NYU | COURANT INSTITUTE OF MATHEMATICAL SCIENCES

Convex optimization

Optimization-Based Data Analysis

http://www.cims.nyu.edu/~cfgranda/pages/OBDA_spring16

Carlos Fernandez-Granda

2/1/2016

Convex sets

A convex set $\mathcal{S}$ is any set such that for any $x, y \in \mathcal{S}$ and $\theta \in (0,1)$

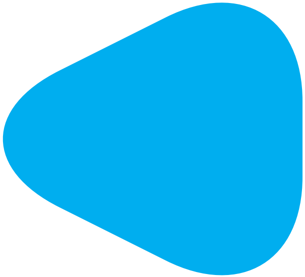
$$\theta x + (1-\theta)\, y \in \mathcal{S}$$

The intersection of convex sets is convex

Convex vs nonconvex

Nonconvex

Convex



Projection onto convex set

The projection of any vector x onto a non-empty closed convex set $\mathcal{S}$

$$\mathcal{P}_{\mathcal{S}}(x) := \arg\min_{s \in \mathcal{S}} ||x - s||_2$$

exists and is unique

Convex combination

Given n vectors $x_1, x_2, \ldots, x_n \in \mathbb{R}^n$,

$$x := \sum_{i=1}^{n} \theta_i x_i$$

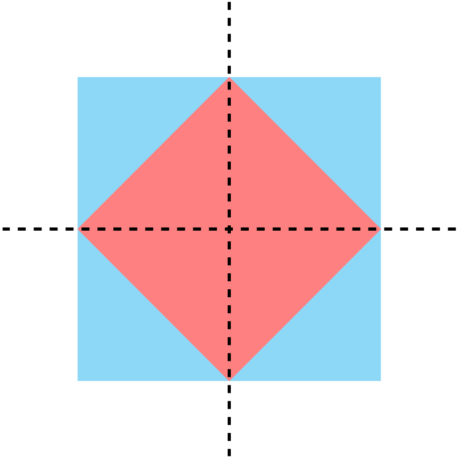
is a convex combination of $x_1, x_2, \ldots, x_n$ if

$$\theta_i \geq 0, \quad 1 \leq i \leq n$$
$$\sum_{i=1}^{n} \theta_i = 1$$

Convex hull

The convex hull of $\mathcal{S}$ is the set of convex combinations of points in $\mathcal{S}$

ℓ_1-norm ball



Convex function

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if for any $x, y \in \mathbb{R}^n$ and any $\theta \in (0, 1)$,

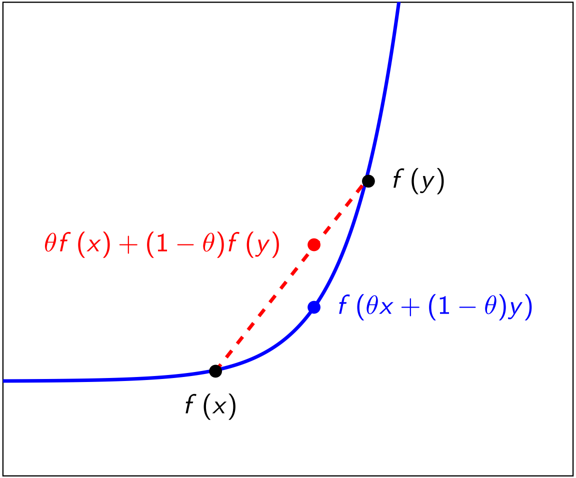
$$\theta f(x) + (1 - \theta) f(y) \geq f(\theta x + (1 - \theta) y)$$

The function is strictly convex if $x \neq y$ implies that

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

A concave function is a function f such that $-f$ is convex

Convex function



Two important properties

Local minima of convex functions are global minima

Sublevel functions $\{x \mid f(x) \leq \gamma, \ \gamma \in \mathbb{R}\}$ of convex function are convex

Norm

Function $||\cdot||$ from a vector space $\mathcal{V}$ to $\mathbb{R}$ that satisfies:

- For all $\alpha \in \mathbb{R}$ and $x \in \mathcal{V}$

$$||\alpha\, x|| = |\alpha|\; ||x||$$

- Triangle inequality

$$||x + y|| \leq ||x|| + ||y||$$

- $||x|| = 0$ implies that x is the zero vector 0

Norms

$$||x||_2 := \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

$$||x||_1 := \sum_{i=1}^{n} |x_i|$$

$$||x||_\infty := \max_{1 \le i \le n} |x_i|$$

ℓ_0 "norm"

The ℓ_0 "norm" is not convex and is not a norm

Equivalent definition of convex functions

$f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for any $x, y \in \mathbb{R}^n$ the 1D function

$$g_{x,y}(\alpha) := f(\alpha x + (1 - \alpha) y)$$

is convex

Epigraph

The graph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the curve in $\mathbb{R}^{n+1}$

