Convexity: Sets and Functions

Lecturer: Jianyong Sun
jy.sun@xjtu.edu.cn

School of Mathematics and Statistics
Xi'an Jiaotong University

Prerequisites: no formal ones, but class will be fairly fast paced.

Assume working knowledge of/proficiency with

- Real Analysis / Matrix theory
- Core problems in Stats/ML
- Programming (Matlab, Python, R, $\cdots$)
- Data structures, computational complexity
- Formal mathematical thinking

Supplementary Books

- Boyd and Vandenberghe, Convex Optimization, 2009
- R.T. Rochafellar, Convex Analysis, 1996
- D.P. Bertsekas, Convex Optimization Theory, 2009

Optimization problems are ubiquitous in Statistics and Machine Learning

Optimization problems underlie most everything we do in Statistics and Machine Learning.

Why convexity? Simply put: because we can broadly understand and solve convex optimization problems, and Nonconvex problems are mostly treated on a case by case basis.

Presumably, other people have already figured out how to solve

$$P : \min_{x \in D} f(x)$$

Why bother?

- Different algorithms can perform better or worse for different problems P (sometimes drastically so)
- Studying P can actually give you a deep understanding of the statistical procedure in question.

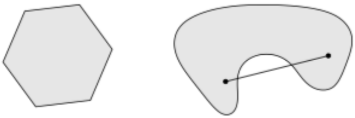
Outline

- Convex sets
- Examples
- Key properties
- Operations preserving convexity
- Same for convex functions.

Convex Sets and Functions

Convex set: $C \subseteq \mathbb{R}^n$ such that

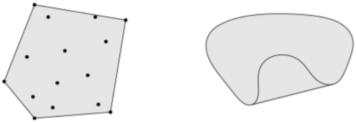
$$x, y \in C \Longrightarrow tx + (1-t)y \in C \text{ for all } 0 \le t \le 1$$



Convex combination of $x_1, \cdots, x_k \in \mathbb{R}^n$: any linear combination

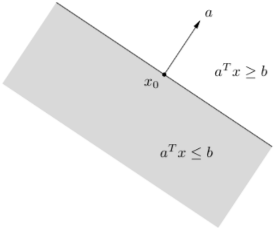
$$\theta_1 x_1 + \cdots + \theta_k x_k$$

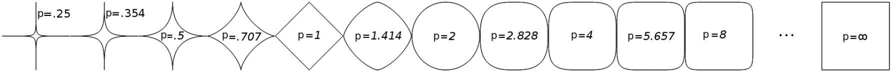
with $\theta_i \geq 0, i = 1, \cdots, k$ and $\sum_i \theta_i = 1$, where Convex hull of a set C, conv(C) is all convex combinations of elements. Always convex!



Examples of convex sets

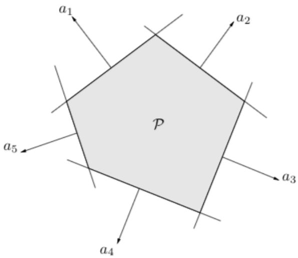
- Trivial-ones: empty set, point, line
- Norm ball: $\{x : \|x\| \leq r\}$, for given norm $\|\cdot\|$, radius r
- Hyperplane: $\{x : a^T x = b\}$ for given a, b
- Halfspace: $\{x : a^T x \leq b\}$
- Affine space $\{x : A^T x = b\}$





Examples of convex sets

- Polyhedron: $\{x : Ax \leq b\}$, where inequality $\leq$ is interpreted component-wise. Note: the set $\{x : Ax \leq b, Cx = d\}$ is also a polyhedron (why?)



- Simplex: special case of polyhedra, given by conv$\{x_0, \cdots, x_k\}$, where these points are affinely independent. The canonical example is the probability simplex

$$\text{conv}\{e_1, \cdots, e_n\} = \{w : w \geq 0, 1^T w = 1\}$$

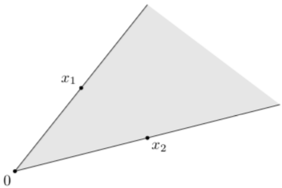
Cones

Cone: $C \subseteq \mathbb{R}^n$ such that

$$x \in C \Rightarrow tx \in C \text{ for all } \; t \geq 0$$

Convex cone: cone that is also convex, i.e.

