31 Jan 2022 — Convexity and Gradient Descent

Announcement:

TA office hours begin today.
See office hour calendar on website.

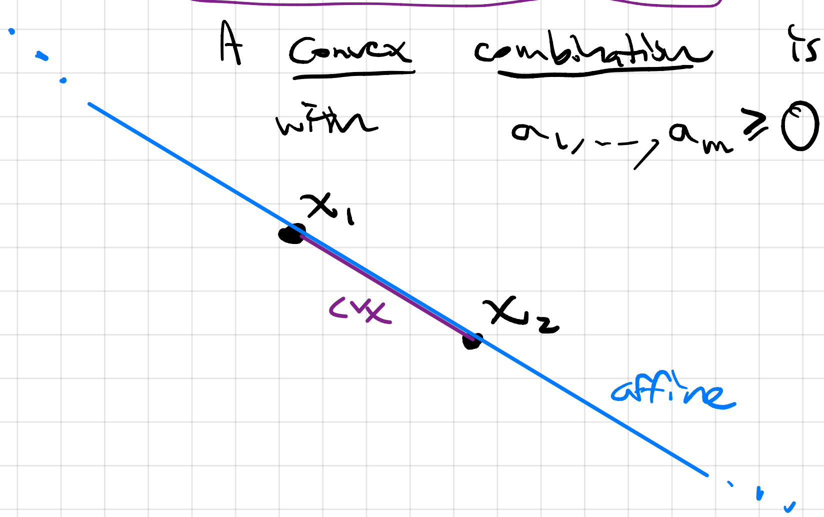
https://cs.cornell.edu/courses/cs4850/2022sp

Fill out OH modality poll. (Pinned post on Ed.)

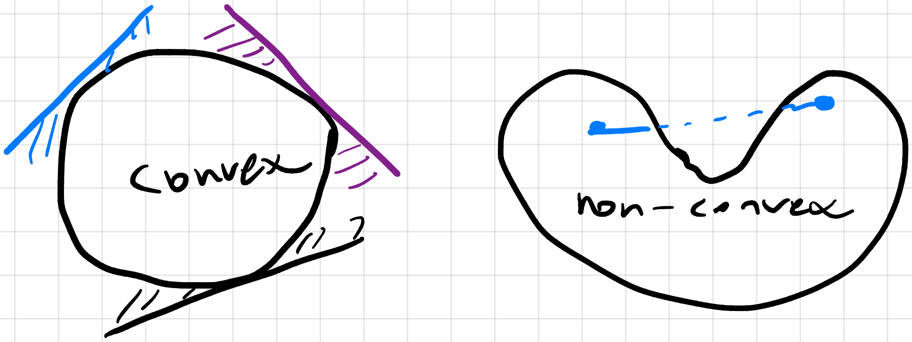
Write an Ed post requesting CMS access if you don't have it yet.

Def. If $x_1, \dots, x_m$ are vectors in vect spc V, an affine combination is any linear combination $a_1 x_1 + \cdots + a_m x_m$ such that $a_1 + a_2 + \cdots + a_m = 1$.

A convex combination is an affine combination with $a_1, \dots, a_m \geq 0$. (weighted average)



A subset of V is convex if it is closed under taking convex combinations.
(Suffices to just test that the line segment joining any 2 vectors in the set remains in the set.)



<u>Ex.</u> A halfspace in V is a set of the form

$$H = \{x \mid f(x) \le \theta\}$$

where $f \in V^*$, $f \neq 0$. Equivalently, if V has a non-degenerate inner product,

$$H = \{x \mid \langle w, x\rangle \le \theta\}$$

where $w \neq 0$.

<u>Pictures:</u> A halfspace in $\mathbb{R}^2$ looks like

<u>Prop.</u> In a finite dimensional vector space, a closed subset K is convex if and only if it is representable as an intersection of (potentially infinitely many) halfspaces.

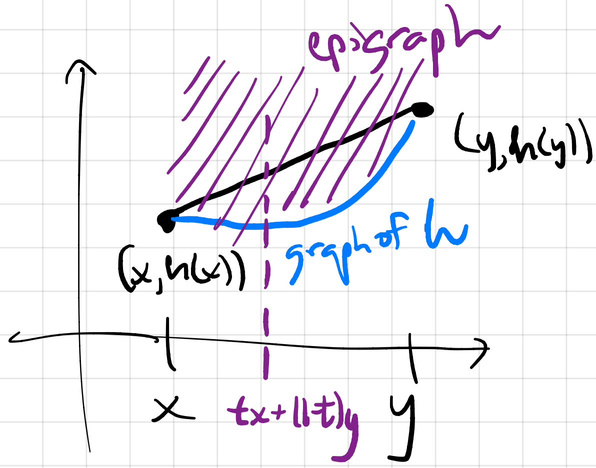
under taking limit points

($K \subseteq V$ convex)