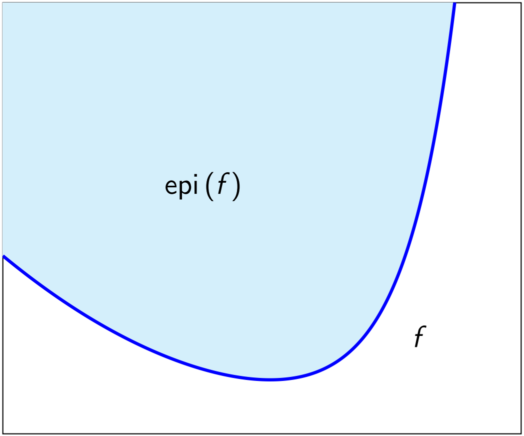
$$\text{graph}\,(f) := \{x \mid f\,(x_{1:n}) = x_{n+1}\}$$

The epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\text{epi}\,(f) := \{x \mid f\,(x_{1:n}) \leq x_{n+1}\}$$

It is a subset of $\mathbb{R}^{n+1}$

Epigraph



A function is convex if and only if its epigraph is convex

Composition of convex and affine function

If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then for any $A \in \mathbb{R}^{n \times m}$ and any $b \in \mathbb{R}^n$

$$h(x) := f(Ax + b)$$

is convex

Consequence: The least-squares cost function

$$||Ax - y||_2$$

is convex

Nonnegative weighted sums

Nonnegative weighted sums of convex functions

$$f := \sum_{i=1}^{m} \alpha_i f_i$$

are convex

Consequence: Regularized least-squares cost functions of the form

$$||Ax - y||_2^2 + ||x|| \,,$$

where $||\cdot||$ is an arbitrary norm, are convex

Pointwise maximum/supremum

The pointwise maximum of m convex functions $f_1, \ldots, f_m$

$$f(x) := \max_{1 \leq i \leq m} f_i(x)$$

is convex

The pointwise supremum of a family of convex functions f_i, $i \in \mathcal{I}$,

$$f(x) := \sup_{i \in \mathcal{I}} f_i(x)$$

is convex

First-order condition

A differentiable function f is convex if and only if for every $x, y \in \mathbb{R}^n$

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

It is strictly convex if and only if

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

Interpretation: First-order Taylor expansion

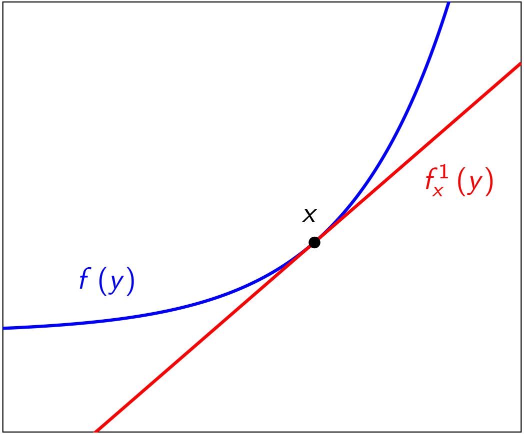
$$f_x^1(y) := f(x) + \nabla f(x)(y - x)$$

is a lower bound on f

Optimality conditions

- If $\nabla f(x) = 0$, x is a global minimum
- If in addition f is strictly convex x is the only minimum

First-order condition



Supporting hyperplane

A hyperplane $\mathcal{H}$ is a supporting hyperplane of a set $\mathcal{S}$ at x if

- $\mathcal{H}$ and $\mathcal{S}$ intersect at x
- $\mathcal{S}$ is contained in one of the half-spaces bounded by $\mathcal{H}$

The hyperplane $\mathcal{H}_{f,x} \subset \mathbb{R}^{n+1}$

$$\mathcal{H}_{f,x} := \left\{ y \mid y_{n+1} = f(x) + \nabla f(x)^T (y_{1:n} - x) \right\}$$

is a supporting hyperplane of $\text{epi}(f)$ at x

Second-order condition in 1D

A twice-differentiable function $g : \mathbb{R} \to \mathbb{R}$ is convex if and only if

$$g''(\alpha) \geq 0 \quad \text{for all } \alpha \in \mathbb{R}$$

Second-order condition

A twice-differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \mathbb{R}^n$$

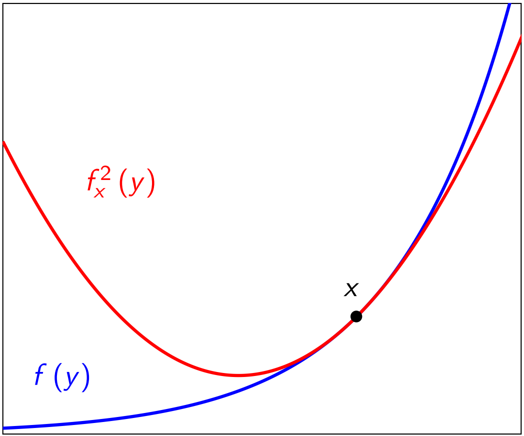
Intuition: Second derivative is nonnegative in every direction

