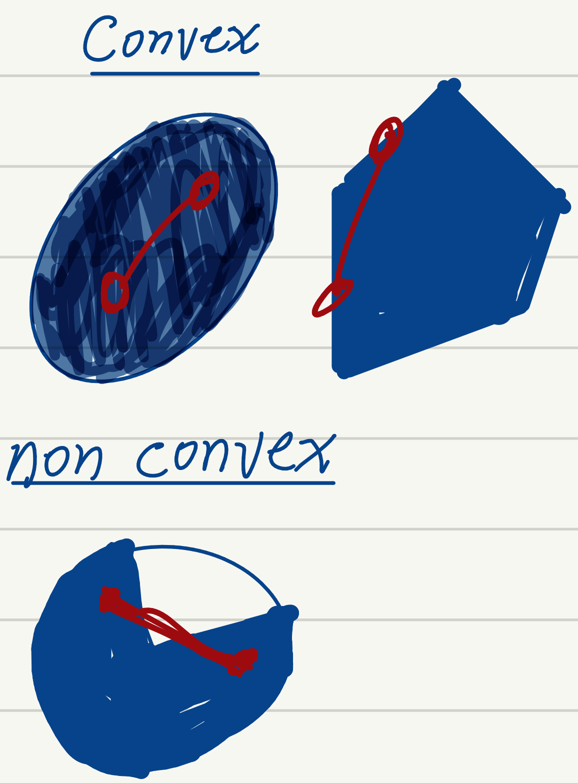
Convex Learning Problems (Reference: Ch 12 Shalev-Schwartz)

Definition (Convexity)

A set C in a vector space, is convex if for any two vectors u, v in C the line segment between u and v is contained in C.

