Chapter A: Convex Analysis

Ernest K. Ryu

MATH 164: Optimization
University of California, Los Angeles
Department of Mathematics

Last edited: February 12, 2025

Line segment

Given $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$,

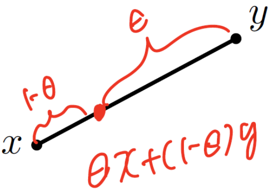
$$\theta x + (1 - \theta) y$$

is a point in between x and y if $\theta \in [0, 1]$.

The set of all points between a given $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$

$$\{\theta x + (1 - \theta) y \,|\, \theta \in [0, 1]\}$$

is called the *line segment* between x and y



Convex combinations

Given $x_1, \ldots, x_k \in \mathbb{R}^n$,

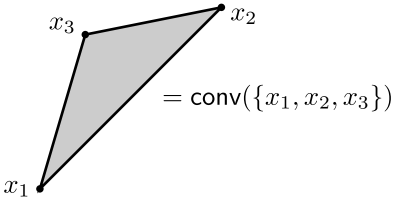
$$\theta_1 x_1 + \cdots + \theta_k x_k$$

is called a *convex combination* or a *weighted average* of $x_1, \ldots, x_k$ if $\theta_1, \ldots, \theta_k \geq 0$ and $\theta_1 + \cdots + \theta_k = 1$.

Given $x_1, \ldots, x_k \in \mathbb{R}^n$, the set of all convex combinations

$$\mathrm{conv}(\{x_1, \ldots, x_k\}) = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid \theta_1, \ldots, \theta_k \geq 0,\ \theta_1 + \cdots + \theta_k = 1\}$$

is called the *convex hull* of $x_1, \ldots, x_k$.



Convex sets

We say a set $C \subseteq \mathbb{R}^n$ is *convex* if

$$\theta x + (1-\theta) y \in C, \qquad \forall\, x, y \in C,\, \theta \in (0,1).$$

In other words, C is convex if $x, y \in C$ implies the line segment connecting x and y is wholly contained in C.

TODO: Add picture

Convex functions

We say a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is *convex* if

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y), \qquad \forall\, x, y \in \mathbb{R}^n,\ \theta \in [0,1].$$

I.e., f is convex if the chord (line segment) connecting $(x, f(x))$ and $(y, f(y))$ lies above the graph of f.

TODO: Picture

We say $f\colon \mathbb{R}^n \to \mathbb{R}$ is *concave* if $-f$ is convex.

Strictly convex functions

Recall that $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex if

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y), \qquad \forall\, x, y \in C,\, x \neq y,\, \theta \in (0,1).$$

(Our prior definition of convexity definition is equivalent to this.)

We say $f\colon \mathbb{R}^n \to \mathbb{R}$ is *strictly convex* if

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y), \qquad \forall\, x, y \in C,\, x \neq y,\, \theta \in (0,1).$$

I.e., f is strictly convex if the chord connecting $(x, f(x))$ and $(y, f(y))$ lies strictly above the graph of f (excluding the endpoints).

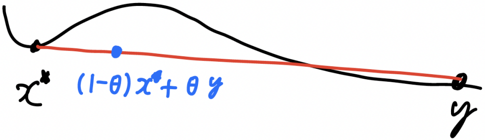
No bad local minima for cvx. functions

Theorem.

Let f be convex. Then any local minimizer is a global minimizer.

Thus, when we minimize convex functions, we never get stuck at bad local minima because there aren't any bad local minima.

Illustration of proof. Let $x_\star$ be a local minimizer. Assume for contradiction that $x_\star$ is not a global minimum.



Draw a contradiction because the chord is below the graph for $\theta \approx 0$.