The second-order or quadratic approximation of f

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

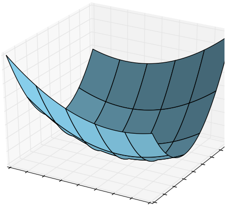
is convex everywhere

Second-order condition

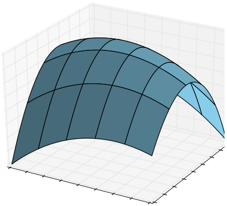


Quadratic forms

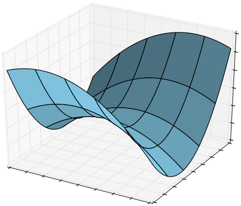
Convex



Concave



Neither



Subgradients

The subgradient of $f : \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$ is a vector $q \in \mathbb{R}^n$ such that

$$f(y) \geq f(x) + q^T(y - x), \quad \text{for all } y \in \mathbb{R}^n.$$

The set of all subgradients is the subdifferential of the function at x

Interpretation: Gradient of supporting hyperplane of f at x

First-order condition for nondifferentiable functions

If f has a non-empty subdifferential everywhere then f is convex

Optimality condition for nondifferentiable functions

If the zero vector is a subgradient of f at x then x minimizes f

Subdifferential of ℓ_1 norm

$q \in \mathbb{R}^n$ is a subgradient of the ℓ_1 norm at $x \in \mathbb{R}^n$ if

$$\begin{aligned} q_i &= \operatorname{sign}(x_i) && \text{if } x_i \neq 0, \\ |q_i| &\leq 1 && \text{if } x_i = 0 \end{aligned}$$

Canonical optimization problem

$$f_0, f_1, \ldots, f_m, h_1, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$$

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad 1 \leq i \leq m, \\ & h_i(x) = 0, \quad 1 \leq i \leq p, \end{aligned}$$

Definitions

- A feasible vector is a vector that satisfies all the constraints
- A solution is any vector x^* such that for all feasible vectors x

$$f_0(x) \geq f_0(x^*)$$

- If a solution exists $f(x^*)$ is the optimal value or optimum of the problem

Convex optimization problem

The optimization problem is convex if

- f_0 is convex
- $f_1, \ldots, f_m$ are convex
- $h_1, \ldots, h_p$ are affine, i.e. $h_i(x) = a_i^T x + b_i$ for some $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$

Linear program

$$\begin{aligned} \text{minimize} \quad & a^T x \\ \text{subject to} \quad & c_i^T x \le d_i, \quad 1 \le i \le m, \\ & Ax = b \end{aligned}$$

ℓ_1-norm minimization as an LP

The optimization problem

$$\begin{aligned} \text{minimize} \quad & ||x||_1 \\ \text{subject to} \quad & Ax = b \end{aligned}$$

can be recast as the LP

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{n} t_i \\ \text{subject to} \quad & t_i \geq x_i, \\ & t_i \geq -x_i, \\ & Ax = b \end{aligned}$$

Quadratic program

For a positive semidefinite matrix $Q \in \mathbb{R}^{n \times n}$

$$\begin{aligned} \text{minimize} \quad & x^T Q x + a^T x \\ \text{subject to} \quad & c_i^T x \leq d_i, \quad 1 \leq i \leq m, \\ & Ax = b \end{aligned}$$

ℓ_1-norm regularized least squares as a QP

The optimization problem

$$\text{minimize} \quad ||Ax - y||_2^2 + \lambda \, ||x||_1$$

can be recast as the QP

$$\begin{aligned} \text{minimize} \quad & x^T A^T A x - 2y^T x + \lambda \sum_{i=1}^{n} t_i \\ \text{subject to} \quad & t_i \geq x_i, \\ & t_i \geq -x_i \end{aligned}$$

Lagrangian

The Lagrangian of a canonical optimization problem is

$$L(x, \lambda, \nu) := f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

$\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$ are called Lagrange multipliers or dual variables

If x is feasible and $\lambda_i \geq 0$ for $1 \leq i \leq m$

$$L(x, \lambda, \nu) \leq f_0(x)$$

Lagrange dual function

The Lagrange dual function of the problem is

$$l(\lambda, \nu) := \inf_{x \in \mathbb{R}^n} f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

Let p^* be an optimum of the optimization problem

$$l(\lambda, \nu) \leq p^*,$$

as long as $\lambda_i \geq 0$ for $1 \leq i \leq n$

Dual problem

The dual problem of the (primal) optimization problem is

$$\begin{aligned} \text{maximize} \quad & l(\lambda, \nu) \\ \text{subject to} \quad & \lambda_i \geq 0, \quad 1 \leq i \leq m. \end{aligned}$$

The dual problem is always convex, even if the primal isn't

Weak duality

If p^* is a primal optimum and d^* a dual optimum

$$d^* \leq p^*$$

Strong duality

For convex problems

$$d^* = p^*$$

under very weak conditions

LPs: The primal optimum is finite

General convex programs (Slater's condition):

There exists a point that is strictly feasible

$$f_i(x) < 0 \quad 1 \leq i \leq m$$