<u>Def.</u> A function $h : K \to \mathbb{R}$ is <u>convex</u> if

$$h(tx + (1-t)y) \le t\,h(x) + (1-t)\,h(y) \quad \forall x, y \in K$$

h

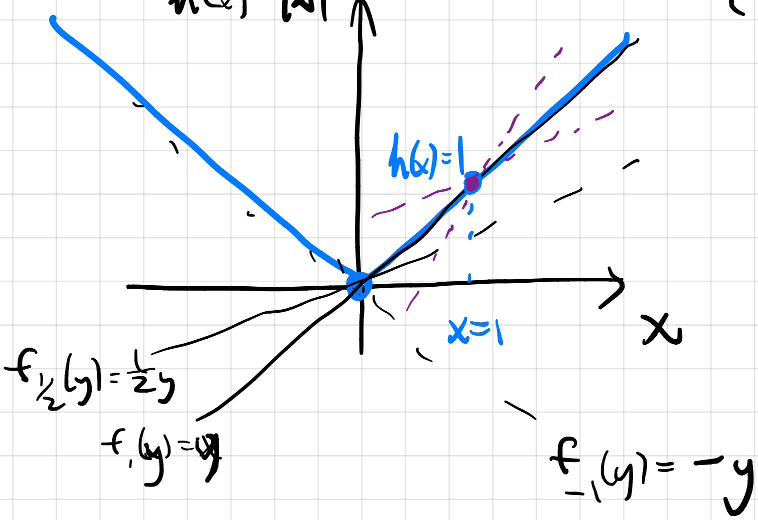


<u>Def.</u> For $K \subseteq V$ convex, $h : K \to \mathbb{R}$,

(1) The epigraph of h is $\left\{ (x,y) \in V \times \mathbb{R} \;\middle|\; \begin{array}{l} x \in K, \\ y \geq h(x) \end{array} \right\}$

(2) The subdifferential of h at $x \in K$ is the subset of V^* defined as

$$\partial h(x) = \{ f \in V^* \mid h(y) \geq h(x) + f(y-x) \quad \forall y \in K \}$$



$$\partial h(\vec{0}) = \{ f \in \mathbb{R}^* \mid \forall y \;\; |y| \geq f(y) \}$$

$$= \{ f(x) = ax \mid -1 \leq a \leq 1 \}.$$

$$\partial h(1) = \{ f \in \mathbb{R}^* \mid \forall y \;\; |y| \geq 1 + f(y-1) \}$$

$$= \{ f(x) = x \}.$$

$\partial h(x)$ = "set of slopes of 'supporting hyperplanes' to the graph of h at $(x, h(x))$."

Theorem. (Proved in lecture notes)
For $K \subseteq V$ convex and $h : K \to \mathbb{R}$
The following are equivalent.

(i) h is a convex function

(ii) The epigraph of h is a convex set.

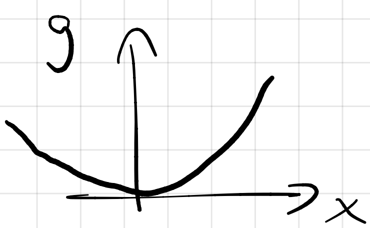
(iii) The subdifferential of h is nonempty at every point.

Differentiable Functions

Def. The norm of a vector in a space with pos. def. inner product is $\|x\| = \langle x, x \rangle^{1/2}$.

E.g. in $\mathbb{R}^n$ with standard inner prod,
$$\|x\| = (x_1^2 + \cdots + x_n^2)^{1/2} = \text{Euclidean length of } \vec{x}$$
$$= 2\text{-norm of } \vec{x}.$$

Def. A function $g : V \to \mathbb{R}$ vanishes to first order at $\vec{0}$ if $\forall \varepsilon > 0\ \exists \delta > 0$ s.t.
$$\frac{g(x)}{\|x\|} < \varepsilon \quad \text{whenever } \|x\| < \delta.$$



Def. $f: V \to \mathbb{R}$ is differentiable at x if there exists an element of V^* called the <u>differential</u> of f, denoted df_x, such that

$$\forall y \quad f(x+y) = f(x) + df_x(y) + g(y)$$

where g vanishes to 1st order at $\vec{0}$.

If h is convex and differentiable at x then

$$\partial h(x) = \{ dh_x \}.$$

For differentiable $f: \mathbb{R}^n \to \mathbb{R}$ the differential is

$$df_x = \left[\left.\frac{\partial f}{\partial x_1}\right|_x \quad \left.\frac{\partial f}{\partial x_2}\right|_x \quad \cdots \quad \left.\frac{\partial f}{\partial x_n}\right|_x \right].$$

The <u>gradient</u> of f is defined when V has a non-degenerate inner prod and f is diff'ble at x. Then ∇f_x is the image of df_x under the isomorphism $V^* \to V$.

Ex. $\mathbb{R}^n$ with standard inner prod.

$(\mathbb{R}^n)^*$ = row vectors length n

$\mathbb{R}^n$ = col vectors " "

isomorphism = transpose

$$\nabla f_x = \begin{bmatrix} \partial f / \partial x_1 \\ \vdots \\ \partial f / \partial x_n \end{bmatrix}$$