That is, for any $\alpha \in [0,1]$

$$\alpha u + (1-\alpha) v \in C$$

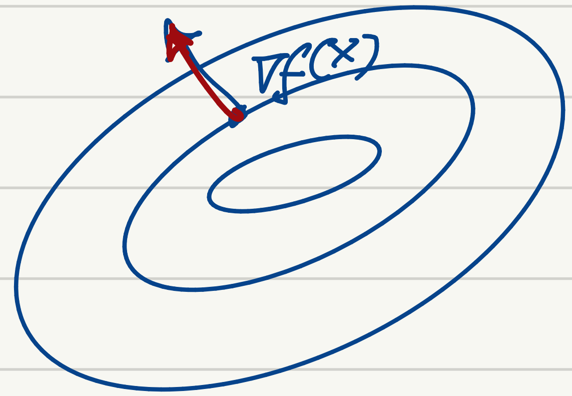


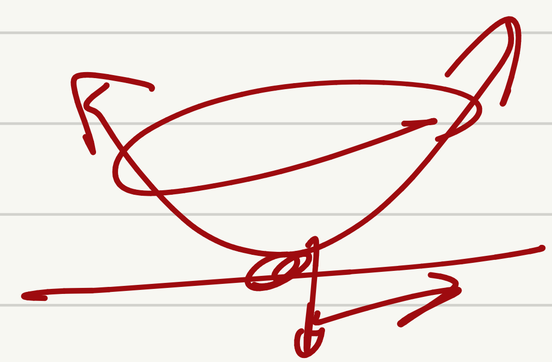
Note $f: \mathbb{R}^n \to \mathbb{R}$

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right)$$

gradient

vector pointing in direction of greatest increase for $f(x)$

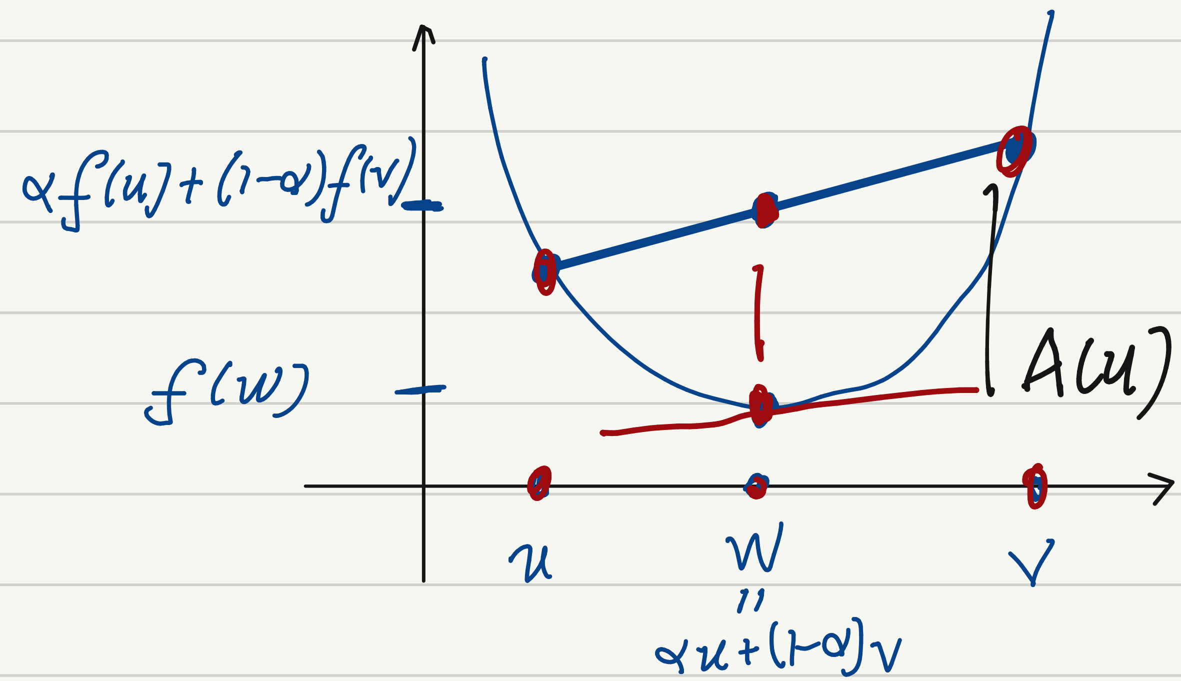




Defn (Convex function)

A function from a convex set C, $f: C \to \mathbb{R}$ is convex if, for every $u, v \in C$ and $\alpha \in [0,1]$

$$f(\underbrace{\alpha u + (1-\alpha) v}_{w}) \leq \alpha f(u) + (1-\alpha) f(v)$$



$$A(u) = g \cdot (u - w) + f(w)$$

Check $\quad A(w) = g \cdot 0 + f(w) = f(w).$

$f(w) = \frac{w^2}{2}$

$w = 3$

$A(u) = 3(u-3) + 9/2$

check $\quad u = 7 \qquad 49/2 = f(7)$

$A(7) = 3 \cdot 4 + 9/2$

Gap at 7 is $\; 24.5 - 16.5 = 8$

Properties of convex functions

Ref Boyd Convex Opt.

- Every local minimum is a global minimum (exercise)
- Supporting hyperplane (tangent) property.

