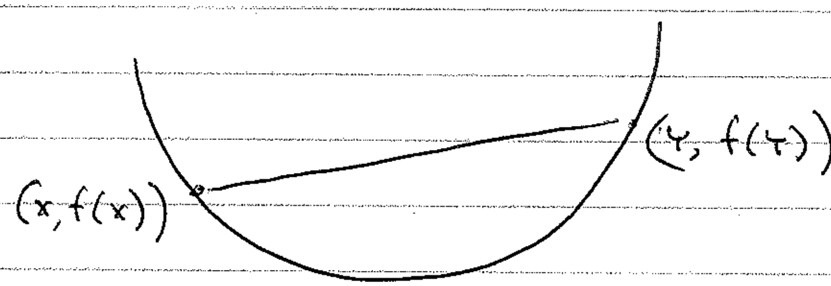
A function $f: \mathbb{R}^n \to \mathbb{R}$ is convex if dom f is a convex set and if for all $x, y \in \mathrm{dom}\, f$, and with $0 \le \theta \le 1$ we have

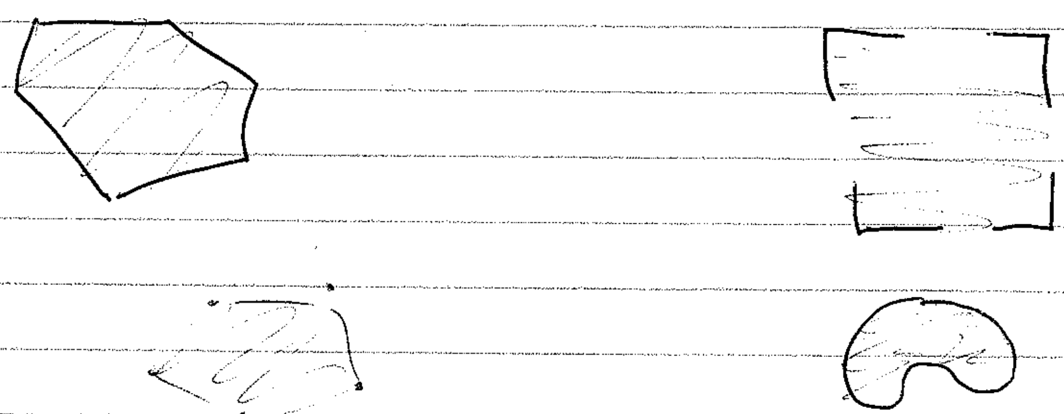
$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$$



What is a convex set? A set is convex if the line segment between any two points in C lies in C. i.e., if for any $x_1, x_2 \in C$ and any θ with $0 \le \theta \le 1$ we have

$$\theta x_1 + (1-\theta) x_2 \in C$$

Example convex sets non-convex sets



The convex hull of a set C, denoted $\mathbf{conv}\, C$, is the set of all convex combinations of points in C.

$$\mathbf{conv}\, C = \{\theta_1 x_1 + \cdots \theta_k x_k \mid x_i \in C,\ \theta_i \geq 0,\ i = 1, \ldots, k,\ \theta_1 + \cdots + \theta_k = 1\}$$

The convex hull of a set of points is always convex. It is the set of all convex combinations of points in C

An example convex set is the hyperplane

$$\{x \mid a^T x = b\} \qquad a \in \mathbb{R}^n,\ a \neq 0,\ b \in \mathbb{R}$$

Analytically this is the set of all nontrivial linear equation solutions among the components of x.

Geometrically this set can be interpretted as a hyperplane with normal vector a; the constant $b \in \mathbb{R}$ determines the offset of the hyperplane from the origin. This can be seen as

$$\{x \mid a^T (x - x_0) = 0\}$$

with x_0 any point that satisfies $a^T x_0 = b$

Operations that preserve the convexity of a set include:

Intersection – if S_1 + S_2 are convex so is $S_1 \cap S_2$

Affine functions – an affine function is one that involves a linear function and a constant.

