Convex Optimization

Part 1: Convex sets and functions

Namhoon Lee

POSTECH

14 Sep 2022

Convex optimization

Consider an optimization problem:

$$\min_{x \in \mathbb{C}} \; f(x)$$

i.e., "minimize a function f subject to x being in the set $\mathbb{C}$".

We call the above a *convex* optimization problem if:

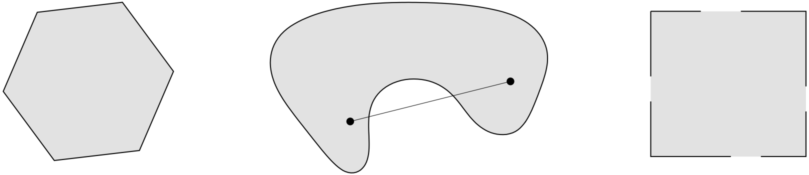
- The set $\mathbb{C}$ is a convex set.
- The function f is a convex function.

Key property: All local minima are global minima.

Convex sets

Convex set: a set $\mathbb{C}$ is convex if, for any $x_1, x_2 \in \mathbb{C}$ and any $0 \le \theta \le 1$, it contains the line segment between x_1 and x_2 in $\mathbb{C}$

$$\theta x_1 + (1 - \theta) x_2 \in \mathbb{C} \ .$$



Convex sets

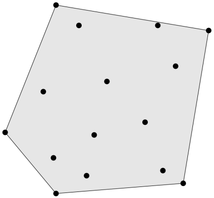
Convex combination of $x_1, ..., x_k$: any point x of the form

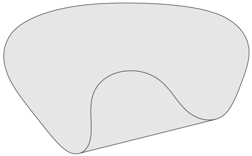
$$x = \theta_1 x_1 + \theta_2 x_2 + ... + \theta_k x_k$$

with $\theta_1 + \theta_2 + ... + \theta_k = 1$ and $\theta_i \geq 0$.

Convex hull of a set $\mathbb{C}$: set of all convex combinations of points in $\mathbb{S}$

$$\{\theta_1 x_1 + ... + \theta_k x_k \mid x_i \in \mathbb{C},\ \theta_1 + ... + \theta_k = 1,\ \theta_i \geq 0\}\ .$$





Cones

Cone: if for every $x \in \mathbb{C}$ and $\theta \geq 0$ we have

$$\theta x \in \mathbb{C} \ .$$

Conic (nonnegative) combination of x_1 and x_2: any point of the form

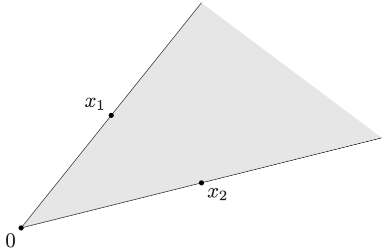
$$x = \theta_1 x_1 + \theta_2 x_2$$

with $\theta_1 \geq 0, \theta_2 \geq 0$.

Cones

Convex cone: a set $\mathbb{C}$ is a convex cone if it is convex and a cone; for any $x_1, x_2 \in \mathbb{C}$ and $\theta_1, \theta_2 \geq 0$, we have

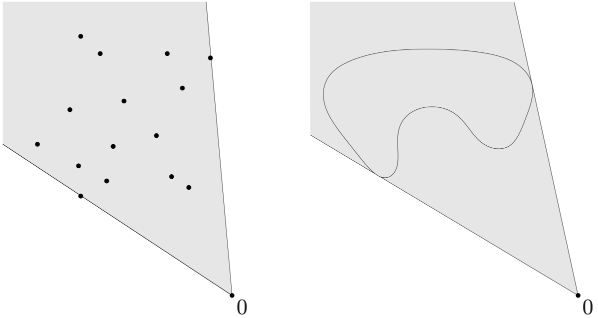
$$\theta_1 x_1 + \theta_2 x_2 \in \mathbb{C} \ .$$



Cones

Conic hull of a set $\mathbb{C}$: the set of all conic combinations of points in $\mathbb{C}$

$$\{\theta_1 x_1 + \ldots + \theta_k x_k \mid x_i \in \mathbb{C},\ \theta_i \geq 0\} \ .$$



Cones

Polar cone of a set $\mathbb{C}$: the set of all points that make $\geq 90^\circ$ with any point in $\mathbb{C}$

$$\{y \mid \langle y, x \rangle \leq 0, \ \forall x \in \mathbb{C}\} \ .$$

Tangent cone: the closure of all directions you are allowed to move in within $\mathbb{C}$.

- If x is an interior to $\mathbb{C}$ then the tangent cone is $\mathbb{R}^n$.

Normal cone: the polar of tangent cone.

- It will be used to characterize optimality conditions later.

