Convexity

Weikai Xu

School of Information Science and Technology
ShanghaiTech University

May 27th, 2020

Outline

Outline

Convex Sets

Combinations and hulls

1. Given a set of points (vectors) in $\mathbb{R}^n$:

$$\mathcal{P} = \{x^{(1)}, \cdots, x^{(m)}\}$$

   the *linear hull* (subspace) generated by these points is the set of all possible linear combinations of the points:

$$x = \lambda_1 x^{(1)} + \cdots + \lambda_m x^{(m)}, \text{ for } \lambda_i \in \mathbb{R}, i = 1, \ldots, m$$

2. The *affine hull*, aff $\mathcal{P}$, of $\mathcal{P}$ is the set generated by taking all possible linear combinations of the points in $\mathcal{P}$, under the restriction that the coeefficients λ_i sum up to one, that is $\sum_{i=1}^m \lambda_i = 1$. aff $\mathcal{P}$ is the smallest affine set containing $\mathcal{P}$.

3. A *convex combination* of the points is a special type of linear combination, in which the coefficients λ_i are restricted to be nonnegative and to sum up to one, that is

$$\lambda_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^m \lambda_i = 1$$

Convex Sets

Convexity

1. Intuitively, a convex combination is a weighted average of the points, with weights given by the λ_i coefficients. The set of all possible convex combination is called the convex hull of the point set:

$$\mathrm{co}(x^{(1)}, \ldots, x^{(m)}) = \{x = \sum_{i=1}^{m} \lambda_i x^{(i)} : \lambda_i \geq 0, i = 1, \ldots, m; \sum_{i=1}^{m} \lambda_i = 1\}$$

2. Similarly, the *conic hull* of a set of points is defined as

$$\mathrm{conic}(x^{(1)}, \ldots, x^{(m)}) = \{x = \sum_{i=1}^{m} \lambda_i x^{(i)}, \lambda_i \geq 0, i = 1, \ldots, m\}$$

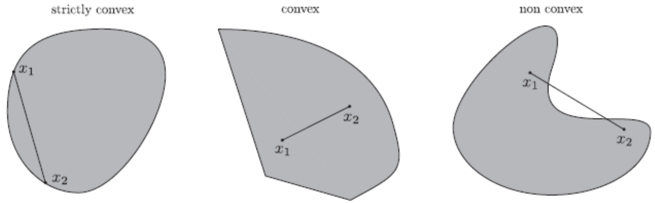
Convex Sets

Convexity

1. A subset $C \in \mathbb{R}^n$ is said to be convex if it contains the line segment between any two points in it:

$$x_1, x_2 \in C, \lambda \in [0,1] \rightarrow \lambda x_1 + (1-\lambda) x_2 \in C$$

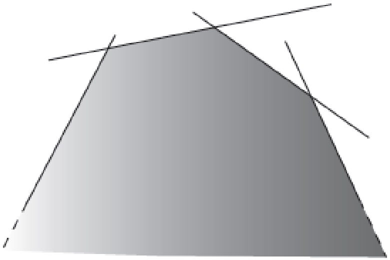
2. Subspaces and affine sets, such as lines and hyperplanes are obviously convex, as they contain the entire line passing through any two points. Half-spaces are also convex.
3. A set C is a cone if $x \in C$, then $\alpha x \in C$, for every $\alpha > 0$. A set C is said to be a convex cone if it is convex and it is a cone. The conic hull of a set is a convex cone.



Operations that preserve convexity

Intersection

1. If $C_1, \ldots, C_m$ are convex sets, then their intersection $C = \bigcap_{i=1,\ldots,m} C_i$ is also a convex set
2. The intersection rule actually holds for possibly infinite families of convex sets: if$C(\alpha), \alpha \in \mathcal{A} \subseteq \mathbb{R}^q$, is a family of convex sets, parameterized by α, then the set $C = \bigcap_{\alpha \in \mathcal{A}} C_\alpha$ is convex.
3. Example: An halfspace $\mathcal{H} = \{x \in \mathcal{R}^n : c^T x \leq d\}, c \neq 0$ is a convex set. The intersection of m halfspaces $\mathcal{H}_i, i = 1, \cdots, m$ is a convex set called a ployhedron.



