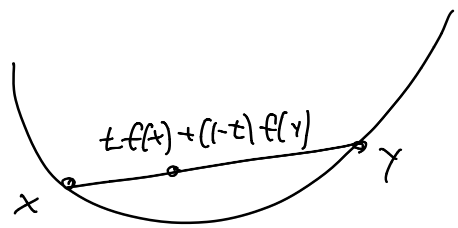
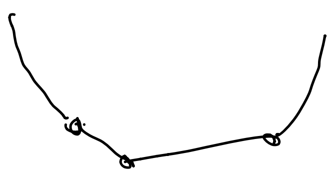
Convex Functions (in one var is $f''(x) \geq 0$)

A function $f: \mathbb{R}^n \to \mathbb{R}$ is <u>convex</u> if for all x and y and all $0 \leq t \leq 1$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \qquad (*)$$

These look like

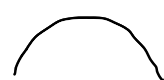


Or  - linear is convex.

f is <u>strictly</u> convex if

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y)$$

$\|x\|_2$ is strictly convex, but $\|x\|_\infty$ and $\|x\|_1$ are merely convex.

f is <u>concave</u> if $-f$ is convex 

I think of a cave.

dom(f), the domain of f, is the set of x for which f is defined.

If $\mathrm{dom}(f) \neq \mathbb{R}^n$, then f is convex if

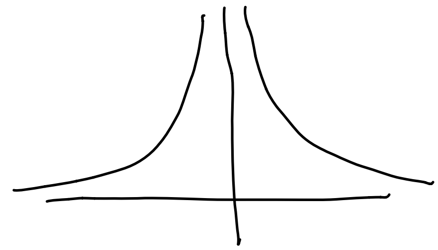
a. dom(f) is a convex set, and

b. $(*)$ holds for all $x, y \in \mathrm{dom}(f)$

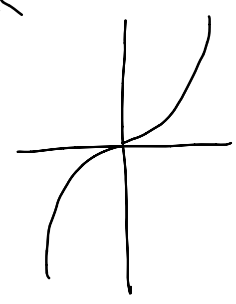
Example: $f(x) = 1/x^2$

is convex if $\mathrm{dom}(f) = (0, \infty]$,

but not on $\mathbb{R}$



Same for x^3



If f is convex but $\mathrm{dom}(f) \neq \mathbb{R}^n$,

the extension of f, $\tilde{f}$, is convex on $\mathbb{R}^n$, where

$$\tilde{f}(x) = \begin{cases} f(x) & \text{for } x \in \mathrm{dom}(f) \\ \infty & \text{for } x \notin \mathrm{dom}(f) \end{cases}$$

Why? We care about convex optimization problems:

$$\min f(x) \quad \text{s.t. } x \in C$$

where C is a convex set and f is a convex fuction.

This is one of the broadest classes of problems that we can solve efficiently.

If f is strictly convex, the minimum is unique:

if $x_1, x_2 \in \arg\min f(x)$, but $x_1 \neq x_2$,

then $f(\frac{1}{2}x_1 + \frac{1}{2}x_2) < \frac{1}{2} f(x_1) + f(x_2)$,

contradicting minimality of x_1, x_2.

Local minima of convex functions are global minima.

x is a <u>local minimum</u> of f if $\exists \varepsilon > 0$ s.t.

$\|x-y\| \le \varepsilon \Rightarrow f(x) \le f(y)$.

x is a <u>global minimum</u> if $\forall y \; f(x) \le f(y)$.

If x were a local minimum but NOT global, then $\exists y$ s.t. $f(y) < f(x)$.

For every $\varepsilon > 0$ there is a $t \in [0,1]$ s.t.

$\|x - ((1-t)x + ty)\| \le \varepsilon$.

For this t, $f((1-t)x+ty) \le (1-t)f(x) + tf(y) < f(x)$.

This contradicts the assertion that x is a local minimum.