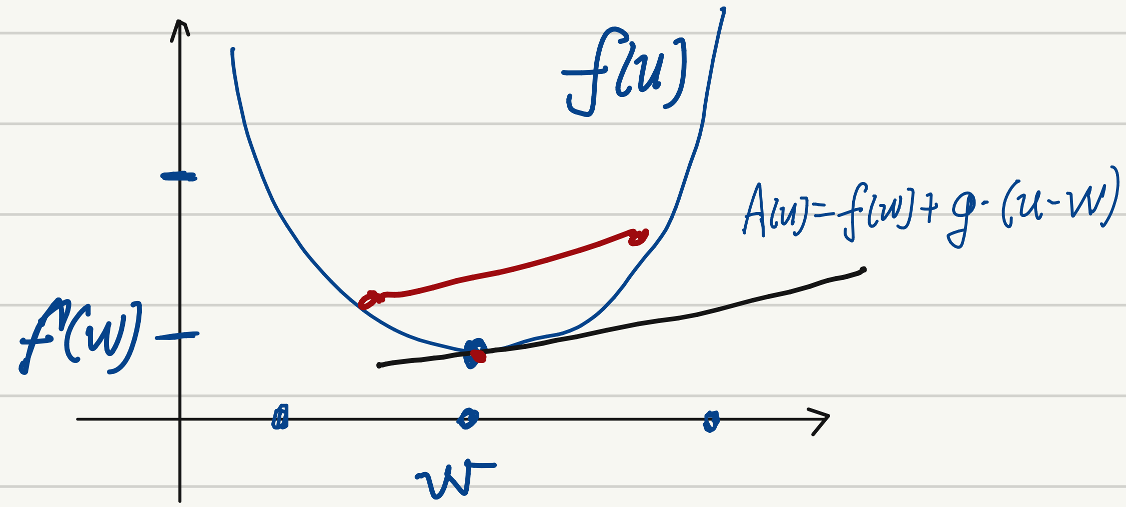
Defn Let $f : C \to \mathbb{R}$ where C is an open, convex set

Define, for vectors g, the affine function

$$A_g(u) = f(w) + g\cdot(u-w)$$

$A(w) = f(w)$

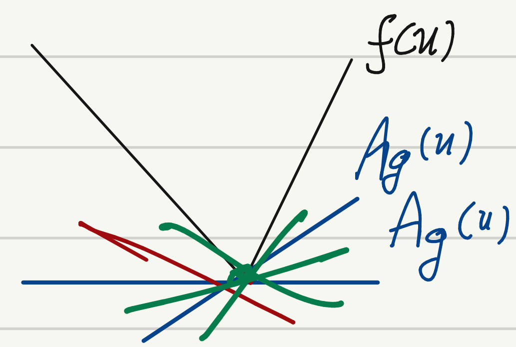
$g = \nabla f(w)$



Lemma: the function $f : C \to \mathbb{R}$, C open, convex, is convex iff for every $w \in C$, there exists g such that

$$f(u) \geq A_g(u) \quad \text{for all } u \in C \quad (*)$$

$A_g(w) = f(w)$



$f(x) = |x|$

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

A vector g that satisfies $*$ is called a subgradient of f at w. $\partial f(w) = \{\text{all subgradients of } f \text{ at } w\}$

proof / discussion of these results.

<u>Ref</u> Ch 1 Convex Analysis.

distance.