Examples

Second-order cone

1. The second-order cone in $\mathbb{R}^{n+1}$:

$$\mathcal{K}_n = \{(x,t), x \in \mathbb{R}^n, t \in \mathbb{R} : ||x||_2 \leq t\}$$

is convex, since it is the intersection of half-spaces:

$$\mathcal{K}_n = \bigcap_{y:||u||_2 \leq 1} \{(x,t), x \in \mathbb{R}^n, t \in \mathbb{R} : u^T x \leq t\}$$

2. Here, we have used the representation of $||\cdot||_2$ based on the Cauchy-Schwarz inequality:

$$||x||_2 = \max_{u:||u||_x \leq 1} u^T x$$

which implies that

$$||x||_2 \leq t \leftrightarrow u^T x \leq t \text{ for every } u \text{ such that } ||u||_2 \leq 1$$

Examples

Set of positice semi-definite matrices

1. Recall that a symmetric matrix $X \in \mathcal{S}^n$ is positive-semidefinite if and only if

$$\forall u \in \mathbb{R}^n : u^T X u \geq 0$$

2. The set of symmetric, positive-semidefinite matrices, $\mathbb{S}_+^n$, is the intersection of (an infinite number of) half-spaces in $\mathbb{S}^n$:

$$\mathbb{S}_+^n = \bigcap_{u \in \mathbb{R}^n} \{X \in \mathbb{S}^n : u^T X u \geq 0\}$$

Hence, $\mathbb{S}_+^n$ is convex. In fact, it is a convex cone, since multiplying a PSD matrix by a positive number results in a PSD matrix.

Operations that preserve convexity

Affine transformation

1. If a map $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine, and $C \subset R^n$ is convex, then the image set

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

is convex.

2. This fact is easily verified: any affine map has a matrix representation

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

Then, for any $y^{(1)}, y^{(2)} \in f(C)$ there exist $x^{(1)}, x^{(2)}$ in C such that $y^{(1)} = Ax^{(1)} + b, y^{(2)} = Ax^{(2)} + b$. Hence, for $\lambda \in [0, 1]$, we have that

$$\lambda y^{(1)} + (1 - \lambda)y^{(2)} = A(\lambda x^{(1)} + (1 - \lambda)x^{(2)}) + b = f(x)$$

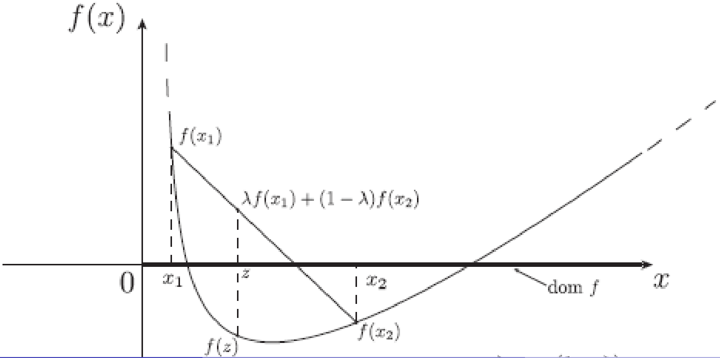
where $x = \lambda x^{(1)} + (1 - \lambda)x^{(2)} \in C$

3. In particular, the projection of a convex set C onto a subspace is representable by means of a linear map, hence the projected set is convex.

Outline

Convex Functions

1. The domain of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set over which the function is well-defined:
$$\text{dom } f = \{x \in \mathbb{R}^n : -\infty < f(x) < \infty\}$$
2. A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if dom f is a convex set, and for all $x, y \in$ dom f and all $\lambda \in [0, 1]$ it holds that
$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$
3. We say that a function f is concave if $-f$ is convex.



