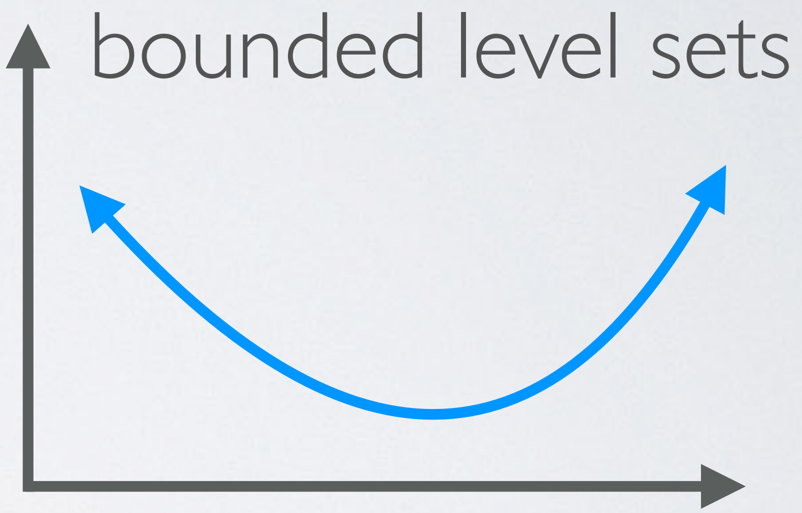
Proper : epigraph is non-empty

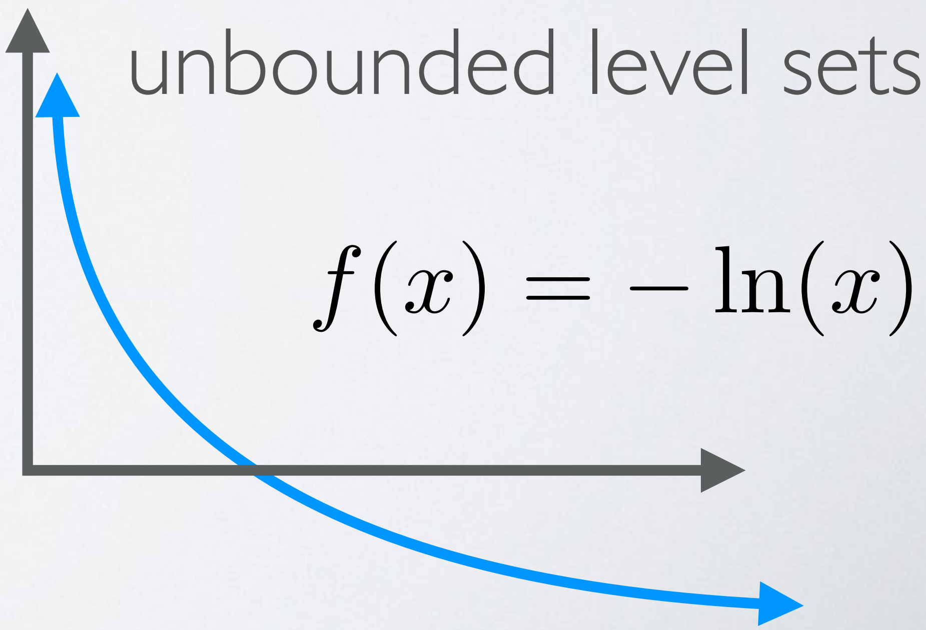
Coercive :

$$\|x\| \to \infty \implies f(x) \to \infty$$

Bounded level sets :

$$\forall \alpha \, \exists r, \; f(x) \leq \alpha \implies \|x\| \leq r$$

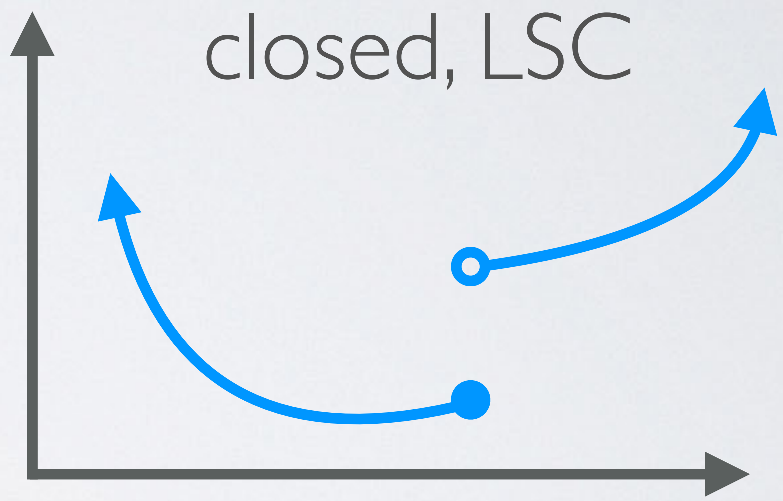




SOME DEFINITIONS

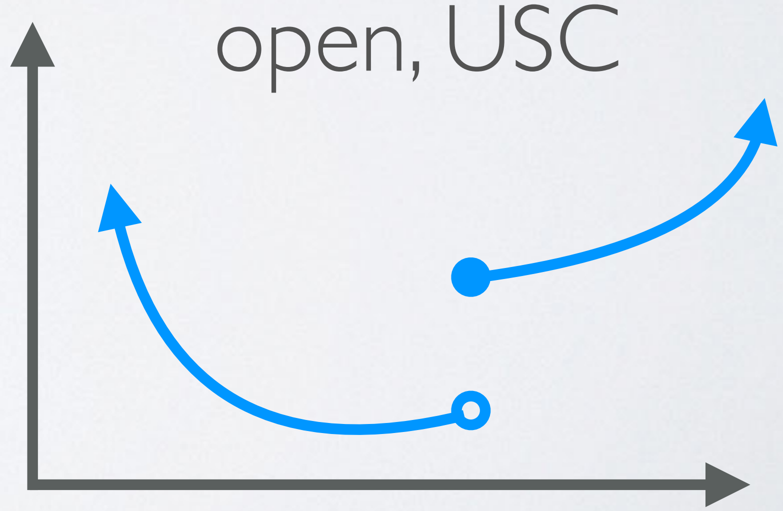
$$\mathrm{epi}(f) = \{(x, y) \mid f(x) \le y\}$$

Closed : epigraph is closed
or level sets are closed



Lower semi-continuous :

$$\forall \epsilon > 0, \exists \delta > 0, |x - x_0| < \delta$$
$$\implies f(x) \ge f(x_0) - \epsilon$$



Proper+closed = LSC

WHY DO WE CARE ABOUT CONVEX FUNCTION?