<u>Fact</u>

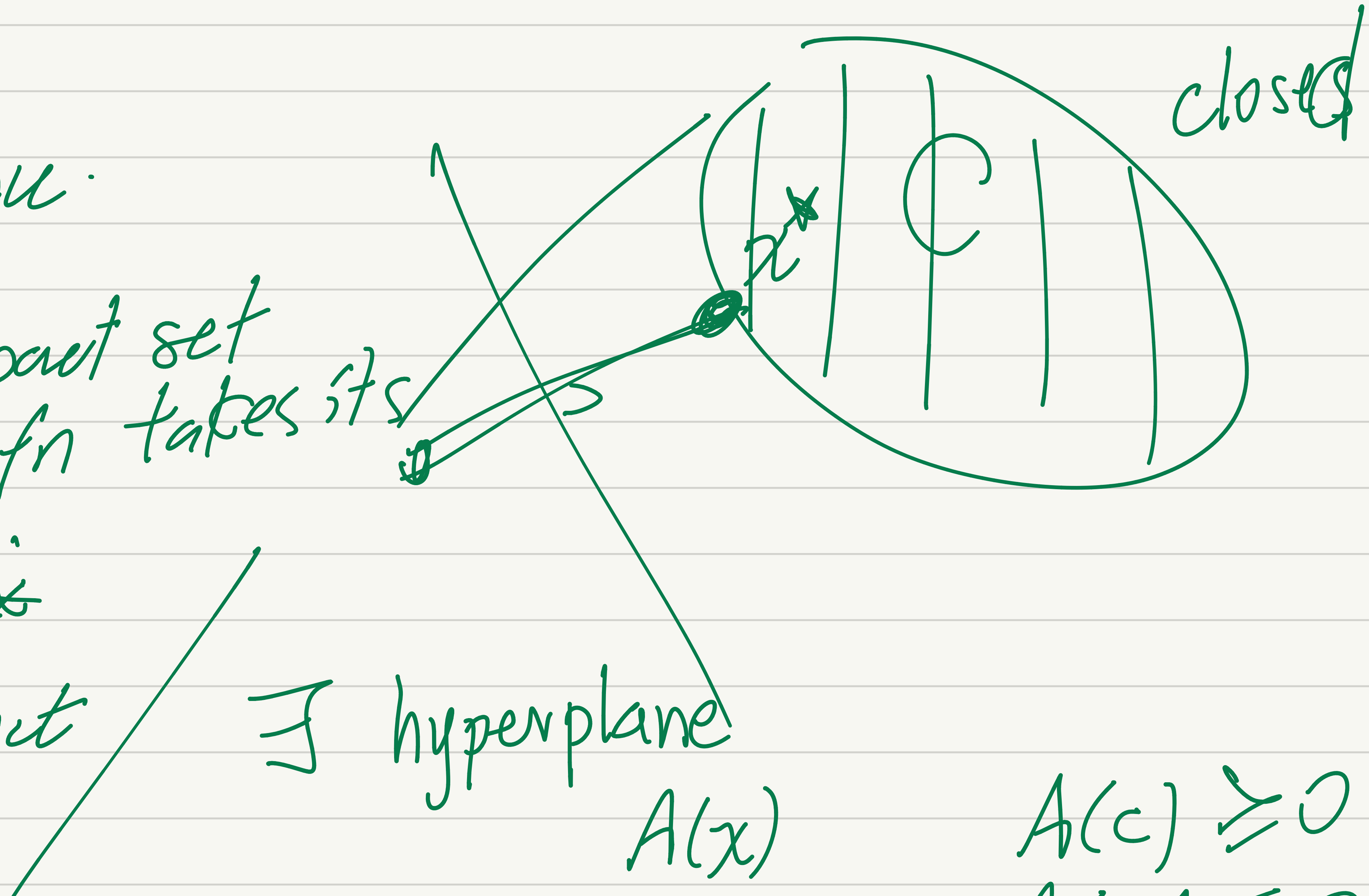
compact set

cts fn takes its

min.

