Convexity I: Sets and Functions

Ryan Tibshirani
Convex Optimization 10-725

See supplements for reviews of

- basic real analysis
- basic multivariate calculus
- basic linear algebra
Why convexity? Simply put: because we can broadly understand and solve convex optimization problems.

Nonconvex problems are mostly treated on a case by case basis.

Reminder: a convex optimization problem is of the form

\[
\min_{x \in D} \quad f(x)
\]

subject to \( g_i(x) \leq 0, \ i = 1, \ldots, m \)

\( h_j(x) = 0, \ j = 1, \ldots, r \)

where \( f \) and \( g_i, i = 1, \ldots, m \) are all convex, and \( h_j, j = 1, \ldots, r \) are affine. Special property: any local minimizer is a global minimizer.
Outline

Today:

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

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

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

In words, line segment joining any two elements lies entirely in set

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

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

with $\theta_i \geq 0$, $i = 1, \ldots, k$, and $\sum_{i=1}^{k} \theta_i = 1$. Convex hull of a set $C$, $\text{conv}(C)$, 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: \( \{ x : Ax = b \} \), for given \( A, b \)
• **Polyhedron:** \( \{ x : Ax \leq b \} \), where inequality \( \leq \) is interpreted componentwise. Note: the set \( \{ x : Ax \leq b, Cx = d \} \) is also a polyhedron (why?)

\[
\begin{align*}
\text{Figure 2.11} & \quad \text{The polyhedron } P \text{ (shown shaded) is the intersection of five halfspaces, with outward normal vectors } a_1, \ldots, a_5. \\
& \quad \text{It will be convenient to use the compact notation } \\
P = \{ x \mid Ax \preceq b, Cx = d \} \quad (2.6)
\end{align*}
\]

\( A = \begin{bmatrix}
a_1^T \\
a_2^T \\
a_3^T \\
a_4^T \\
a_5^T
\end{bmatrix}, C = \begin{bmatrix}
c_1^T \\
c_2^T \\
c_3^T \\
c_4^T \\
c_5^T
\end{bmatrix} \)

• **Simplex:** special case of polyhedra, given by 
\[
\text{conv}\{x_0, \ldots, x_k\}, \text{ where these points are affinely independent.}
\]

The canonical example is the **probability simplex**, 
\[
\text{conv}\{e_1, \ldots, e_n\} = \{ w : w \geq 0, 1^T w = 1 \}
\]
Cones

Cones: \( C \subseteq \mathbb{R}^n \) such that

\[
x \in C \implies tx \in C \text{ for all } t \geq 0
\]

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

\[
x_1, x_2 \in C \implies t_1 x_1 + t_2 x_2 \in C \text{ for all } t_1, t_2 \geq 0
\]

Conic combination of \( x_1, \ldots, x_k \in \mathbb{R}^n \): any linear combination

\[
\theta_1 x_1 + \cdots + \theta_k x_k
\]

with \( \theta_i \geq 0, i = 1, \ldots, k \). Conic hull collects all conic combinations.
Examples of convex cones

- **Norm cone:** \( \{(x, t) : \|x\| \leq t\} \), for a norm \( \| \cdot \| \). Under the \( \ell_2 \) norm \( \| \cdot \|_2 \), called second-order cone

- **Normal cone:** given any set \( C \) and point \( x \in C \), we can define

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

  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 between hyperplane them

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

\[
C \subseteq \{x : a^T x \leq b\} \\
D \subseteq \{x : a^T x \geq b\} 
\]
• **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 $x_0 \in \text{bd}(C)$, then there exists $a$ such that

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

Both of the above theorems (separating and supporting hyperplane theorems) have partial converses; see Section 2.5 of BV.
Operations preserving convexity

- **Intersection**: the intersection of convex sets 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 $A_1, \ldots, A_k, B \in \mathbb{S}^n$, a linear matrix inequality is of the form

$$x_1A_1 + x_2A_2 + \cdots + x_kA_k \preceq B$$

for a variable $x \in \mathbb{R}^k$. Let's prove 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 $f : \mathbb{R}^k \rightarrow \mathbb{S}^n$, $f(x) = B - \sum_{i=1}^k x_iA_i$. Note that $C = f^{-1}(\mathbb{S}_+^n)$, affine preimage of convex set.
More operations 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),

  \[ P(x, z) = \frac{x}{z} \]

  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 if $f(C)$, and if $D$ is convex then so is $f^{-1}(D)$.
Example: conditional probability set

Let $U, V$ be random variables over $\{1, \ldots, n\}$ and $\{1, \ldots, m\}$. 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

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

where $f$ is a linear-fractional function, hence $D$ is convex
Convex functions

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

\[
f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y)
\]

for \( 0 \leq t \leq 1 \)

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

In words, function lies below the line segment joining \( f(x), f(y) \)

Concave function: opposite inequality above, so that

\[
f \text{ concave} \iff -f \text{ convex}
\]
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 convex and has greater curvature than a linear function.

- **Strongly convex** with parameter $m > 0$: $f - \frac{m}{2} \|x\|_2^2$ is convex. In words, $f$ is at least as convex as a quadratic function.