$$x_1, x_2 \in C \Rightarrow t_1 x_1 + t_2 x_2 \in C \text{ for all } \; t_1, t_2 \geq 0$$



Conic combination of $x_1, \cdots, x_k \in \mathbb{R}^n$: any linear combination

$$\theta_1 x_1 + \cdots + \theta_k x_k$$

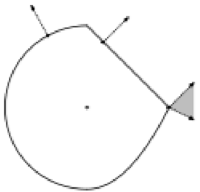
with $\theta_i \geq 0, i = 1, \cdots, k$. Conic hull collects all conic combinations

Examples of convex cones

- Norm cone: $\{(x,t) : \|x\| \le t\}$, for a norm $\|\cdot\|$. Under ℓ_2-norm $\|\cdot\|_2$, called second-order cone
- Normal cone: Given any set C and a point $x \in C$, we can define

$$\mathcal{N}_C(x) = \{g : g^T x \ge g^T y, \text{ for all } y \in C\}$$

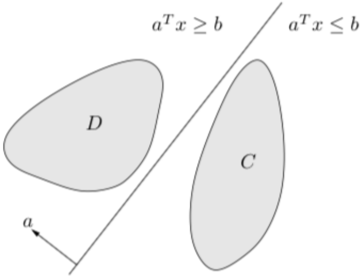
  This is always a convex cone, regardless of C



- Positive semidefinite cone: $\mathbb{S}^n_+ = \{X \in \mathbb{S}^n : X \succeq 0\}$, where $X \succeq 0$ means that X is positive semidefinite (and $\mathbb{S}^n$ is the set of $n \times n$ symmetric matrices).

Key properties of convex sets

- Separating hyperplane theorem: two disjoint convex sets have a separating hyperplane between them

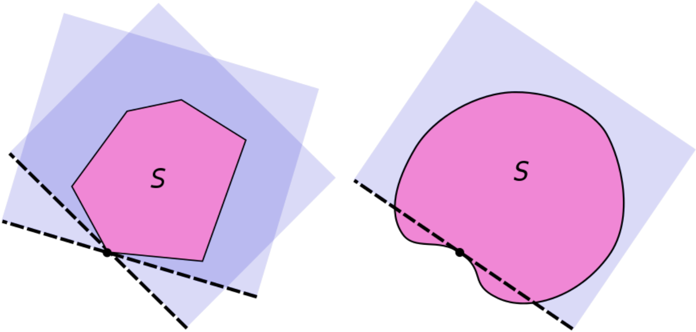


Formally, if C, D are nonempty convex set with $C \cap D = \emptyset$, then there exists a, b such that

$$
\begin{aligned}
C &\subseteq \{x : a^T x \leq b\} \\
D &\subseteq \{x : a^T x \geq b\}
\end{aligned}
$$

Key properties of convex sets

- Supporting hyperplane theorem: a boundary point of a convex set has a supporting hyperplane passing through it



Formally, if C is a nonempty convex set, and $x_0 \in \mathbf{bd}(C)$, then there exists a such that

$$C \subseteq \{x : a^T x \leq a^T x_0\}$$

Operations preserving convexity

- Intersection: the intersection of convex set is convex
- Scaling and translation: if C is convex, then

$$aC + b = \{ax + b : x \in C\}$$

  is convex for any a, b
- Affine images and preimages: if $f(x) = Ax + b$ and C is convex then

$$f(C) = \{f(x) : x \in C\}$$

  is convex and if D is convex, then

$$f^{-1}(D) = \{x : f(x) \in D\}$$

  is convex.

Example: linear matrix inequality solution set

- Given $A_1, \cdots, A_k, B \in \mathbb{S}^n$, a linear matrix inequality is of the form

$$x_1 A_1 + x_2 A_2 + \cdots + x_k A_k \preceq B$$

  for a variable $x \in \mathbb{R}^k$. Let's prove that the set C of points x that satisfy the above inequality is convex.
- Approach 1: directly verify that $x, y \in C \Rightarrow tx + (1-t)y \in C$. This follows by checking that, for any v

$$v^T \left(B - \sum_{i=1}^{k} (tx_i + (1-t)y_i) A_i \right) v \geq 0$$

- Approach 2: let $f : \mathbb{R}^k \to \mathbb{S}^n, f(x) = B - \sum_{i=1}^{k} x_i A_i$. Note that $C = f^{-1}(\mathbb{S}^n_+)$: affine preimage of convex set

Example: fantope

- Given some integer $k \geq 0$, the fantope of order k is

$$\mathcal{F} = \{Z \in \mathbb{S}^n : 0 \preceq Z \preceq I, \mathrm{tr}(Z) = k\}$$

where the trace operator $\mathrm{tr}(Z) = \sum_i Z_{ii}$ is the sum of the diagnoal entries. Prove that $\mathcal{F}$ is convex.