Any closed function with bounded level sets has a minimizer (by compactness)

…but, for a convex function,

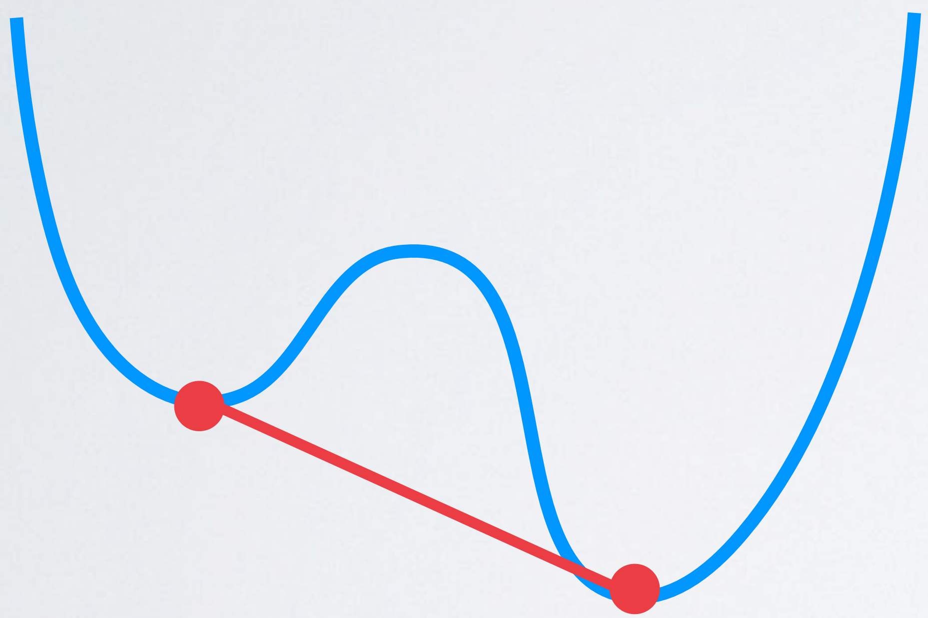
Any minimizer is **global** (why)

Set of minima is **convex** (why)

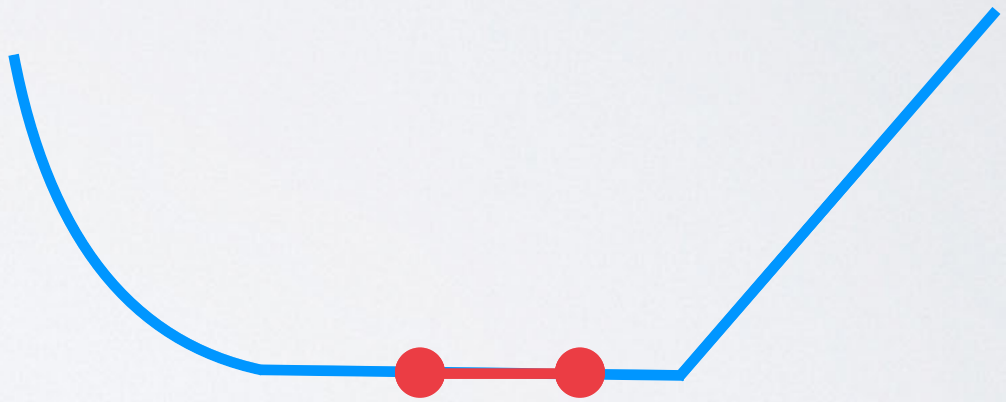
Therefore, we can find global minimizers

WHY DO WE CARE ABOUT CONVEX FUNCTIONS?

This can't happen!

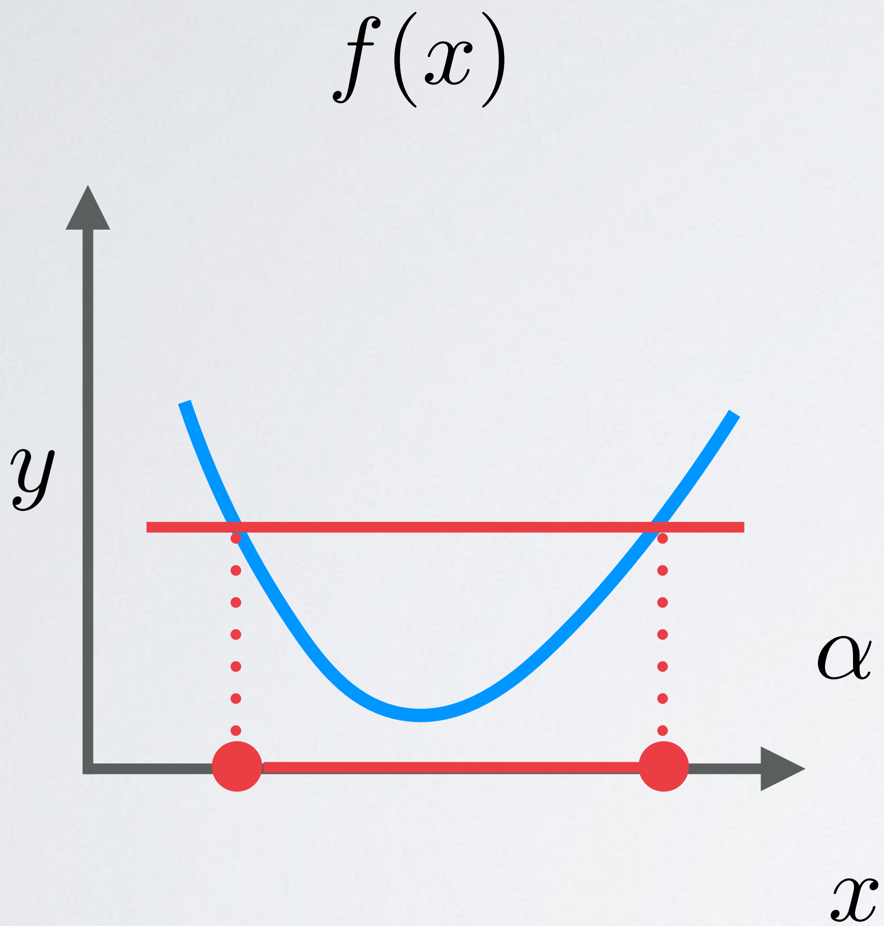


Minimizers of **convex** problems form a **convex** set



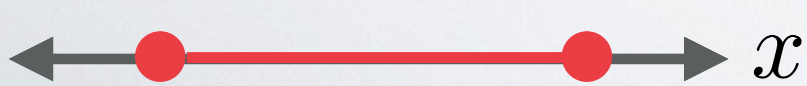
If you found one you found them all!