Examples of convex functions.

One variable:

e^{ax} for $a \in \mathbb{R}$

x^α $\alpha \ge 1$ or $\alpha \le 0$, $x \in [0,\infty]$

$-x^\alpha$ $0 \le \alpha \le 1$, $x \in [0,\infty]$

$-\log(x)$, $x \in [0,\infty]$, because $\log(x)$ is concave

Vectors

norms: $\|x\|$

Indicator functions: $I(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C, \end{cases}$

C is convex. (could replace ∞ with 1)

Affine functions : $f(x) = a^T x + b$

Rules:

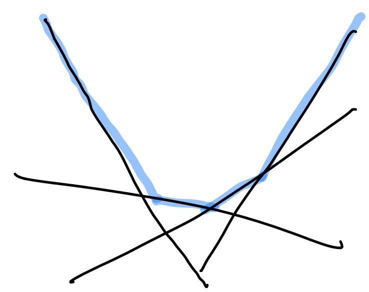
Non-negative sums:

$f_1, \dots, f_n$ convex, $w_1, \dots, w_n \geq 0 \Rightarrow \sum w_i f_i$ is convex

Maximum: the maximum of convex is convex

$$f(x) = \max\left(f_1(x), f_2(x)\right)$$

Ex. maximum of linear is convex



Ex. $\max\left(x(1), x(2), \dots, x(n)\right)$

Ex. Sum of largest k components

$$= \max_{|S|=k} \sum_{i \in S} x(i)$$

Affine composition:

if g is convex, then so is $f(x) = g(Ax+b)$

Examples: $\| \ \|_2$ is convex $\Rightarrow$ $\|Ax\|_2$ convex

$\Rightarrow$ $x^T(A^TA)x$ is convex.

Every Positive Semidefinite $Q = A^TA$, for some A.

So, x^TQx is convex.

Least squares: $\|Ax-b\|_2^2$ is convex

l_1-regularized: $\|Ax-b\|_2 + \lambda\|x\|_1$ is convex

Logistic loss: $\sum_i \log(1+e^{y_i(a_i^Tx+b)}) = f(x)$

$g(z) = \log(1+e^{yz})$ is convex in g, because

$$g''(z) = \frac{y^2e^{yz}}{(1+e^{yz})^2}$$

a_i^Tx+b is affine.

Composition rules (one of many)

If $f(x) = h(g(x))$ where $h: \mathbb{R}\to\mathbb{R}$, $g: \mathbb{R}^n\to\mathbb{R}$,

g convex & h convex, non-decreasing $\Rightarrow$ f convex

idea: in one var, $f'(x) = g'(x)\, h'(g(x))$

$$f''(x) = \underset{\geq 0}{g''(x)}\, \underset{\geq 0}{h'(g(x))} + \underset{\geq 0}{g'(x)^2}\, \underset{\geq 0}{h''(g(x))}$$

Functions & Sets:

The α-sublevel set of f is $\{x : f(x) \le \alpha\}$.

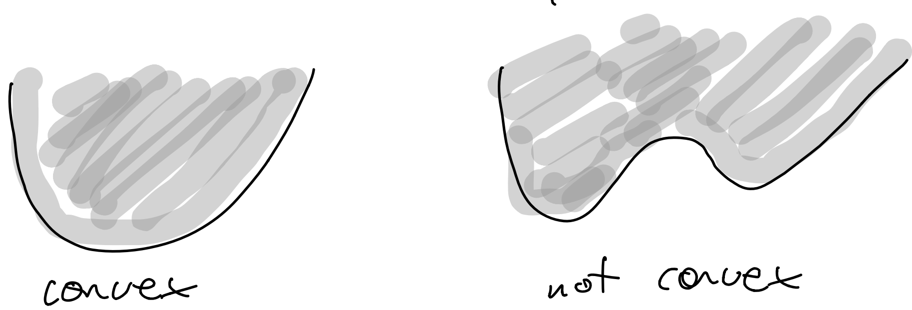
If f is convex then so are its α-sublevel sets.