- Approach 1: verify that $0 \preceq Z, W \preceq I$ and $\mathrm{tr}(Z) = \mathrm{tr}(W) = k$, implies that the same for $tZ + (1 - t)W$
- Approach 2: recognize the fact that

$$\mathcal{F} = \{Z \in \mathbb{S}^n : Z \succeq 0\} \cap \{Z \in \mathbb{S}^n : Z \preceq I\} \cap \{Z \in \mathbb{S}^n : \mathrm{tr}(Z) = k\}$$

i.e. the intersection of linear equality constraints (hence like a polyhedron but for matrices).

More operations that preserving convexity

- Perspective images and preimages: the perspective function is $P : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n$ (where $\mathbb{R}_{++}$ denotes positive reals)

$$P(x, z) = x/z$$

  for $z > 0$. If $C \subseteq \text{dom}(P)$ is convex then so is $P(C)$, and if D is convex, then so is $P^{-1}(D)$.

- Linear-fractional images and preimages: the perspective map composed with an affine function

$$f(x) = \frac{Ax + b}{c^T x + d}$$

  is called a linear-fractional function, defined on $c^T x + d > 0$. If $C \subseteq \text{dom}(f)$ is convex then so is $f(C)$, and if D is convex then so is $f^{-1}(D)$

Example: conditional probability set

Let U, V be random variables over $\{1, \cdots, n\}$ and $\{1, \cdots, m\}$. Let $C \subseteq \mathbb{R}^{nm}$ be a set of joint distributions for U, V, i.e., each $p \in C$ defines joint probabilities

$$p_{ij} = \mathbb{P}(U = i, V = j)$$

Let $D \subseteq \mathbb{R}^{nm}$ contain corresponding conditional distributions, i.e. each $q \in D$ defines

$$q_{ij} = \mathbb{P}(U = i | V = j)$$

Assume C is convex. Let's prove that D is convex,. Write

$$D = \left\{ q \in \mathbb{R}^{nm} : q_{ij} = \frac{p_{ij}}{\sum_{k=1}^{n} p_{kj}}, \text{ for some } p \in C \right\} = f(C)$$

where f is a linear-fractional function, hence D is convex.

Convex functions

Convex function: $f : \mathbb{R}^n \to \mathbb{R}$ such that $\text{dom}(f) \subset \mathbb{R}^n$ is convex and

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y), \text{ for } 0 \le t \le 1$$

and all $x, y \in \text{dom}(f)$

In words, f lies below the line segment joining $f(x)$ and $f(y)$

Concave function: opposite inequality above, so that

$$f \text{ concave} \iff -f \text{ convex}$$

Convex functions

Important modifiers:

- Strictly convex: $f(tx + (1-t)y) < tf(x) + (1-t)f(y)$ for $x \neq y$ and $0 < t < 1$. In words, f is a convex and has greater curvature than a linear function
- Strongly convex with parameter $m > 0$: $f - \frac{m}{2}\|x\|_2^2$ is convex (m-strongly convex). In words, f is at least as convex as a quadratic function.

Note: strongly convex $\Rightarrow$ strictly convex $\Rightarrow$ convex

Analogously for concave function

Strictly convex $\not\Rightarrow$ strongly convex, example?

Examples of Convex Functions

- Univariate functions
  - Exponential function: e^{ax} is convex for any $a \in \mathbb{R}$
  - Power function: x^a is convex for $a \geq 1$ or $a < 0$ over $\mathbb{R}_+$ (nonnegative reals)
  - Power function is concave for $0 \leq a < 1$ over $\mathbb{R}_+$
  - Logarithmic function: $\log x$ is concave over $\mathbb{R}_+$
- Affine function: $x^T a + b$ is both convex and concave
- Quadratic function: $\frac{1}{2} x^T Q x + b^T x + c$ is convex provided that $Q \succeq 0$
- Least square loss: $\|y - AX\|_2^2$ is always convex (since $A^T A$ is always positive semidefinite.