CONVEX FUNCTIONS HAVE CONVEX SUB-LEVEL SETS

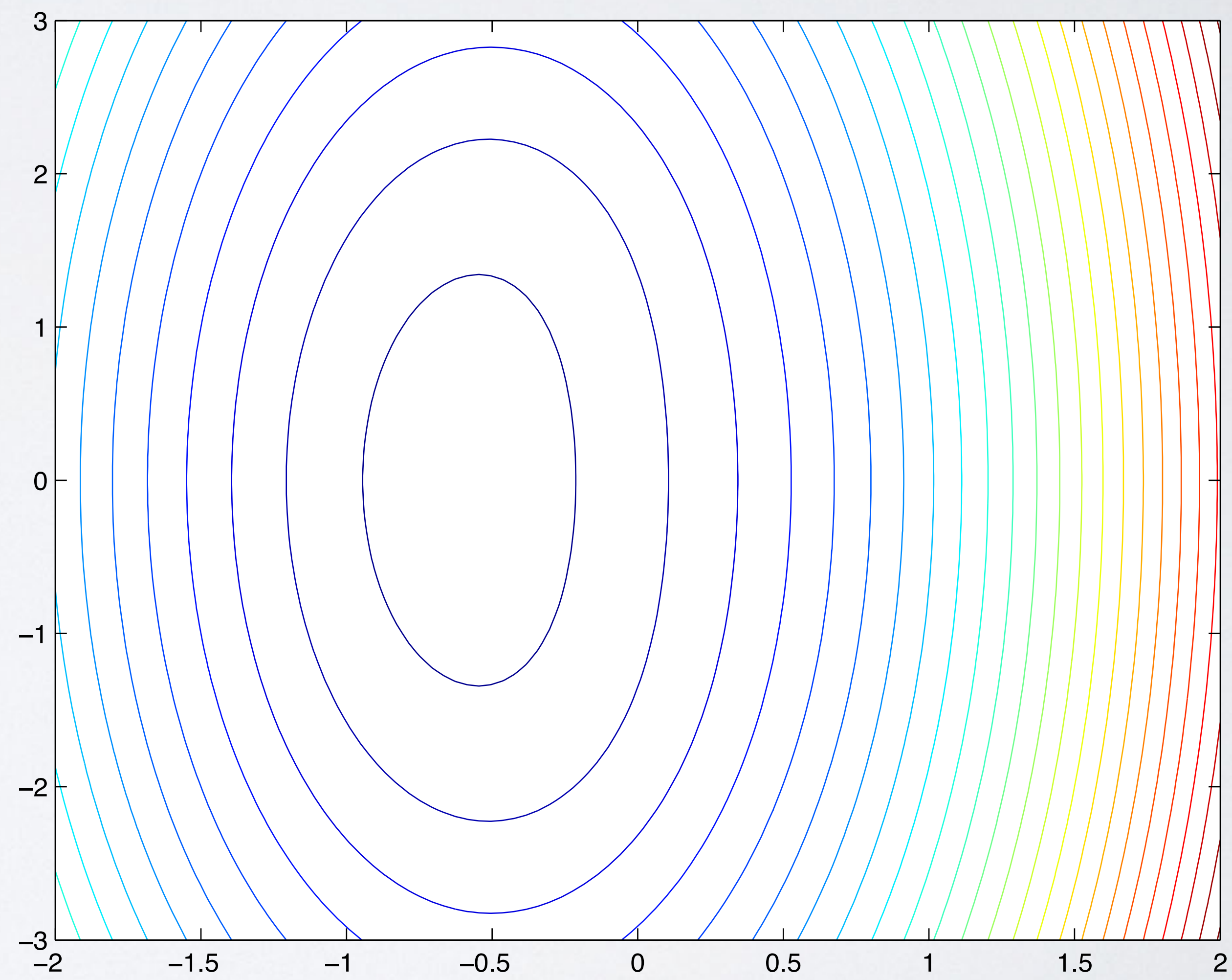


sub-level set

$$f_\alpha = \{x \mid f(x) \leq \alpha\}$$



convex contours



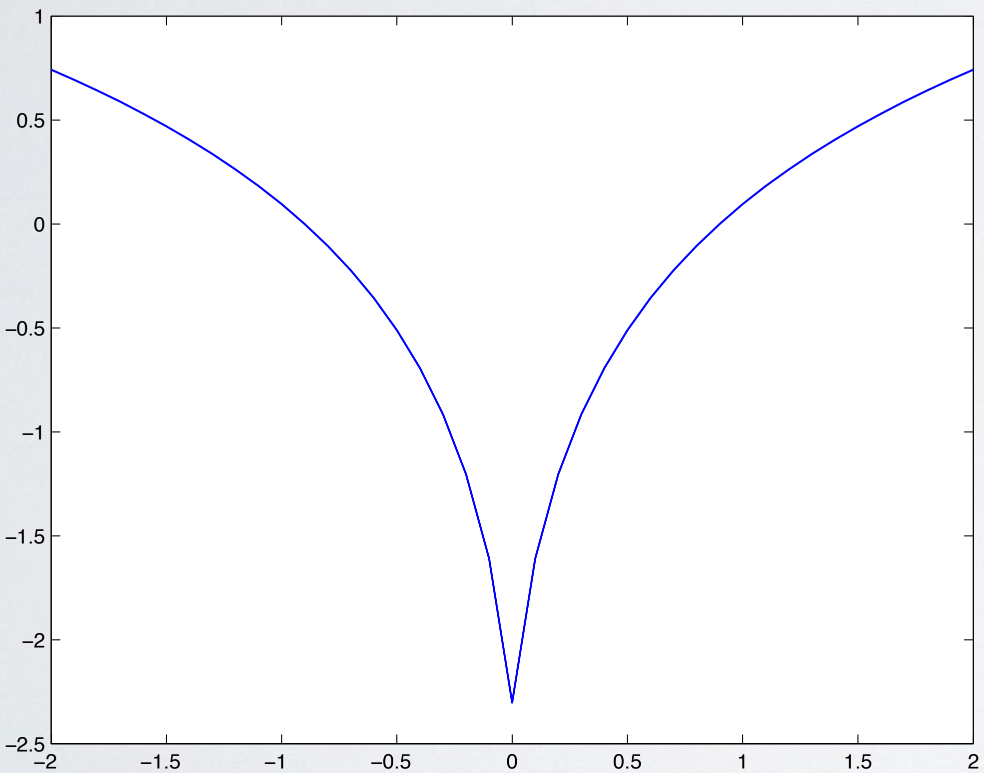
QUASI-CONVEX

Non-convex problems can still have nice properties…

convex contours = quasi-convex < convex

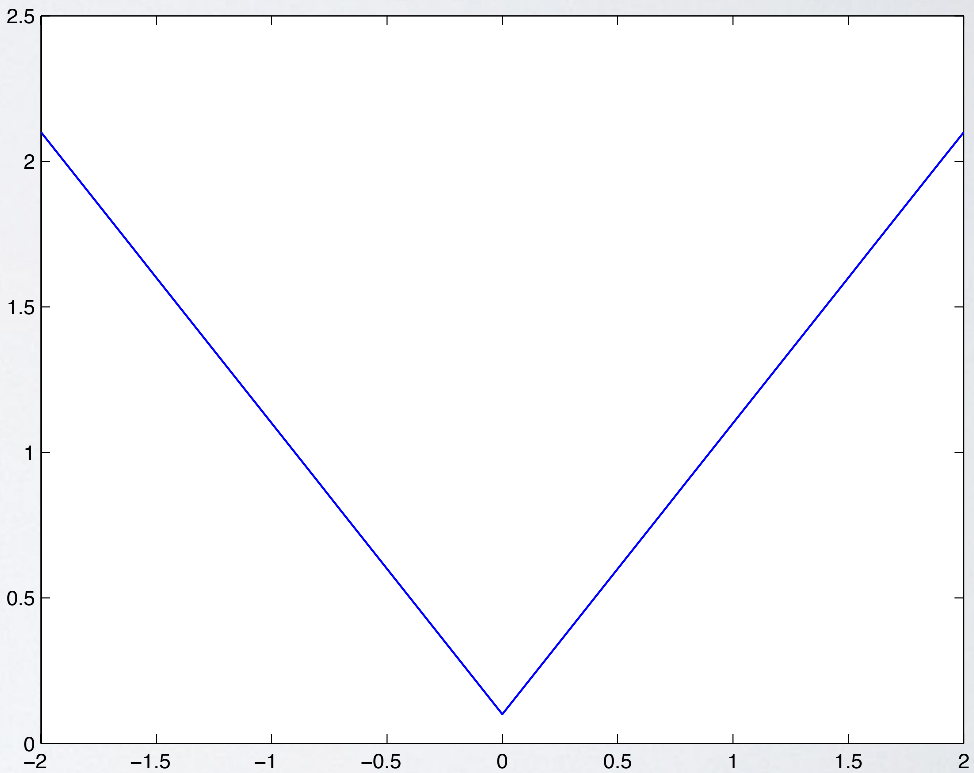
strict minima and always global mins

$$f(x) = \log(|x| + .1)$$



$$e^f$$

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



This function is "**log-convex**"

WHY DO WE CARE ABOUT CONVEX FUNCTION?

Any minimizer is **global** (why)

Set of minima is **convex** (why)

Therefore, we can find global minimizers

↓ But why convex?