The <u>graph</u> of a function f is $\{(x, f(x)) : x \in dom(f)\}$.

The <u>epigraph</u> of f is the region above:

$$\{(x,t) : x \in dom(f),\ f(x) \le t\}$$

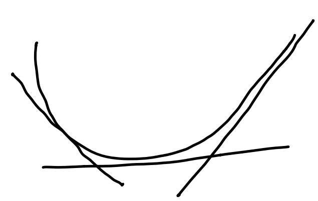
like the ice cream cone, or:



f is a convex function iff its epigraph is a convex set.

In one variable, a twice-differentiable function $f(x)$ is convex iff $f''(x) \ge 0,\ \forall x \in dom(f)$.

This says that the function always lies above tangent lines, like



Thm 1

In $\mathbb{R}^n$, if f is differentiable then f is convex iff

$\forall x, y \in dom(f), \quad f(y) \geq f(x) + \nabla f(x)^T (y-x) \quad (*)$

Note that $h(y) \triangleq f(x) + \nabla f(x)^T (y-x)$ is the linear function in y such that $h(x) = f(x)$,

and is a supporting hyperplane of epigraph at $(x, f(x))$

proof of Thm 1

We first prove $(*) \Rightarrow$ convex

Let $x, y \in dom(f)$, $0 < \lambda < 1$, $z = \lambda x + (1-\lambda) y$.

Let $g = \nabla f(z)$.

$$(*) \Rightarrow \quad \lambda f(x) \geq \lambda f(z) + \lambda g^T (x-z)$$

$$(1-\lambda) f(y) \geq (1-\lambda) f(z) + (1-\lambda) g^T (y-z)$$

$$\Rightarrow \lambda f(x) + (1-\lambda) f(y) \geq f(z) + \underbrace{g^T(\lambda x - \lambda z + (1-\lambda) y - (1-\lambda) z)}_{0}$$

convex $\Rightarrow (*)$

For $x, y \in dom(f)$, $0 < \lambda < 1$,

$$(1-\lambda) f(x) + \lambda f(y) \geq f((1-\lambda) x + \lambda y) = f(x + \lambda (y-x))$$

$$\Rightarrow f(y) \geq f(x) + \frac{f(x + \lambda(y-x)) - f(x)}{\lambda}$$

taking lim as $\lambda \to 0$ gives

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

Thm 2 If f is twice differentiable,

f is convex iff $\nabla^2 f(x)$ is psd
where $\nabla^2 f(x)$ is matrix with entries $\frac{\partial^2}{\partial x(i)\, \partial x(j)} f$

Computing $\|A\|_2^2 = \max_{\|x\|=1} \|Ax\|_2^2 = \max_{\|x\|=1} x^T (A^T A) x$

Does not seem a convex program, because
i. are maximizing
ii. $\|x\|=1$ is not a convex set.
Restricting to $\|x\| \le 1$ solves ii but not i.

Solution: write as
$\min t$ s.t. $tI - A \succeq 0$

($M \succeq 0$ iff M is positive semidefinite).
Is a convex cone.

Can check if $M \succeq 0$ by trying to compute
a Cholesky factorization: L s.t. $LL^T = M$.

S_n^+ = set of symmetric $n \times n$ positive semidefinite.

If A is symmetric, but $A \notin S^n_+$,

$\exists v$ s.t. $v^T A v < 0$

in space of symmetric matrices

A hyperplane separating A from S^n_+ is given by

$\{ \text{symmetric } X : v^T X v = 0 \}$

$v^T X v \geq 0$ for $X \in S^n_+$. $v^T A v < 0$.

And is a hyperplane because

$$v^T X v = \sum_{1 \leq i,j \leq n} X(i,j)\, v(i)\, v(j) \quad \text{is linear in } X.$$