Examples of convex sets

Some simple ones:

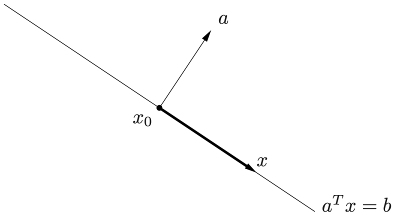
- empty set
- single point
- line and line segment
- subspace and the whole space

Examples of convex sets

Hyperplane:

$$\mathbb{H} = \{x \in \mathbb{R}^n : a^\top x = b\}$$

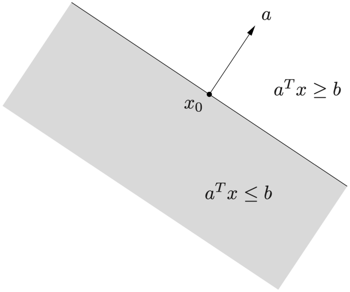
where $a \neq 0$ and $b \in \mathbb{R}$.



Examples of convex sets

Half space:

$$\mathbb{H}^{+} = \{x \in \mathbb{R}^n : a^\top x \geq b\} \quad \text{or} \quad \mathbb{H}^{-} = \{x \in \mathbb{R}^n : a^\top x \leq b\} \ .$$



Examples of convex sets

Norm ball with center x_c and radius r:

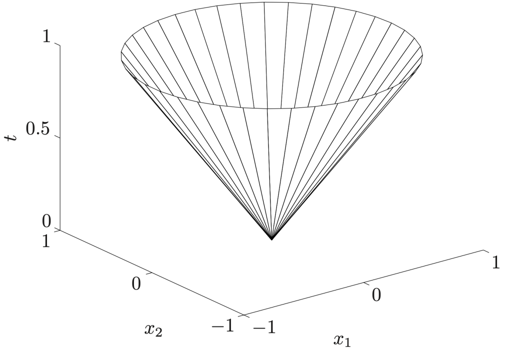
$$\{x : \|x - x_c\|_p \leq r\} \ .$$

- Examples of different $1 \leq p \leq \infty$.
- Prove Euclidean balls (*i.e.*, $p = 2$) are convex.

Examples of convex sets

Norm cone:

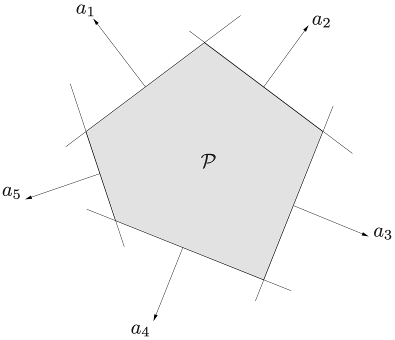
$$\{(x, t) : \|x\|_p \leq t\}$$



Examples of convex sets

Polyhedron: the solution set of a finite number of linear equalities and inequalities

$$\mathbb{P} = \{x : a_i^\top x \le b_i, i = 1, ..., m,\ c_j^\top x = d_j, j = 1, ..., p\}\ .$$



Convexity preserving operations

Intersections of convex sets are convex.

Let $C_i, i \in \mathbb{I}$ be convex sets, where $\mathbb{I}$ is a index set. Then $C = \cap_{i\in\mathbb{I}} C_i$ is a convex set.

Example: linear program with linear inequalities constraints $Ax \leq b$.

- Each constraint $a_i^\top x \leq b_i$ defines a half-space.
- Half-spaces are convex sets.
- So the set of x satisfying $Ax \leq b$ is the intersection of convex sets.

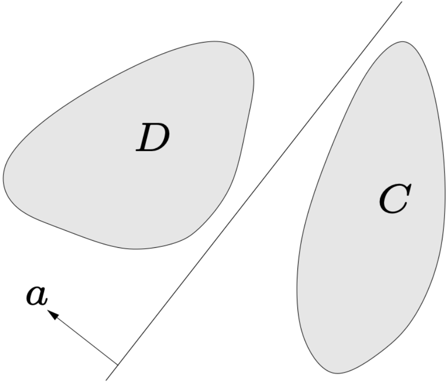
Other operations that preserve convexity

- Affine images and inverse images (*e.g.*, scaling and translation)
- Perspective images and inverse images (*e.g.*, pin-hole camera)
- Linear-fractional images and inverse images (*e.g.*, projective transformation)

Properties of convex sets

Separating hyperplane theorem

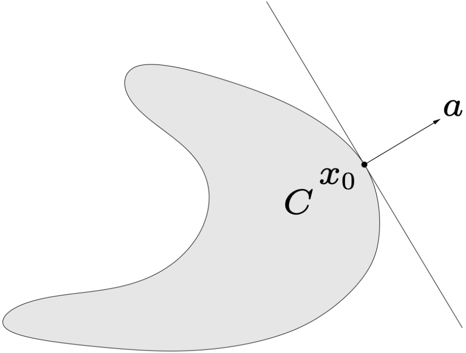
If C and D are nonempty disjoint convex sets, there exist $a \neq 0, b$ such that $a^\top x \leq b$ for $x \in C$ and $a^\top x \geq b$ for $x \in D$.