No bad local minima for cvx. functions

Proof. Let $x_\star \in \mathbb{R}^n$ be a local minimizer of f. Assume for contradiction that there is $y \in \mathbb{R}^n$ such that $f(y) < f(x_\star)$, i.e., assume for contradiction that $x_\star$ is not a global minimizer. By convexity,

$$f((1-\theta)x_\star + \theta y) \leq (1-\theta)f(x_\star) + \theta f(y) < f(x_\star)$$

for any $\theta \in (0, 1)$, even for θ very close to 0. However, $x_\star$ is a local minimizer, so $f((1-\theta)x_\star + \theta y) \geq f(x_\star)$ for θ sufficiently close to 0, and we have a contradiction. Thus we conclude that such y cannot exist, i.e., $x_\star$ is a global minimizer. $\square$

Gradient provides global lower bound for cvx. functions

Theorem.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Assume f is differentiable at x. Then,

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

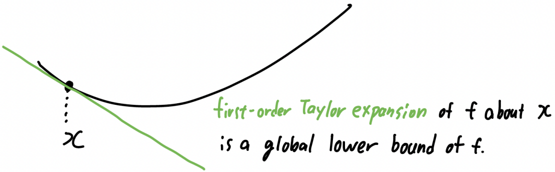
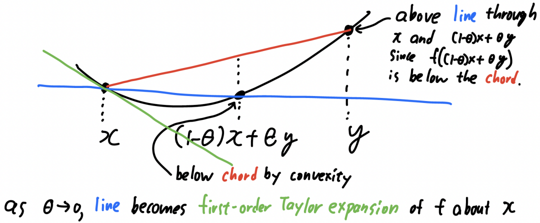


Illustration of proof.



Gradient provides global lower bound for cvx. functions

Theorem.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ *be convex. Assume* f *is differentiable at* x*. Then,*

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

Proof. By convexity,

$$f(x + \theta(y - x)) \leq (1 - \theta) f(x) + \theta f(y), \qquad \forall\, \theta \in (0, 1).$$

Reorganizing, we get

$$f(y) \geq f(x) + \frac{f(x + \theta(y - x)) - f(x)}{\theta}, \qquad \forall\, \theta \in (0, 1).$$

By taking $\theta \to 0$, we get the desired inequality. $\square$

Gradient provides global lower bound for cvx. functions

The inequality

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

is called the *convexity inequality*.

It turns out that the convexity inequality is equivalent to convexity, i.e., a differentiable $f \colon \mathbb{R}^n \to \mathbb{R}$ is convex if and only if it satisfies the convexity inequality.

Basic calculus of convex sets and functions

Theorem.
The intersection of convex sets is convex.

Theorem.
A nonnegative combination of convex functions is convex.

Theorem.
A sublevel set of a convex function is convex.

Basic calculus of convex sets and functions

Theorem.
The intersection of convex sets is convex.

So if $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^n$ are convex sets, then $A \cap B$ are convex.

- The intersection can be arbitrary, i.e., the intersection can be over countably or uncountably infinite convex sets.
- To clarify, an empty set is defined to be a convex set, and the intersection of convex sets can be empty.

Basic calculus of convex sets and functions

Theorem.
A nonnegative combination of convex functions is convex.

I.e., if $\alpha_1, \dots, \alpha_k$ are nonnegative scalars and $f_1, \dots, f_k$ are convex functions, then $\alpha_1 f_1 + \cdots + \alpha_k f_k$ is convex.

- If f is convex, then αf is convex and $-\alpha f$ is concave if $\alpha \geq 0$.
- Often, one shows that an f is convex by arguing that $f = g + h$ and showing that g and h are convex.

Basic calculus of convex sets and functions

Theorem.
A sublevel set of a convex function is convex.

For any $f\colon \mathbb{R}^n \to \mathbb{R}$ and $\alpha \in \mathbb{R}$, the α-sublevel set of f is defined as

$$\{x \mid f(x) \leq \alpha\} \subseteq \mathbb{R}^n,$$

which is the set of x attaining function value better than α.