Examples of Convex Functions

- Norm: $\|x\|$ is convex for any norm, e.g. ℓ_p-norm for $p \geq 1$

$$\|x\|_p = \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} \quad \text{for} \quad p \geq 1, \|x\|_\infty = \max_{i=1,\cdots,n} |x_i|$$

and also operator (spectral) and trace (nuclear) norms:

$$\|X\|_{\text{op}} = \sigma_1(X), \qquad \|X\|_{\text{tr}} = \sum_{i=1}^{r} \sigma_i(X)$$

where $\sigma_1(X) \geq \sigma_2(X) \cdots \sigma_r(X) \geq 0$ are the singular value of the matrix X.

Examples of Convex Functions

- Indicator function: if C is a convex, then its indicator function

$$I_C(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases}$$

  is convex

- Support function: for any set C (convex or not), its support function

$$I_C^*(x) = \max_{y \in C} x^T y$$

  is convex

- Max function:

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

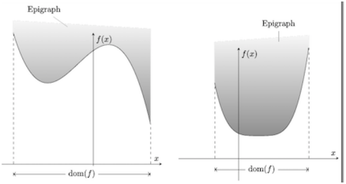
  is convex

Key Properties of convex functions

- A function is convex only and only if its restriction to any line is convex
- Epigraph characterization: A function is convex if and only if epigraph

$$\text{epi}(f) = \{(x, t) \in \text{dom}(f) \times R : f(x) \leq t\}$$

is a convex set



- Convex sublevel set: if f is convex, then its sublevel set

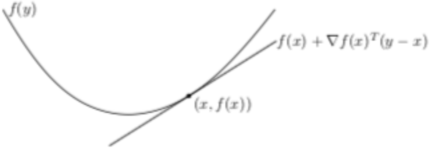
$$\{x \in \text{dom}(f) : f(x) \leq t\}$$

are convex for all $t \in \mathbb{R}$. The converse is not true.

- First-order characterization: if f is differentiable, then f is convex if and only if $\mathrm{dom}(f)$ is convex and

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

  for all $x, y \in \mathrm{dom}(f)$. Therefore for a differentiable convex function $\nabla f(x) = 0 \iff x$ minimizes f



- Second-order characterization: if f is twice differentiable, then f is convex if and only if $\mathrm{dom}(f)$ is convex and $\nabla^2 f(x) \succeq 0$ for all $x \in \mathrm{dom}(f)$
- Jensen's Inequality: if f is convex, and X is a random variable supported on $\mathrm{dom}(f)$, then $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$

Operations preserving convexity

- Nonnegative linear combination: $f_1, \cdots, f_m$ convex implies $a_1 f_1 + \cdots + a_m f_m$ convex for any $a_1, \cdots, a_m \geq 0$

- Pointwise maximization: if f_s is convex for any $s \in S$, then $f(x) = \max_{x \in S} f_s(x)$ is convex. Note that the set S here (number of functions f_s) can be infinite.

- Partial minimization: if $g(x, y)$ is convex in x, y and C is convex, then
$$f(x) = \min_{y \in C} g(x, y)$$
is convex.

Example: distances to a set

Let C be an arbitrary set, and consider the maximum distance to C under arbitrary norm $\|\cdot\|$:

$$f(x) = \max_{y \in C} \|x - y\|$$

Let's check this is convex: $f_y(x) = \|x - y\|$ is convex in x for any fixed y, so by pointwise maximization rule, f is convex.

Now let C be convex and consider the minimum distance to C

$$f(x) = \min_{y \in C} \|x - y\|$$

Let's check this is convex,. $g(x, y) = \|x - y\|$ is convex in x, y jointly, and C is assumed convex, so apply partial minimization rule.