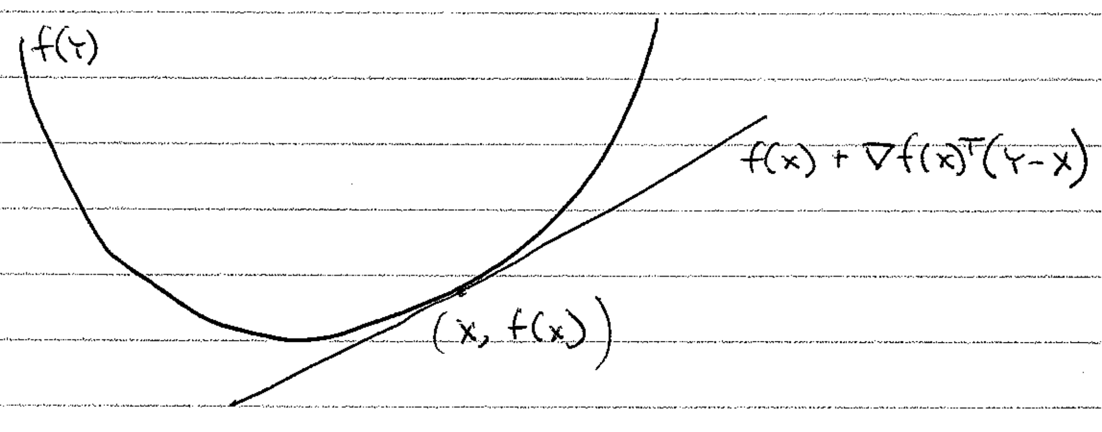
<u>Back to Convex functions:</u>

First order conditions for a convex function:

Suppose f is differentiable (ie. is gradient ∇f exists at each point in $\underline{\text{dom}}\ f$) Then f is convex if and only if $\underline{\text{dom}}\ f$ is convex and

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

holds for all $x, y \in \underline{\text{dom}}\ f$.



This function (affine function of y) is the first order Taylor approximation of f near x.

This is a global underestimator given local information.

Also, when $\nabla f(x) = 0$ then $\forall y \in \mathbf{dom}\, f$ $f(y) \geq f(x)$, i.e. x is a global minimizer of the

Proof of first-order convexity condition

First consider $n = 1$ (dimensionality). We show that $f: \mathbb{R} \to \mathbb{R}$ is convex if and only if

$$f(y) \geq f(x) + f'(x)(y-x) \quad \forall x, y \in \mathbf{dom}\, f$$

Assume f is convex and $x, y \in \mathbf{dom}\, f$. Since $\mathbf{dom}\, f$ is convex then $\forall\ 0 < t \leq 1$, $x + t(y-x) \in \mathbf{dom}\, f$. Then by convexity of f

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

If we divide through by t we get

$$f(y) \geq \frac{f(x + t(y-x)) - (1-t) f(x)}{t}$$

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

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

which, in the limit of $t \to 0$ yields

$$f(y) \geq f(x) + f'(x)(y-x). \qquad (1)$$

To show sufficiency, assume the function satisfies (1) $\forall\, x, y \in \underline{\text{dom}}\, f$ (an interval). Choose $x \neq y$ and $0 \leq \theta \leq 1$ and let $z = \theta x + (1-\theta) y$. Applying (1) twice yields.

$$f(x) \geq f(z) + f'(z)(x-z), \quad f(y) \geq f(z) + f'(z)(y-z)$$

mult. by θ and $1-\theta$

$\cancel{\theta f(x) \geq \theta f(z) + \theta f'(z)(x-z)}$ won't do alg.

yields

$$\theta f(x) + (1-\theta) f(y) \geq f(z)$$

which proves that f is convex.

Now we prove the general case with $f: \mathbb{R}^n \to \mathbb{R}$. Let $x, y \in \mathbb{R}^n$ and consider f restricted to the line passing through them. i.e. the function defined by

$$g(t) = f(ty + (1-t)x)$$

so

$$g'(t) = \nabla f(ty + (1-t)x)(y-x)$$

First assume f is convex, this implies g is convex. So by

$$f(y) \geq f(x) + f'(x)(y-x)$$ for general convex function f

$$g(1) \geq g(0) + g'(0)$$ $y=1$, $x=0$

which means

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

The other direction (ie. if $f(y) \geq f(x) + \nabla f(x)^T (y-x) \Rightarrow$ convexity) is left as an excercise (Boyd & VB pg 70).