$\Rightarrow x^*$

constant



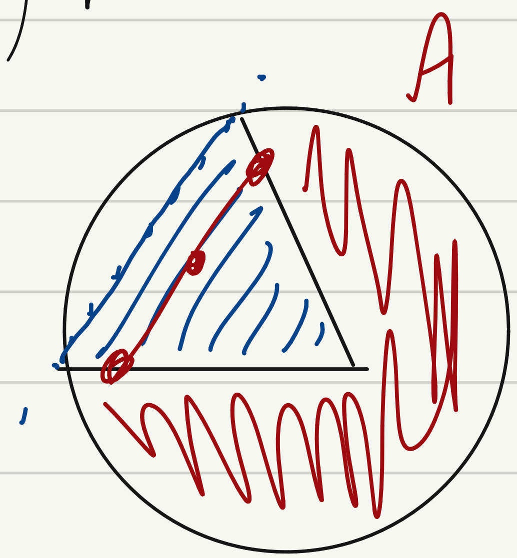
$\exists$ hyperplane $A(x)$

$A(c) \geq 0 \quad \forall c \in C$

$A(x) < 0$

Definition Convex Hull (of a set) A

smallest convex set containing A



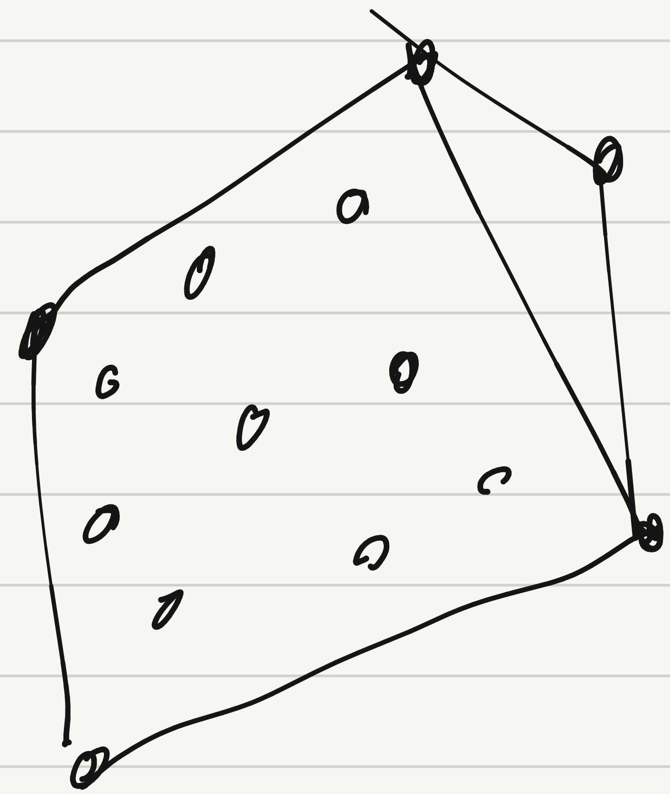
$$co(A) = \left\{ \bigcap C \;\middle|\; \begin{array}{l} C \supset A, \\ C \text{ convex} \end{array} \right\}$$

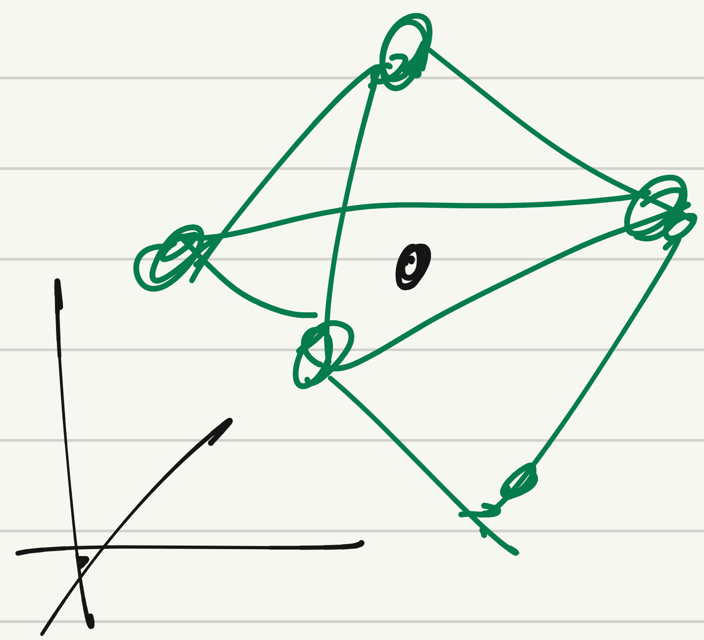
can write every $c \in co(A)$ as $\quad A \subset \mathbb{R}^d$

$$c = \left\{ \sum_{i=1}^{n} w_i a_i \quad \begin{array}{l} \vec{w} \text{ weight vector} \\ \sum_{i=1}^{n} w_i = 1 \quad w_i \geq 0 \end{array} \right\}$$

Mazur's Lemma

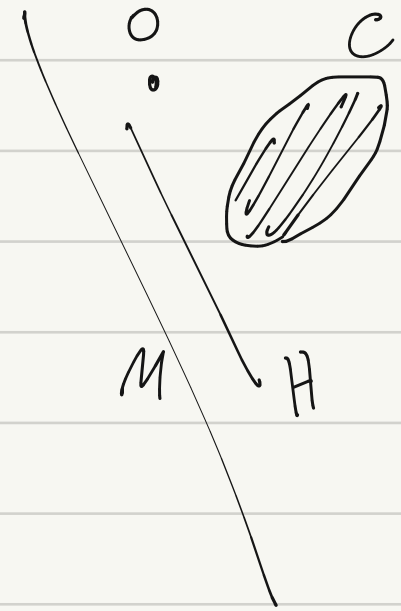
EX





Separation Thm

Difference between finite & ∞ dim.