- In particular, this implies that the set of minimizers of a convex function is convex.
- Often, one shows that a set is convex by showing that it is a sublevel set of a convex function.

Convexity via monotonicity

For differentiable f, convexity is monotonicity of f'.

Theorem.

A differentiable univariate function $f\colon \mathbb{R} \to \mathbb{R}$ is convex if and only if f' is non-decreasing.

(To clarify, convex functions need not be differentiable.)

Proof. ($\Rightarrow$) Assume f is convex. Then, by the convexity inequality,

$$f(y) \geq f(x) + f'(x)(y - x)$$
$$f(x) \geq f(y) + f'(y)(x - y)$$

for all $x, y \in \mathbb{R}$. Adding the two, we get

$$(f'(x) - f'(y))(x - y) \geq 0,$$

which implies $f'(x) \geq f'(y)$ if $x > y$, i.e. f' is non-decreasing.

Convexity via monotonicity

($\Leftarrow$) Assume $f' \colon \mathbb{R} \to \mathbb{R}$ is non-decreasing. Let $z = \theta x + (1-\theta)y$ with $\theta \in [0,1]$. Then,

$$f(y) - f(z) = \int_z^y f'(t)\, dt \geq \int_z^y f'(z)\, dt = f'(z)(y-z)$$

$$f(x) - f(z) = -\int_x^z f'(t)\, dt \geq -\int_x^z f'(z)\, dt = f'(z)(x-z)$$

multiplying the first inequality by $(1-\theta)$ and the second my θ and adding them gives us

$$\theta(x) + (1-\theta)f(y) - f(z) \geq f'(z) \underbrace{\big(\theta x + (1-\theta)y - z\big)}_{=0} = 0,$$

which is the definition of convexity. $\square$

Convexity via curvature

For twice-differentiable f, convexity is positive (nonnegative) curvature.

Theorem.
A twice-differentiable univariate function $f : \mathbb{R} \to \mathbb{R}$ *is convex if and only if* $f''(x) \geq 0$ *for all* $x \in \mathbb{R}$*.*

Proof. From the previous theorem, f is convex if and only if f' is non-decreasing. Since f' is assumed to be differentiable, this holds if and only if $f'' \geq 0$. □

For multivariate convex functions, the curvature condition is given by eigenvalues of the Hessian. (We omit the proof.)

Theorem.
A twice-differentiable multivariate function $f : \mathbb{R}^n \to \mathbb{R}$ *is convex if and only* $\nabla^2 f(x)$ *has nonnegative eigenvalues for all* $x \in \mathbb{R}$*.*

Affine functions are convex

Theorem.
An affine function is convex.

(A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is said to be affine is $f(x) = \langle a, x\rangle + b$ for some $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$.)

Proof 1. An affine function has 0 curvature, which is nonnegative. $\square$

Proof 2. If f is affine,

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \langle a, \theta x + (1-\theta)y\rangle + b \\ &= \theta\langle a, x\rangle + \theta b + (1-\theta)\langle a, y\rangle + (1-\theta)b \\ &= \theta f(x) + (1-\theta)f(y). \end{aligned}$$

$\square$

Cocoercivity inequality for smooth convex functions

Theorem.

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex and L-smooth. Then,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L} \|\nabla f(y) - \nabla f(x)\|^2, \qquad \forall\, x, y \in \mathbb{R}^n.$$

Proof. Let

$$g(y) = f(y) - \langle \nabla f(x), y \rangle.$$

Then x is a minimizer of g, since g is convex and $\nabla g(x) = 0$ by construction. Since g is L-smooth, the L-smoothness lemma gives us

$$g(x) \leq g(y + \delta) \leq g(y) + \langle \nabla g(y), \delta \rangle + \frac{L}{2} \|\delta\|^2, \qquad \forall \delta \in \mathbb{R}^n.$$

Now let $\delta = -\frac{1}{L} \nabla g(y) = -\frac{1}{L}(\nabla f(y) - \nabla f(x))$ and we get

$$f(x) - \langle \nabla f(x), x \rangle \leq f(y) - \langle \nabla f(x), y \rangle - \frac{1}{2L} \|\nabla g(y)\|^2.$$

Rearranging the terms, we conclude the statement. $\square$

Cocoercivity inequality for smooth convex functions

This inequality

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L} \|\nabla f(y) - \nabla f(x)\|^2$$

is called the *cocoercivity inequality* for smooth convex functions.