2nd order condition

If f is twice differentiable then is 2nd derivative $\nabla^2 f$ exists at each point in $\mathbf{dom}\, f$. f is convex if and only if $\mathbf{dom}\, f$ is convex and its Hessian is positive semi definite: for all $x \in \mathbf{dom}\, f$

$$\nabla^2 f(x) \succeq 0$$

An important example (arises in log of multivariate normal for example) is the quadratic function $f: \mathbb{R}^n \to \mathbb{R}$ with $\underline{\mathrm{dom}}\ f = \mathbb{R}^n$, given by

$$f(x) = \tfrac{1}{2} x^T P x + q^T x + r$$

with $P \in S^n$, $q \in \mathbb{R}^n$ and $r \in \mathbb{R}$. Since $\nabla^2 f(x) = P$ $\forall x$ then f is convex iff $P \succeq 0$

Interesting convex functions: on $\mathbb{R}$

$\cancel{e^{ax}}$, $\cancel{x^a}$, $\cancel{|x|^p}$

e^{ax} on $\mathbb{R}$, $a \in \mathbb{R}$

x^a on $\mathbb{R}_{++}$ $a \geq 1$, $a \leq 0$

$\log x$ on $\mathbb{R}_{++}$

interesting functions on $\mathbb{R}^n$

Norms: Every norm on $\mathbb{R}^n$ is convex

Max: Every max function is convex on $\mathbb{R}^n$

$$f(x) = \max\{x_1, \ldots, x_n\}$$

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$$

Proof for max function

The function $f(x) = \max_i x_i$ satisfies, for $0 \le \theta \le 1$

$$f(\theta x + (1-\theta) y) = \max_i (\theta x_i + (1-\theta) y_i)$$

$$\le \theta \max_i x_i + (1-\theta) \max_i y_i$$

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

later

Consider regression case, we have

$$Q(\vec{b}) = (y - b^T x)^T (y - b^T x)$$

$$= \sum_i (y_i - (b_1 x_i + b_0))^2$$

Jensen's inequality

The basic inequality

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

is sometimes called Jensen's inequality.

It can be extended to convex combinations of more than two points. If f is convex, $x_1, \dots, x_k \in \mathrm{dom}\, f$, and $\theta_1, \dots, \theta_k \ge 0$ with $\theta_1 + \dots + \theta_k = 1$ then

$$f(\theta_1 x_1 + \dots + \theta_k x_k) \le \theta_1 f(x_1) + \dots + \theta_k f(x_k)$$

As in the case convex sets the inequality extends to infinite sums, integrals, and Expected values;

For example, if $p(x) \geq 0$ on $S \subseteq \mathrm{dom}\, f$, and $\int_S p(x)\, dx = 1$ (i.e. think of p as a probability measure) then

$$f\left(\int_S p(x)\, x\, dx \right) \leq \int_S f(x)\, p(x)\, dx$$

Digression

For instance we have

$$f(\mathbb{E}(x)) \leq \mathbb{E}(f(x))$$

This is super important because we often want to bound

$$\log(\mathbb{E}(x))$$

$$\begin{bmatrix} Y \end{bmatrix} \quad \begin{bmatrix} & X & \end{bmatrix} \begin{bmatrix} b^T \end{bmatrix}$$

Now to regression with

$$\begin{aligned}
Q(b) &= (Y - Xb^T)^T (Y - Xb^T) \\
&= Y^TY - Y^TXb^T - (Xb^T)^TY + (Xb^T)^T Xb^T \\
&= Y^TY - Y^TXb^T - (((Xb^T)^TY)^T)^T + bX^TXb^T \\
&= Y^TY - Y^TXb^T - Y^TXb^T + bX^TXb^T \\
&= Y^TY - 2Y^TXb^T + bX^TXb^T
\end{aligned}$$

Quadratic form in b

$$\nabla Q(b) = -2Y^TX + (X^TX + X^TX)b$$

$$\nabla^2 Q(b) = 2X^TX$$

$$\text{if } 2X^TX \succeq 0 \Rightarrow Q(b) \text{ convex.}$$

Alternatively one can ~~prove~~ that

$$\cancel{Q(b_0) \geq Q(b_0') + (b_0 - b_0') \cdot \nabla}$$

Gradient Descent

if $F(x)$ is defined and differentiable in a neighborhood of a point a, the $F(x)$ decreases fastest if one goes from a in the direct of the negative gradient of F at a.

Assume function is locally convex then for $a, b \in$ convex domain of F then

$$F(b) \geq F(a) + (b-a)\nabla F(a)$$

reparameterizing $v = b-a$, $b = a+v$

$$F(a+v) \geq F(a) + v \nabla F(a)$$

which makes it clear that we should move in the opposite direction of $\nabla F(a)$ if we want to minimize $F(a+v)$.

~~Line Search~~

~~Evaluate~~ $\cancel{F(a) - v \nabla F(a)}$

First order Taylor approximation"

$$f(x+v) \approx \hat{f}(x+v) = f(x) + \nabla f(x)^T v$$

From convexity we know that

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

So the search direction must satisfy

$$\nabla f(x)^T \Delta x < 0$$