About the domain of a convex function

1. Convex functions must be $+\infty$ outside their domains, so that the convex property remains valid even if x or $y \notin \operatorname{dom} f$. The function

$$f(x) = \begin{cases} -\sum_{i=1}^{n} \log x_i & \text{if } x > 0, \\ +\infty & \text{otherwise,} \end{cases}$$

is convex, but the function

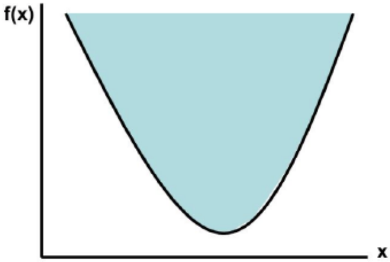
$$f(x) = \begin{cases} -\sum_{i=1}^{n} \log x_i & \text{if } x > 0, \\ -\infty & \text{otherwise,} \end{cases}$$

is not.

Epigraph

1. Given a function $f : \mathbb{R}^n \to (-\infty, +\infty]$, its epigraph (i.e., the set of points lying above the graph of the function) is the set

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



Fact: f is a convex function if and only if epi f is a convex set.

Example

1. Consider the "log-sum-exp" function arising in logistic regression:

$$x \in \mathbb{R}^n \to f(x) = \log(\sum_{i=1}^{n} e^{x_i})$$

2. The epigraph is the set of pairs (x, t) characterized by the inequality $t \geq f(x)$, which can be re-written as

$$\exp f = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : \sum_{i=1}^{n} e^{x_i - t} \leq 1\}$$

which is convex, due to the convexity of the exponential function.

Operations that preserve convexity

Nonnegative linear combinations

1. If $f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, m$,are convex functions, then the function

$$f(x) = \sum_{i=1}^{m} \alpha_i f_i(x), \alpha_i \geq 0, i = 1, \ldots, m$$

   is also convex over $\cap_i \mathrm{dom}\ f_i$

2. This fact easily follows from the definition of convexity, since for any $x, y \in \mathrm{dom}\ f$ and $\lambda \in [0, 1]$,

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \sum_{i=1}^{m} \alpha_i f_i(\lambda x + (1-\lambda)y) \leq \sum_{i=1}^{m} \alpha_i(\lambda f_i(x) + (1-\lambda) f_i(y)) \\ &= \lambda f(x) + (1-\lambda) f(y) \end{aligned}$$

3. Example: the negative entropy function with values for $x \in \mathbb{R}^n_{++}$

$$f(x) = \sum_{i=1}^{n} x_i \log x_i$$

Operations that preserve convexity

Affine variable transformation

1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex, and define

$$g(x) = f(Ax + b), A \in \mathbb{R}^{n,m}, b \in \mathbb{R}^n$$

Then, g is convex over dom $g = \{x : Ax + b \in \text{dom} f\}$

Examples:

1. $f(z) = -\log(z)$, is convex over dom$f = \mathbb{R}_{++}$, hence $f(x) = -\log(ax + b)$ is also convex over $ax + b > 0$
2. For any convex function $\mathcal{L} : \mathbb{R} \to \mathbb{R}$, the function

$$(w, b) \in \mathbb{R}^n \times \mathbb{R} \to \sum_{i=1}^{m} \mathcal{L}(w^T x_i + b)$$

where $x_1, \ldots, x_m \in \mathbb{R}^n$ are given data points, is convex. (Such functions arise as "loss" functions in machine learning.)

Outline

First-order conditions

1. If f is differentiable (that is, dom f is open and the gradient exists everywhere on the domain), then f is convex if and only if

$$\forall x, y \in \text{dom } f, f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

2. **Proof.** Assume that f is convex. Then, the definition implies that for any $\lambda \in (0, 1]$

$$\frac{f(x + \lambda(y - x)) - f(x)}{\lambda} \leq f(y) - f(x),$$

which, for $\lambda \to 0$ yields $\nabla f(x)^T (y - x) \leq f(y) - f(x)$.

3. Conversely, take and $x, y \in \text{dom } f$ and $\lambda \in [0, 1]$, and let $z = \lambda x + (1 - \lambda) y$:

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

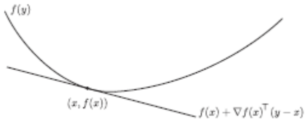
Taking a convex combination of these inequalities, we get

$$\lambda f(x) + (1 - \lambda) f(y) \geq f(z) + \nabla f(z)^T 0 = f(z)$$

First-order conditions

Geometric interpretation

$$\forall x, y \in \text{dom } f, f(y) \geq f(x) + \nabla f(x)^T (y - x)$$



The graph of f is bounded below everywhere by anyone of its tangent hyperplanes.