Convex functions are **closed under many operations**

POSITIVE WEIGHTED SUM

Sums of convex functions are convex

$$g(x) = \sum_i f_i(x)$$

Example: $$f(x) = \|x\|^2 = \sum_i x_i^2$$

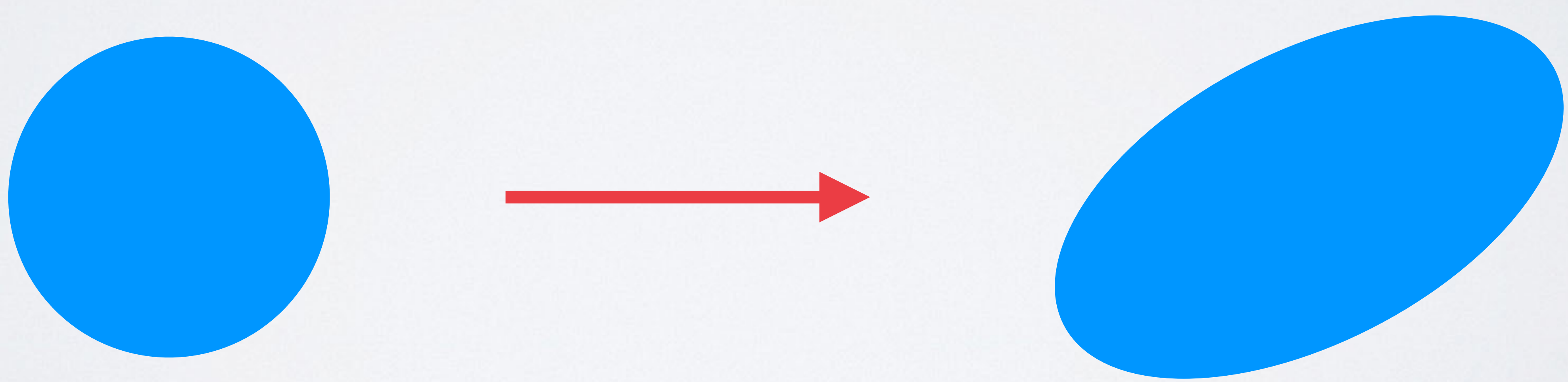
Example: $$f(x) = x + x^2 + x^6 + |x|$$

AFFINE COMPOSITION

Affine composition with convex function = convex

$$f(x) \longrightarrow f(Ax+b)$$

What does this say about epigraphs?



EXAMPLES

Least squares $$\|Ax - b\|^2 = \sum_i (A_i^T x - b)^2$$

SVM $$\frac{1}{2}\|w\|^2 + C\sum_i h(1 - d_i^T w)$$

Logistic Regression $$\sum_i \log(1 + \exp(-\ell_i d_i^T x))$$

POINTWISE-MAX

preserves convexity

$$g(x) = \max_i f_i(x)$$

Absolute value $\quad |x| = \max\{x, -x\}$

Infinity norm $\quad \|x\|_\infty = \max_i |x_i|$

Max eigenvalue $\quad \|A\|_2 = \max_v v^T A v$

What does this say about epigraphs?

WHY ARE THESE CONVEX?

Trace $f(X) = \text{trace}(A^T X)$ **Linear operator**

Distance over set $f(x) = \max_{y \in C} \|x - y\|$ **Max over convex**

Distance to set $f(x) = \min_{y \in C} \|x - y\|$ **Min of convex (special case)**

Max eigenvalue $f(x) = \|b + \sum_i A_i x_i\|_2$ **Affine comp**

If g(x,y) is convex, then minimizing for y preserves convexity

WHY ARE THESE NON-CONVEX?

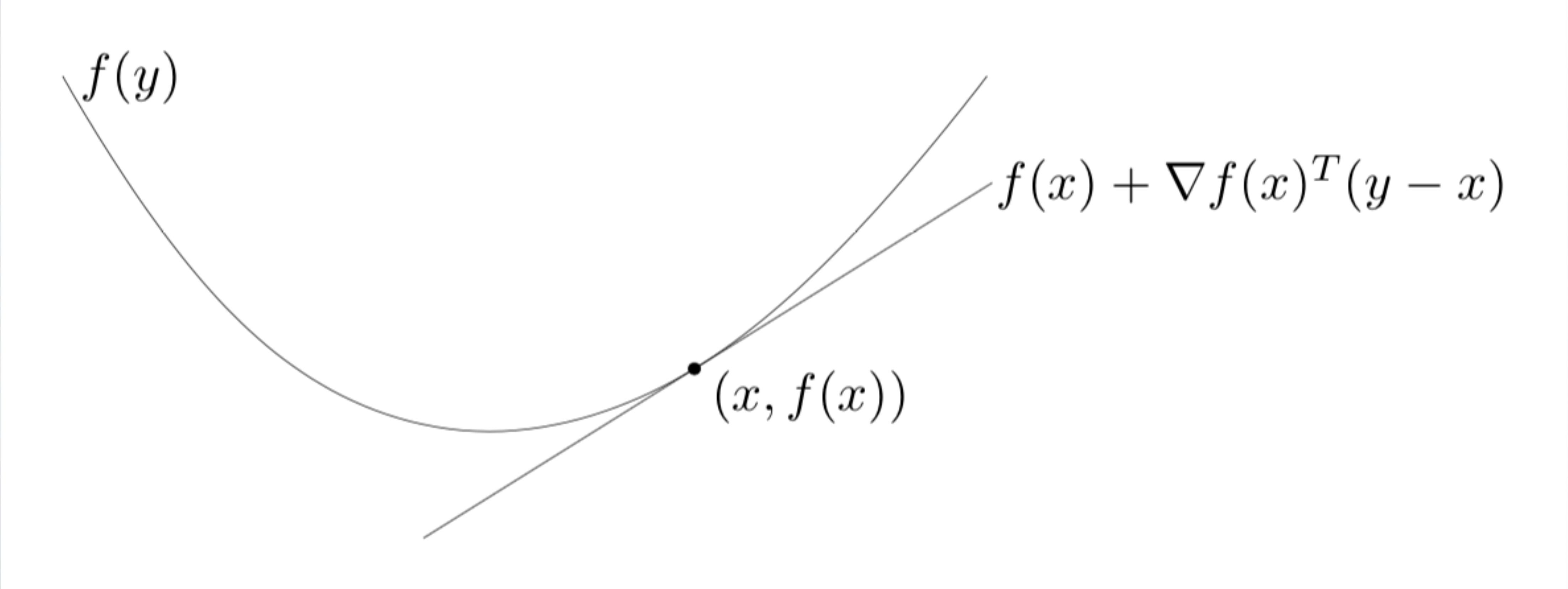
Neural Net $y = \sigma(X_3\sigma(X_2\sigma(X_1 D)))$ **Comp of Convex**