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

(Analogously for concave functions)
Examples of convex functions

- **Univariate functions:**
  - Exponential function: $e^{ax}$ is convex for any $a$ over $\mathbb{R}$
  - Power function: $x^a$ is convex for $a \geq 1$ or $a \leq 0$ over $\mathbb{R}_+$ (nonnegative reals)
  - Power function: $x^a$ is concave for $0 \leq a \leq 1$ over $\mathbb{R}_+$
  - Logarithmic function: $\log x$ is concave over $\mathbb{R}_{++}$

- **Affine function:** $a^T x + b$ is both convex and concave

- **Quadratic function:** $\frac{1}{2}x^T Q x + b^T x + c$ is convex provided that $Q \succeq 0$ (positive semidefinite)

- **Least squares loss:** $\|y - Ax\|_2^2$ is always convex (since $A^T A$ is always positive semidefinite)
• **Norm:** $\|x\|$ is convex for any norm; e.g., $\ell_p$ norms,

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

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

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

where $\sigma_1(X) \geq \ldots \geq \sigma_r(X) \geq 0$ are the singular values of the matrix $X$
• **Indicator function:** if $C$ is 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:** $f(x) = \max\{x_1, \ldots, x_n\}$ is convex
Key properties of convex functions

- A function is convex if and only if its restriction to any line is convex.

- **Epigraph characterization**: a function $f$ is convex if and only if its epigraph
  
  $$\text{epi}(f) = \{(x, t) \in \text{dom}(f) \times \mathbb{R} : f(x) \leq t\}$$

  is a convex set.

- **Convex sublevel sets**: if $f$ is convex, then its sublevel sets
  
  $$\{x \in \text{dom}(f) : f(x) \leq t\}$$

  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 $\text{dom}(f)$ is convex, and

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

for all $x, y \in \text{dom}(f)$. 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 $\text{dom}(f)$ is convex, and $\nabla^2 f(x) \succeq 0$ for all $x \in \text{dom}(f)$

• **Jensen’s inequality:** if $f$ is convex, and $X$ is a random variable supported on $\text{dom}(f)$, then $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$
Operations preserving convexity

- **Nonnegative linear combination:** If $f_1, \ldots, f_m$ are convex, then $a_1 f_1 + \cdots + a_m f_m$ is convex for any $a_1, \ldots, a_m \geq 0$.

- **Pointwise maximization:** If $f_s$ is convex for any $s \in S$, then $f(x) = \max_{s \in S} f_s(x)$ 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 an arbitrary norm $\| \cdot \|:

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

Let's check convexity: $f_y(x) = \| x - y \|$ 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 convexity: $g(x, y) = \| x - y \|$ is convex in $x, y$ jointly, and $C$ is assumed convex, so apply partial minimization rule
More operations preserving convexity

- Affine composition: if \( f \) is convex, then \( g(x) = f(Ax + b) \) is convex

- General composition: suppose \( f = h \circ g \), where \( g : \mathbb{R}^n \rightarrow \mathbb{R} \), \( h : \mathbb{R} \rightarrow \mathbb{R} \), \( f : \mathbb{R}^n \rightarrow \mathbb{R} \). Then:
  - \( f \) is convex if \( h \) is convex and nondecreasing, \( g \) is convex
  - \( f \) is convex if \( h \) is convex and nonincreasing, \( g \) is concave
  - \( f \) is concave if \( h \) is concave and nondecreasing, \( g \) concave
  - \( f \) is concave if \( h \) is concave and nonincreasing, \( g \) convex

How to remember these? Think of the chain rule when \( n = 1 \):

\[
f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)
\]
Vector composition: suppose that

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

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, $g$ is convex
- $f$ is convex if $h$ is convex and nonincreasing in each argument, $g$ 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 nonincreasing in each argument, $g$ is convex
Example: log-sum-exp function

Log-sum-exp function: \( g(x) = \log(\sum_{i=1}^{k} e^{a_i^T x + b_i}) \), for fixed \( a_i, b_i \), \( i = 1, \ldots, k \). Often called “soft max”, as it smoothly approximates \( \max_{i=1, \ldots, k} (a_i^T x + b_i) \).

How to show convexity? First, note it suffices to prove convexity of \( f(x) = \log(\sum_{i=1}^{n} e^{x_i}) \) (affine composition rule).

Now use second-order characterization. Calculate

\[
\nabla_i f(x) = \frac{e^{x_i}}{\sum_{\ell=1}^{n} e^{x_{\ell}}}
\]
\[
\nabla_{ij}^2 f(x) = \frac{e^{x_i} e^{x_j}}{\sum_{\ell=1}^{n} e^{x_{\ell}}} 1\{i = j\} - \frac{e^{x_i} e^{x_j}}{(\sum_{\ell=1}^{n} e^{x_{\ell}})^2}
\]

Write \( \nabla^2 f(x) = \text{diag}(z) - zz^T \), where \( z_i = e^{x_i} / (\sum_{\ell=1}^{n} e^{x_{\ell}}) \). This matrix is diagonally dominant, hence positive semidefinite.
References and further reading