1. The gradient of a convex function at a point $x \in \mathbb{R}^n$ (if it is nonzero) divides the whole space in two halfspaces:

$$\mathcal{H}_{++}(x) = \{y : \nabla f(x)^T (y - x) > 0\}$$
$$\mathcal{H}_{-}(x) = \{y : \nabla f(x)^T (y - x) \leq 0\}$$

and any point $y \in \mathcal{H}_{++}(x)$ is such that $f(y) > f(x)$.

2. This is a key fact exploited by the so-called "gradient" algorithms for minimizing a convex function.

Second-order conditions

1. If f is twice differentiable, then f is convex if and only if its Hessian matrix $\nabla^2 f$ is positive semi-definite everywhere on the (open) domain of f, that is if and only if $\nabla^2 f \succeq 0$ for all $x \in \text{dom } f$.
2. Example: a generic quadratic function

$$f(x) = \frac{1}{2}x^T Hx + c^T x + d$$

has Hessian $\nabla^2 f(x) = H$. Hence f is convex if and only if H is positive semidefinite.

Restriction to a line

1. A function f is convex if and only if its restriction to any line is convex.
2. By restriction to a line we mean the function

$$g(t) = f(x_0 + tv)$$

   of scalar variable t, for fixed $x_0 \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$
3. This rule gives a very powerful criterion for proving convexity of certain functions.
4. Example: for the log-determinant function $f(X) = -\log detX$ over $X \succ 0$, it holds that

$$\begin{aligned} g(t) &= -\log\det(X_0 + tV) = -\log\det X_0(1 + tX_0^{-1/2}VX_0^{-1/2}) \\ &= -\log\det X_0 \prod_{i=1,\dots,n}(1 + t\lambda_i(X_0^{-1/2}VX_0^{-1/2})) \\ &= -\log\det X_0 + \sum_{i=1}^{n} -\log(1 + t\lambda_i(X_0^{-1/2}VX_0^{-1/2})) \end{aligned}$$

Pointwise maximum

1. If $(f_\alpha)_{\alpha\in\mathcal{A}}$ is a family of convex functions indexed by parameter α, and $\mathcal{A}$ is a set, then the pointwise max function

$$f(x) = \max_{\alpha\in\mathcal{A}} f_\alpha(x)$$

   is convex over the domain $\{\cap_{\alpha\in\mathcal{A}} \mathrm{dom}\ f_\alpha\} \cap \{x : f(x) < \infty\}$

2. **Proof:** The epigraph of f is the set of pairs (x, t) such that

$$\forall\alpha \in \mathcal{A} : f_\alpha(x) \le t$$

   hence, the epigraph of f is the intersection of the epigraphs of all the functions involved, therefore f is convex.

Pointwise maximum rule

Example: functions arising in SOCP

1. The function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, with values

$$f(y, t) = ||y||_2 - t$$

is convex since it's the pointwise maximum of linear function of (y, t):

$$f(y, t) = \max_{u:||u||_2 \leq 1} u^T y - t$$

2. Using the rule of affine variable transformation, we obtain that for any matrices A, C, vector b and scalar d, the function

$$x \to ||Ax + b||_2 - (c^T x + d)$$

is also convex.

Pointwise maximum rule

Example: sum of k largest elements

1. Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ with values

$$f(x) = \sum_{i=1}^{k} x_{[i]}$$

   where $x_{[i]}$ denotes the i-th largest element in x.

2. We have:

$$f(x) = \max_u u^T x : u \in \{0,1\}^n, 1^T u = k$$

   For every $u, x \to u^T x$ is linear, hence f is convex.

Pointwise maximum rule

Example: largest eigenvalue of a symmetric matrix

1. Consider the function $f : \mathbb{S}^n \to \mathbb{R}$ with values for a given $X = X^T \in \mathbb{S}^n$ given by
$$f(x) = \lambda_{\max}(X)$$
where $\lambda_{\max}$ denotes the largest eigenvalue.
2. The function is the pointwise maximum of linear functions of X:
$$f(x) = \max_{u:||u||_2=1} u^T X u$$
Hence, f is convex.