Dictionary learning $f(X,Y) = \|XY - B\|_{fro}$ **NOT affine**

DIFFERENTIAL PROPERTIES

IMPORTANT PROPERTY

Convex functions lie **ABOVE** their linear approximation



OPTIMALITY CONDITIONS

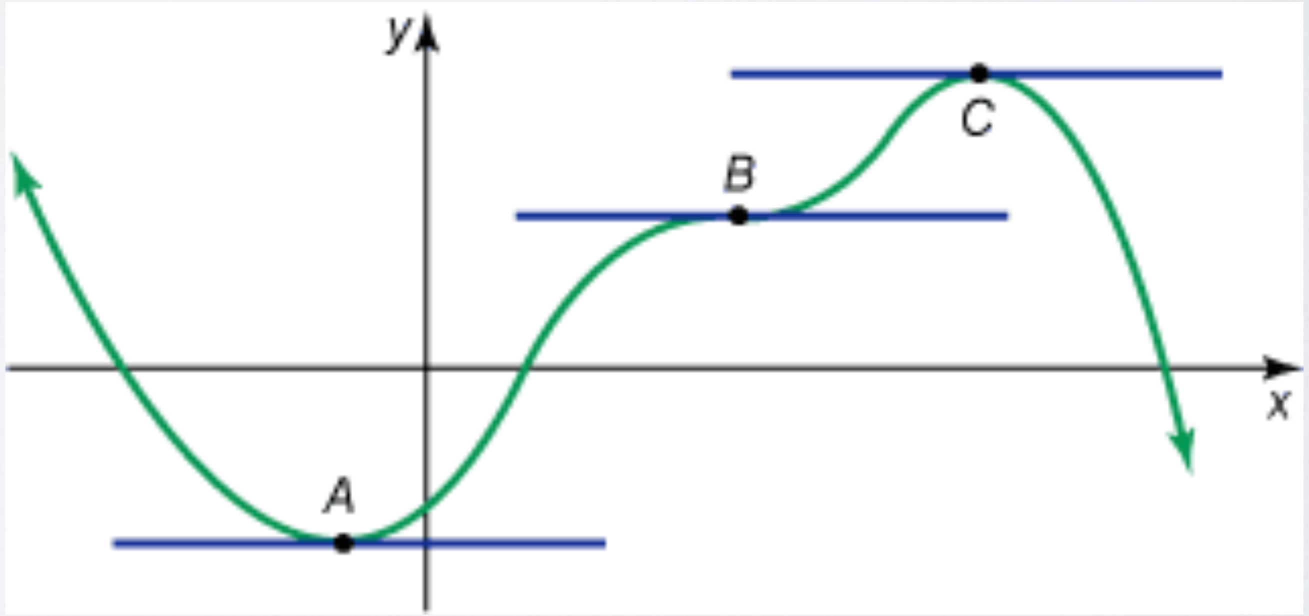
First-order conditions = optimality for convex funcs only

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

=0

$$f(y) \geq f(x)$$

This doesn't happen!!

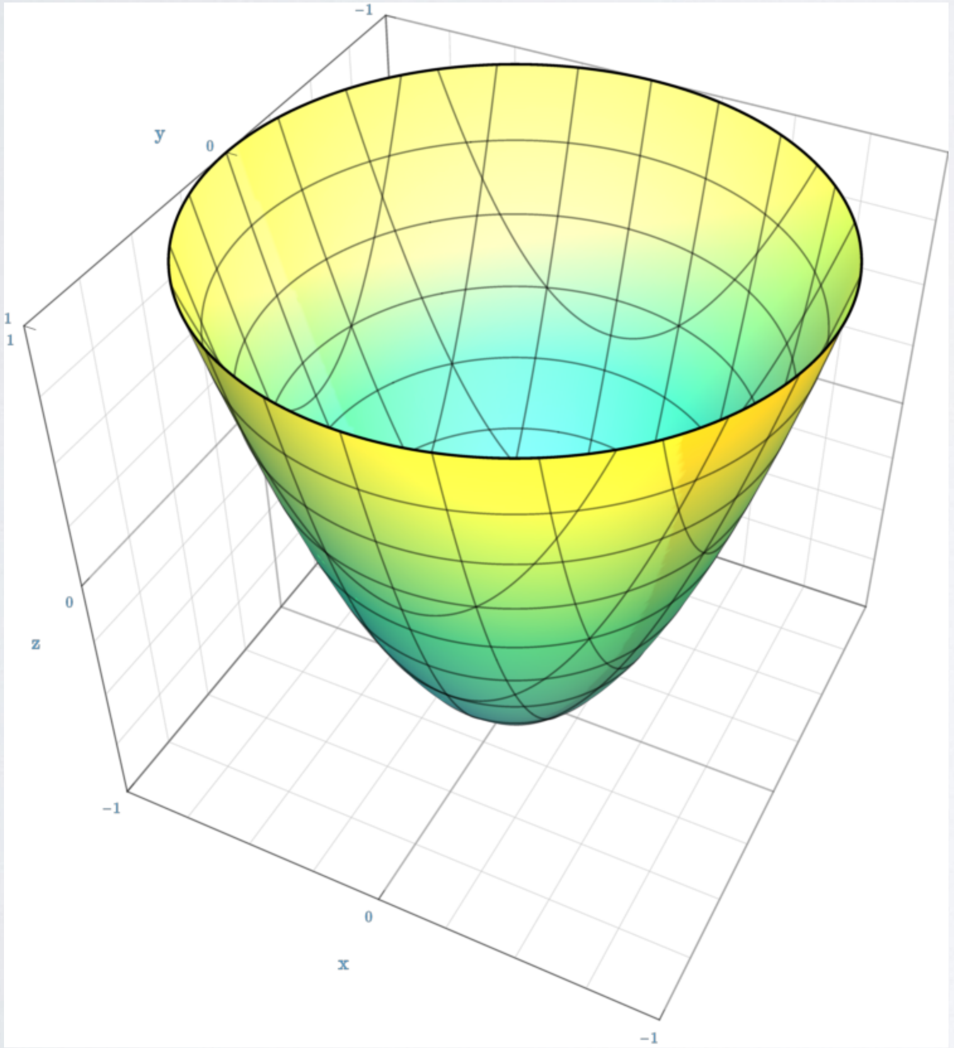


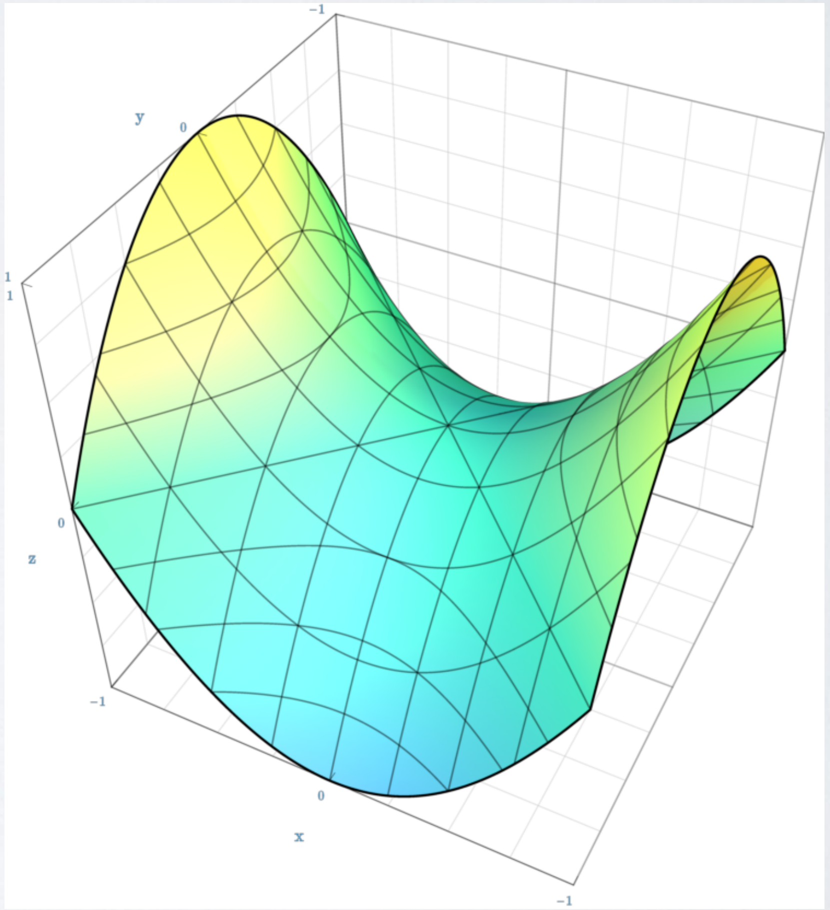
Any minima is a **global** minima

SECOND ORDER CONDITIONS

A smooth function is convex iff $\nabla^2 f(x) \succeq 0, \ \forall x$

For non-convex functions, minima satisfy: $\nabla^2 f(x^\star) \succeq 0$





remember - the Hessian is a good local model of a smooth function

STRONG CONVEXITY

What about when there's no Hessian?