V vector space (could be ∞-dimensional)

[e.g. ℓ^2 $\{(x_1, \dots, x_n, \dots)\}$, $\|x\|_2^2 = \sum x_i^2 < \infty$]

Hahn Banach Thm

V vector space

C open convex set non-empty

M non-empty affine subspace, $C \cap M = \emptyset$

Then there exists a separating hyperplane H, given by $A(x)$

$A(x) = a \cdot x + b$

separates if $A(c) \geq 0$ $\forall c \in C$

$A(a) \leq 0$ $\forall a \in A$

(non strict)

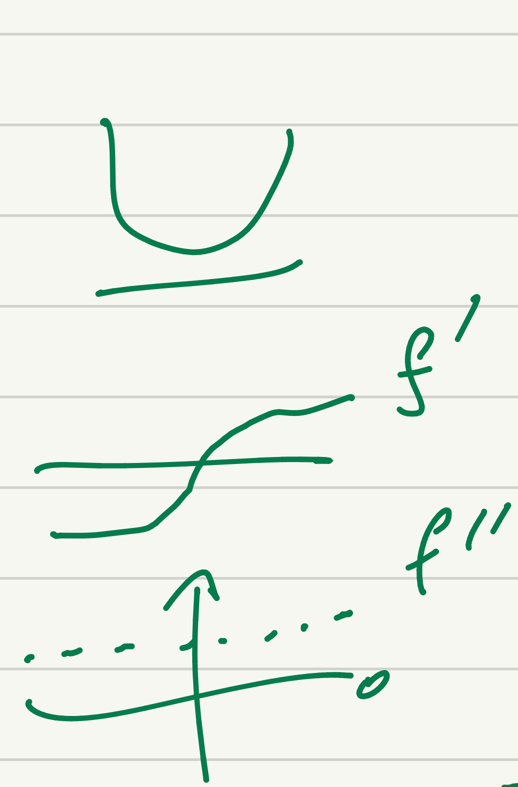
strict: > 0

< 0

More properties of convex functions.

Lemma Let $f: \mathbb{R} \to \mathbb{R}$ be twice differentiable.
the following are equivalent (TFAE)

- 1. f is convex
- 2. f' is monotonically non-decreasing
- 3. f'' is non-negative.



$f: \mathbb{R}^n \to \mathbb{R}$ $\quad C^2$. TFAE

1. f is convex
3. $D^2f(x)$ is non-negative definite
2. $(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq 0$

note: 2 extends to non-differentiable case
replacing gradient with subgradients

3. $D^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right]$

$D^2 f(x) \geq 0$ pos. definite

cond. local min

crit pt. $\nabla f(x) = 0$

Examples of Convex functions:

- $Q(x) = x^T P x + bx$ quadratic where P is positive definite

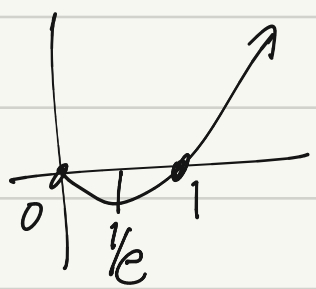
- $f(x) = x^2/2$

- $f(x) = \underline{x \log x}$ (on $x>0$)