Properties of convex sets

Supporting hyperplane theorem

If C is convex, then there exist a supporting hyperplane at every boundary point of C.



Convex functions

C^0 definition of convex functions:

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \ , \quad \forall x, y \in \mathbb{R}^n \text{ and } 0 \leq t \leq 1 \ .$$

- A C^0 function is convex iff the function is below its chord between any two points.

Convex functions

C^1 definition of convex functions:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \ , \quad \forall x, y \in \mathbb{R}^n \ .$$

- A C^1 function is convex iff the function is above its tangent planes at any point.

Convex functions

C^2 definition of convex functions:

$$\nabla^2 f(x) \succeq 0 \ , \quad \forall x \in \mathbb{R}^n \ .$$

- A C^2 function is convex iff it is curved upwards everywhere.

Show that the C^2 definition is equivalent to the C^1 definition.

First, let's recall the fundamental theorem of calculus:

$$\int_0^1 F'(t)dt = F(1) - F(0)$$

Now consider the following:

$$\int_0^1 (x-y)^\top \nabla^2 f(tx + (1-t)y)dt = \int_0^1 \frac{d}{dt}\Big(\nabla f(tx + (1-t)y)\Big)dt = \nabla f(x) - \nabla f(y)$$

Convex functions

Multiplying by $x - y$ both sides gives

$$\int_0^1 (x-y)^\top \nabla^2 f(tx + (1-t)y)(x-y)dt = \langle \nabla f(x) - \nabla f(y), x - y \rangle$$

By applying C^2 definition, we obtain

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$$

- It's called function is monotone; *i.e.*, C^2 function is monotone.
- You can also show that C^1 function is monotone.

Convex functions

Next, consider the following:

$$\int_0^1 \nabla f((y-x)t+x)^\top (y-x)dt = \int_0^1 \frac{d}{dt}\Big(f((y-x)t+x)\Big)dt = f(y)-f(x)$$

Rearranging it gives

$$f(y) = f(x) + \int_0^1 \nabla f((y-x)t+x)^\top (y-x)dt$$

We want to relate this to the C^1 definition, while using monotonicity.

Convex functions

From the monotonicity, we can show that the integrand is smallest at $t = 0$, *i.e.*,

$$\langle \nabla f((y-x)t + x) - \nabla f(x), (y-x)t + x - x \rangle \geq 0$$
$$\langle \nabla f((y-x)t + x) - \nabla f(x), y - x \rangle \geq 0$$

Therefore, we can say

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

- If function is monotone, it's convex.
- More rigorous proofs exist.

Convex functions: examples

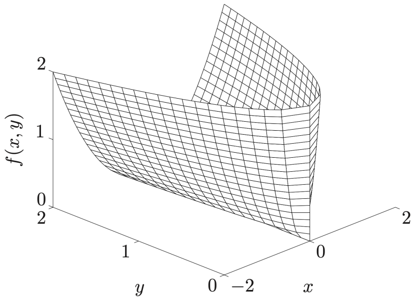
Example: For $f(x) = x^\top Qx$ where Q is postive semidefinite, show f is convex using definitions of convex functions.

Convex functions: examples

Example: Show p-norm is convex.

Convex functions: examples

Example: Show $f(x, y) = x^2/y$ is convex.



Convexity preserving operations

Nonnegative weighted sum

If $\alpha, \beta \geq 0$ and f_1, f_2 convex, $\alpha f_1 + \beta f_2$ is convex.

Pointwise maximum

If $f_1, ..., f_m$ are convex, $\max\{f_1(x), ..., f_m(x)\}$ is convex.

Composition with affine map

If f is convex, $f(Ax + b)$ is convex.

Partial minimization

If $g(x, y)$ is convex in x, y, and $\mathbb{C}$ is convex, then $f(x) = \min_{y \in \mathbb{C}} g(x, y)$ is convex.

Convex functions

More on convex functions..

- Notice from C^1 definition that $\nabla f(x) = 0$ implies $f(y) \geq f(x)$ for all y, so x is a global minimizer; this further explains why least squares can be solved by setting the derivative equal to zero.
- Strictly-convex function have at most one global minimum; w and v can't both be global minima if $w \neq v$; it would imply convex combinations u of w and v would have $f(u)$ below the global minimum.

Convex functions

For strictly convex objective f there can be at most one global optimum.

Proof:

1. Suppose x^* is a local minimum and also there exists another local minimum $x^\#$ $(\neq x^*)$.
2. Since f is convex (because it is strictly convex), $f(x^*)$ and $f(x^\#)$ are both global minima, and $f(x^*) = f(x^\#)$.
3. The C^0 definition for $y = \theta x^* + (1-\theta)x^\#$, *i.e.*,

$$f(y) < \theta f(x^*) + (1-\theta) f(x^\#) = \theta f(x^*) + (1-\theta) f(x^*) = f(x^*)$$

   contradicts that x^* is a global minimum.
4. This means that for $x^\#$ to be a local minimum, it must be that $x^\# = x^*$.

Any questions?