$$f(y) \geq \underline{f(x) + (y-x)^T \nabla f(x)} + \frac{m}{2}\underline{\|y-x\|^2}$$

holds for any convex f | min curvature

When Hessian exists…

$$f(y) \approx f(x) + (y-x)^T \nabla f(x) + \frac{1}{2}(y-x)^T \nabla^2 f(x)(y-x)$$

$$\geq f(x) + (y-x)^T \nabla f(x) + \frac{\lambda_{min}}{2}\|y-x\|^2$$

…and so $\nabla^2 f(x) \succeq mI$

UPPER BOUND ON CURVATURE

Lower $$f(y) \geq f(x) + (y-x)^T \nabla f(x) + \frac{m}{2}\|y-x\|^2$$

upper $$f(y) \leq f(x) + (y-x)^T \nabla f(x) + \frac{M}{2}\|y-x\|^2$$

$$f(y) \approx f(x) + (y-x)^T \nabla f(x) + \frac{1}{2}(y-x)^T \nabla^2 f(x)(y-x)$$
$$\leq f(x) + (y-x)^T \nabla f(x) + \frac{\lambda_{max}}{2}\|y-x\|^2$$

LIPSCHITZ CONSTANT

$$\|\nabla f(x) - \nabla f(y)\| \leq M\|y - x\|$$

upper

$$f(y) \leq f(x) + (y - x)^T \nabla f(x) + \frac{M}{2}\|y - x\|^2$$

We use this when there's no Hessian (ex: Huber function)