check $f'(x) = \frac{x}{x} + \log x$

$= 1 + \log x$ increasing

$f''(x) = \frac{1}{x} > 0$



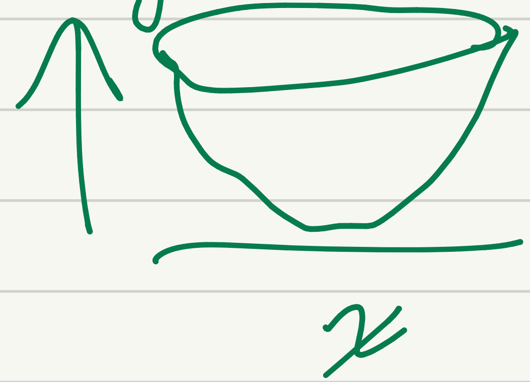
$\log x = -1$
$x = \frac{1}{e}$

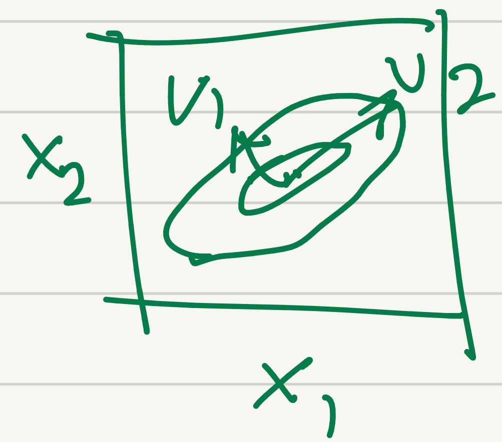
properties of convex functions

- if g is convex and $f(x)$ is affine, then $g \circ f$ is convex
- max of convex functions is convex

See Boyd & Vandenberg for more properties & examples

$y = f(x)$





$P = O^T \Lambda O$

$O^T O = I$ coln of O $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$

$P v_1 = \lambda_1 v_1$

$P v_2 = \lambda_2 v_2$

Key properties for convex optimization.

Lipschitz & "Smooth"

Defn Let $f: C \subset \mathbb{R}^d \mapsto \mathbb{R}^k$. Suppose

$$\| f(w_1) - f(w_2) \| \leq \rho \| w_1 - w_2 \| \quad \forall w_1, w_2 \in C$$

then say f is ρ-Lipshitz continous over C.

Note if f is $\nabla f(x)$ exists & $\| \nabla f(x) \| \leq \rho \quad \forall x \in C$
then f is ρ-Lipschitz.

Why? Mean-Value Thm.

$$f(w_1) - f(w_2) = \nabla f(z) \cdot (w_1 - w_2)$$

for some z.

Smoothness

Defn: A differtiable function $f : \mathbb{R}^d \to \mathbb{R}$ is β-smooth if its gradient is β-Lipschitz.

$$\|\nabla f(v) - \nabla f(w)\| \leq \beta \|v - w\| \quad \forall v, w$$

Note smoothness implies

$$f(v) \leq \underbrace{f(w) + \langle \nabla f(w), v - w \rangle}_{A_w(v)} + \frac{\beta}{2}\|v - w\|^2$$

convexity:

$$A_w(v) \leq f(v)$$

Both $\Rightarrow$ $\underline{A_w(v) \leq f(v) \leq A_w(v) + \frac{\beta}{2}\|v - w\|^2}$

means: upper & lower bounds on affine approximation

Note

$$f(v) \leq A_w(v) + \frac{\beta}{2}\|v-w\|^2$$

set $v = w - \frac{1}{\beta}\nabla f(w)$

$$f(v) \leq f(w) + \left\langle \nabla f(w), \frac{1}{\beta}\nabla f(w) \right\rangle - \frac{\beta}{2}\frac{1}{\beta^2}\|\nabla f\|^2$$

$$\Rightarrow \frac{1}{2\beta}\|\nabla f(w)\|^2 \leq f(w) - f(v)$$

if, in addition $f(v) \geq 0$ $\forall v$, then

$$\|\nabla f(w)\|^2 \leq 2\beta f(w) \quad \forall w$$

" self bounded " function