More operations preserving convexity

- Affine composition: f convex implies $g(x) = f(Ax + b)$ convex
- General composition: suppose $f = h \circ g$ where $g : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R} \to \mathbb{R}, f : \mathbb{R}^n \to \mathbb{R}$, Then
  - f is convex if h is convex and nondecreasing, and g is convex
  - f is convex if h is convex and nonincreasing, and g is concave
  - f is concave if h is concave and nondecreasing, g is concave
  - f is concave if h is concave and nonincreasing, g is convex

How to remember these? Think of the chain rule when $n = 1$

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

More operations preserving convexity

- Vector composition: suppose that

$$f(x) = h(g(x)) = h(g_1(x), \cdots, g_k(x))$$

where $g : \mathbb{R}^n \to \mathbb{R}^k, h : \mathbb{R}^k \to \mathbb{R}, f : \mathbb{R}^n \to \mathbb{R}$. Then

  - f is convex if h is convex and nondecreasing in each argument, and g is convex
  - f is convex if h is convex and nonincreasing in each argument, and g is concave
  - f is concave if h is concave and nondecreasing in each argument, g is concave
  - f is concave if h is concave and nonicreasing in each argument, g is convex.

Example: log-sum-exp function

Log-sum-exp function:

$$g(x) = \log\left(\sum_{i=1}^{k} e^{a_i^T x + b_i}\right)$$

for fixed $a_i, b_i, i = 1, \cdots, k$. Often called "soft-max", as it smoothly approximates $\max_{i=1,2,\cdots,k}(a_i^T x + b_i)$. Convex function!

How to show convexity? First, note it suffices to prove convexity of $f(x) = \log\left(\sum_{i=1}^{k} e_i^x\right)$ (affine composition rule)

Now use second-order characterization. Calculate

$$\begin{aligned} \nabla_i f(x) &= \frac{e^{x_i}}{\sum e^{x_j}} \\ \nabla_{ij}^2 f(x) &= \frac{e^{x_i}}{\sum_\ell e^{x_\ell}} I\{i = j\} - \frac{e^{x_i} e^{x_j}}{(\sum_\ell e^{x_\ell})} \end{aligned}$$

Write $\nabla^2 f(x) = \text{diag}(z) - zz^T$ where $z_i = e^{x_i}/(\sum_j e^{x_j})$. This matrix is diagonally dominant, hence positive semidefinite.

Check

$$\max\left\{\log\left(\frac{1}{(a^T x + b)^7}\right), \|Ax + b\|_1^5\right\}$$

convex?

Convex Optimization Problems

Optimization problem:

$$\begin{array}{ll} \min_{x \in D} & f(x) \\ \text{subject to} & g_i(x) \leq 0, i = 1, 2, \cdots, m \\ & h_j(x) = 0, j = 1, 2, \cdots, r \end{array}$$

Here $D = \text{dom}(f) \cap \bigcap_{i=1}^{m} \text{dom}(g_i) \cap \bigcap_{j=1}^{p} \text{dom}(h_j)$, common domain of all the functions.

This is a convex optimization problem provided the functions f and $g_i, i = 1, 2, \cdots, m$ are convex and $h_j, j = 1, 2, \cdots, p$ are affine:

$$h_j(x) = a_j^T x + b_j, j = 1, \cdots, p$$

Local minima and global minima

For convex optimization problems, local minima are global minima

Formally, if x is feasible — $x \in D$, and satisfies all constraints — and minimizes f in a local neighborhood,

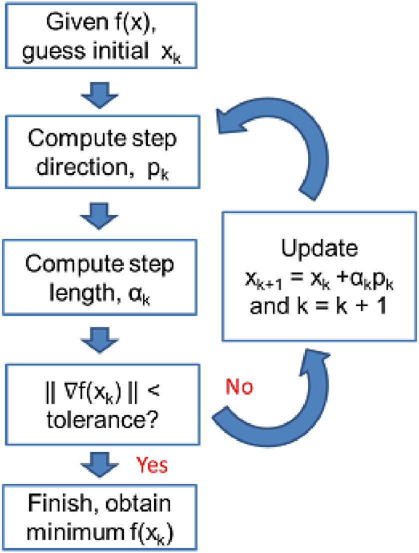
$$f(x) \leq f(y) \text{ for all feasible } y, \|x - y\|_2 \leq \rho,$$

then

$$f(x) \leq f(y) \text{ for all feasible } \; y$$

This is a very useful fact and will save us a lot of trouble!

Line search methods



Questions?