OBJECTIVE ERROR BOUNDS

Lower $$f(y) \geq f(x) + (y-x)^T \nabla f(x) + \frac{m}{2}\|y-x\|^2$$

upper $$f(y) \leq f(x) + (y-x)^T \nabla f(x) + \frac{M}{2}\|y-x\|^2$$

We can bound the objective error
in terms of distance from minimizer

$$f(y) - f(x^\star) \geq \frac{m}{2}\|y-x^\star\|^2$$

$$f(y) - f(x^\star) \leq \frac{M}{2}\|y-x^\star\|^2$$

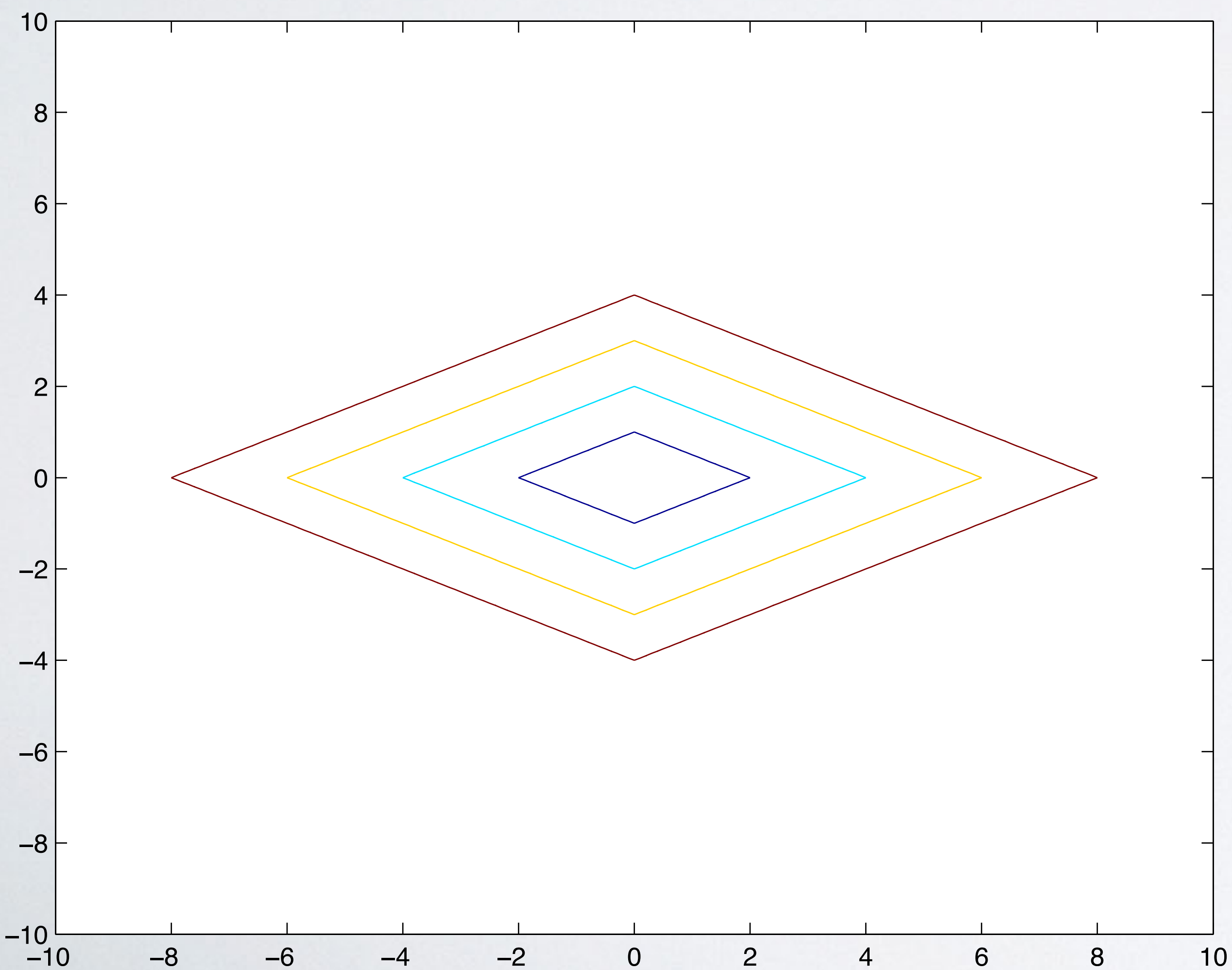
CONDITION NUMBER

For any function → $$\kappa = \frac{\text{major axis}}{\text{minor axis}}$$

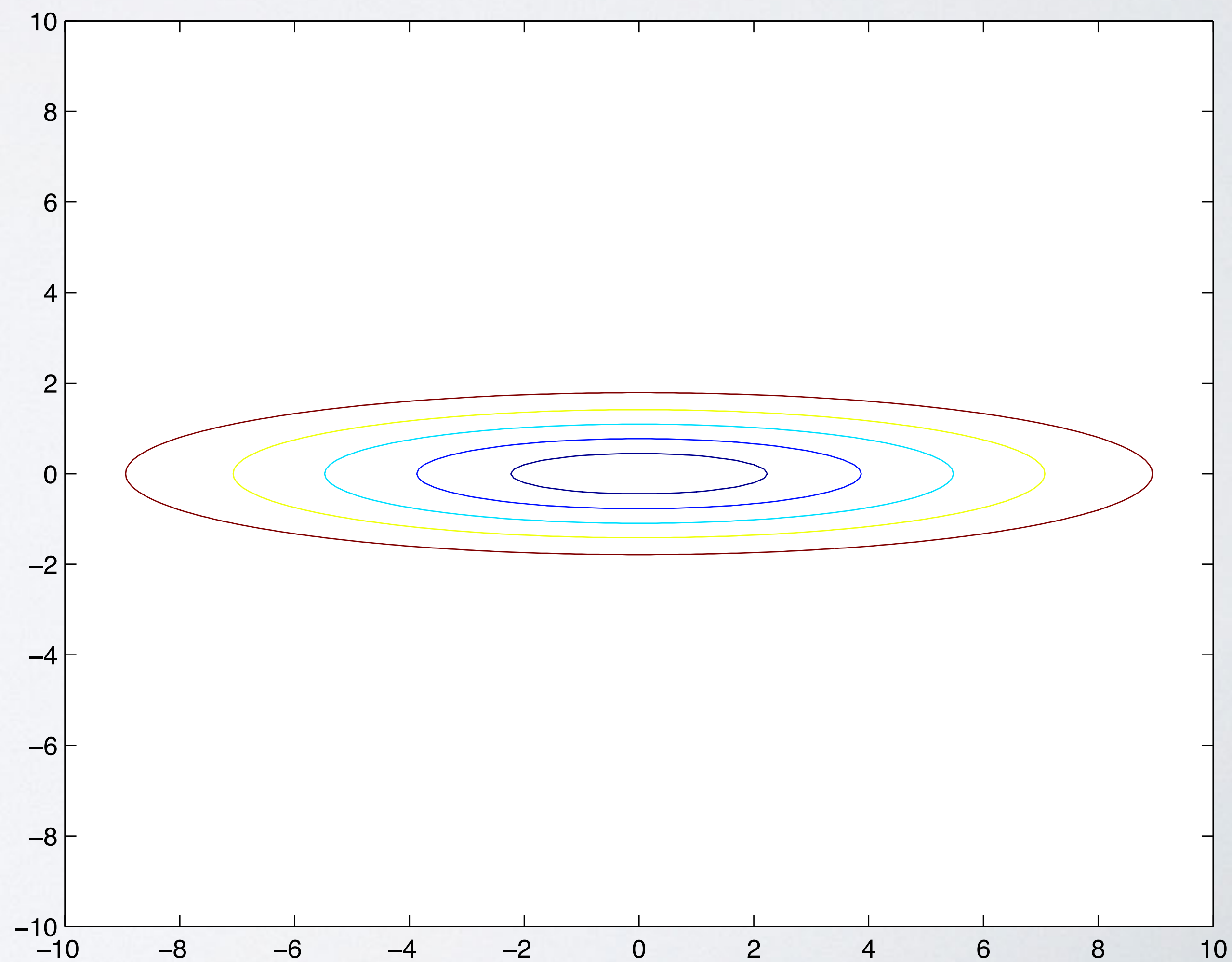
For smooth functions → $$\kappa \approx \text{cond}(\nabla^2 f(x))$$

For differentiable functions → $$\kappa = M/m$$

$\kappa = 1.5$

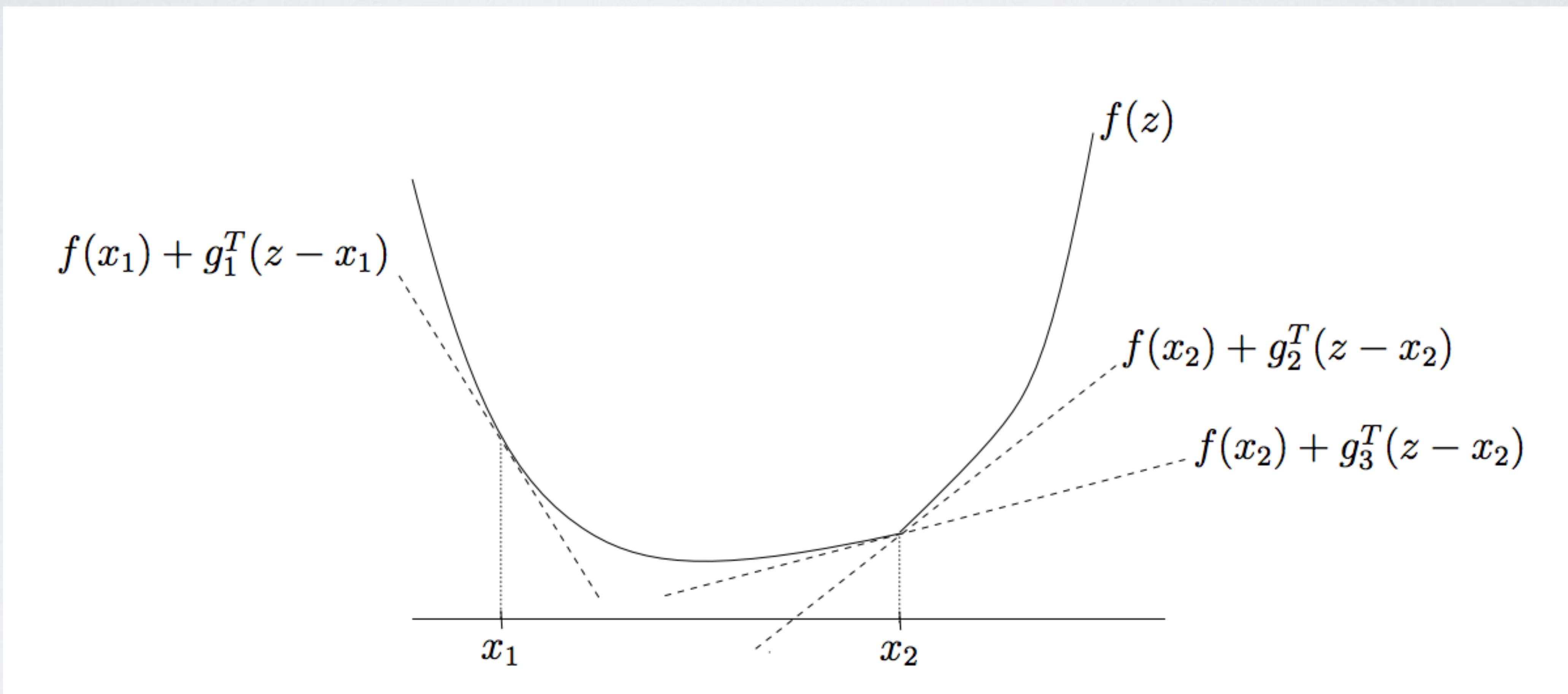


$\kappa = 5$



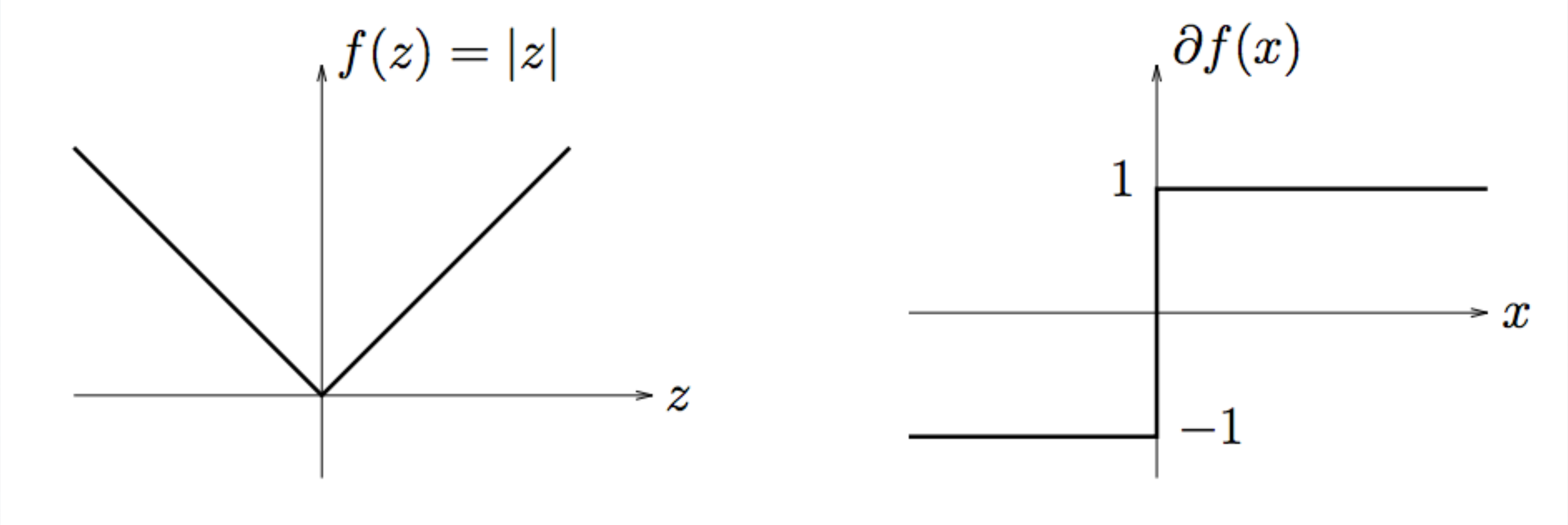
SUB-DIFFERENTIAL

$$\partial f(x) = \{g : f(y) > f(x) + (y - x)^T g,\ \forall y\}$$



Optimality: $0 \in \partial f(x^*)$

EXAMPLE



$$\partial f(0) = [-1, 1]$$

MONOTONICITY

The (sub) gradient of any convex function is **monotone**

$$\langle y - x, \nabla f(y) - \nabla f(x) \rangle \geq 0$$

or

$$\langle y - x, g_y - g_x \rangle \geq 0$$

for any

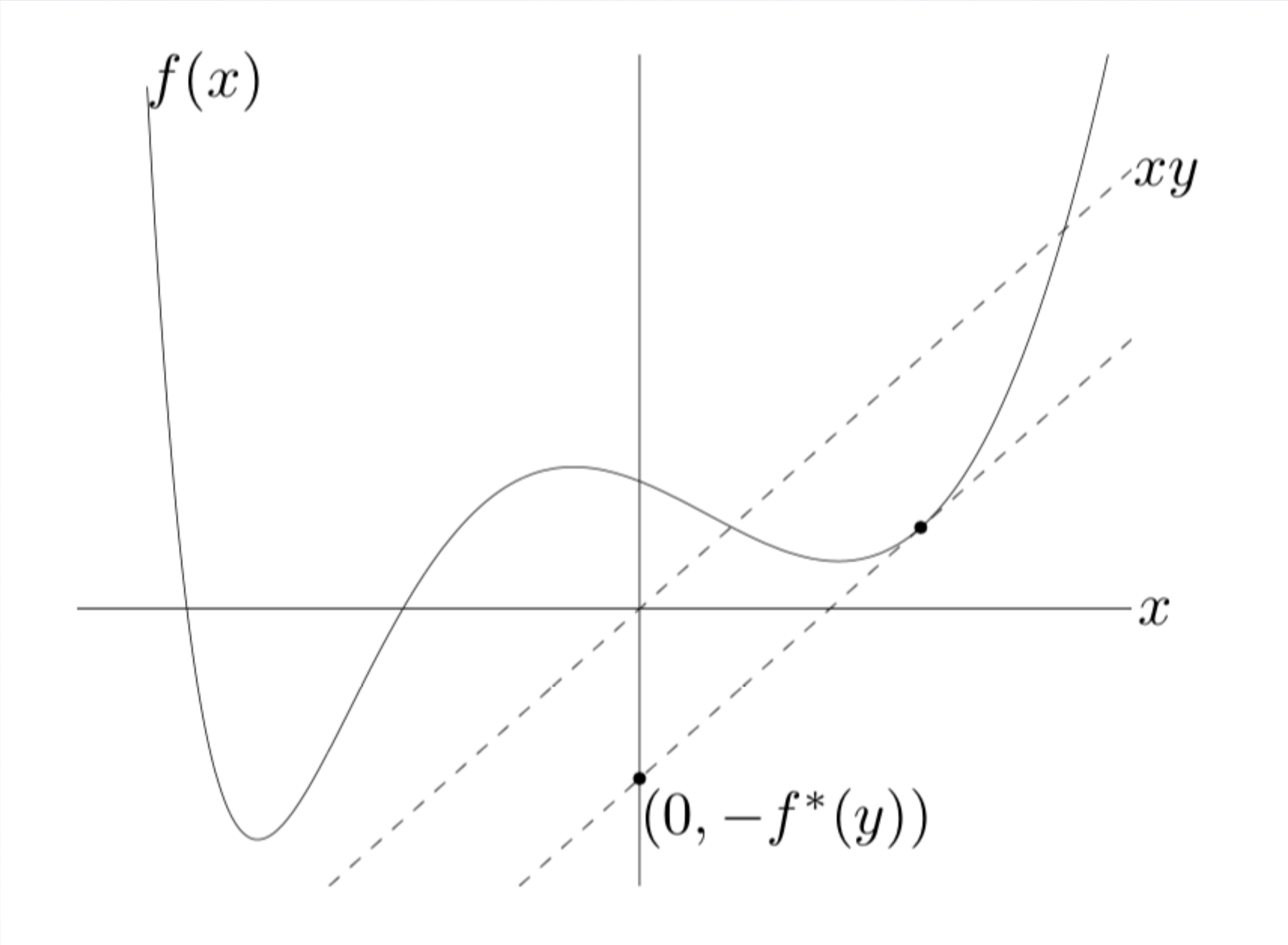
$$g_x \in \partial f(x), \text{ and } g_y \in \partial f(y)$$

This generalizes the concept of PSD Hessian

$$(y - x)^T (\nabla f(y) - \nabla f(x)) = (y - x)^T H (y - x) \geq 0$$

CONJUGATE FUNCTION

$$f^*(y) = \max_x \; y^T x - f(x)$$



Is it convex?

EXAMPLE: QUADRATIC

$$f(x) = \frac{1}{2}x^T Q x$$

$$f^*(y) = \max_x \; y^T x - \frac{1}{2}x^T Q x$$

$$y - Qx^\star = 0$$

$$x^\star = Q^{-1}y$$

$$f^*(y) = y^T Q^{-1} y - \frac{1}{2} y^T Q^{-1} Q Q^{-1} y$$

$$f^*(y) = \frac{1}{2} y^T Q^{-1} y$$

CONJUGATE OF NORM

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

$$f^*(y) = \max_x \; y^T x - \|x\|$$

dual norm

$$\|y\|_* \triangleq \max_x y^T x / \|x\|$$

Holder inequality

$$y^T x \le \|y\|_* \|x\|$$

$$\|y\|_* \le 1 \longrightarrow f^* = 0$$

why?

$$\|y\|_* > 1 \longrightarrow f^* = \infty$$

why?

$$f^*(y) = \begin{cases} 0, & \|y\|_* \le 1 \\ \infty, & \text{otherwise} \end{cases}$$

EXAMPLES

Holder inequality for p-norms

$$|\langle x, y\rangle| \leq \|x\|_p \|y\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1$$

$$f(x) = \|x\|_2, \qquad f^*(y) = \mathcal{X}_2(y) = \begin{cases} 0, & \|x\|_2 \leq 1 \\ \infty, & \|x\|_2 > 1 \end{cases}$$

$$f(x) = |x|, \qquad f^*(y) = \mathcal{X}_\infty(y) = \begin{cases} 0, & \|x\|_\infty \leq 1 \\ \infty, & \|x\|_\infty > 1 \end{cases}$$

$$f(x) = \|x\|_\infty, \qquad f^*(y) = \mathcal{X}_1(y) = \begin{cases} 0, & |x| \leq 1 \\ \infty, & |x| > 1 \end{cases}$$

Useful when we study **duality**

PROPERTIES OF CONJUGATE

conjugate

$$f^*(y) = \max_x \; y^T x - f(x)$$

x is the point where f has gradient y (why?)

$$y \in \partial f(x)$$

The gradient of conjugate = the "adjoint" of the gradient (why?)

$$y = \nabla f(x) \iff \nabla f^*(y) = x$$

…and in general

$$y \in \partial f(x) \iff x \in \partial f^*(y)$$

also important for convergence proofs!