Note that this is stronger than the convexity inequality

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

which holds for differentiable convex functions.

Projection onto convex sets

Projection[1] of $p \in \mathbb{R}^n$ onto C is the point within C that is closest to p. Is this notion well-defined?

Theorem.

Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set and let $p \in \mathbb{R}^n$. Then

$$\Pi_C(p) = \operatorname*{argmin}_{x \in C} \|x - p\|,$$

where $\|\cdot\|$ is the standard Euclidean norm, uniquely exists.

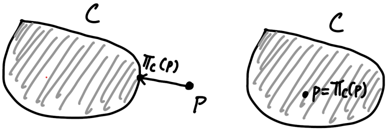


Illustration when C is nonempty closed convex. (Setting of the theorem)

[1] In linear algebra, our notion of projection corresponds to orthogonal projections but not oblique projections.

Projection onto convex sets

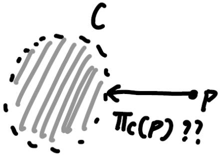


Illustration when C is open. The projection is not attained.

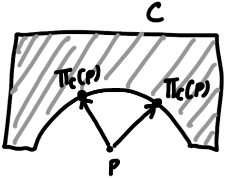


Illustration when C is not convex. Projection may not be unique.

Projection onto convex sets

Proof. Clearly,

$$\Pi_C(p) = \underset{x \in C}{\operatorname{argmin}} \|x - p\| = \underset{x \in C}{\operatorname{argmin}} \|x - p\|^2.$$

Since $\|x-p\|^2$ is a strictly convex function of f, a minimizer, if exists, must be unique. (So, there are 0 or 1 minimizers.)

Let $\{x_k\}_k$ be a sequence such that

$$\|x_k - p\| \to \inf_{x \in C} \|x - p\|.$$

Since $\{x_k\}_k$ is bounded, it has a convergent subsequence $x_{k_j} \to x_\infty \in C$ by the Bolzano–Weierstrass theorem and closedness of C. By continuity of $\|x-p\|^2$ as a function of x, we conclude

$$\|x_\infty - p\|^2 = \inf_{x \in C} \|x - p\|^2,$$

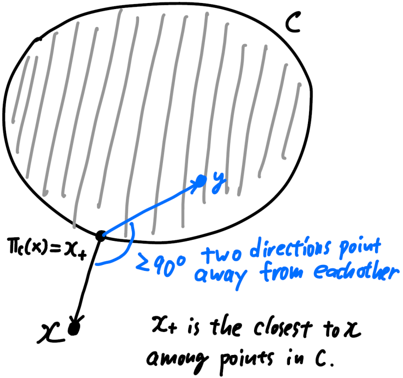
i.e., x_∞ is a minimizer. (So, there are more than 0 minimizers.) $\square$

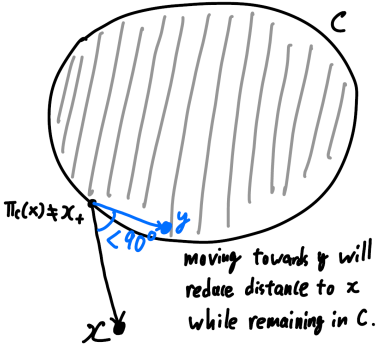
Projection theorem

Theorem.
Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then, $x_+ = \Pi_C(x)$ if and only if

$$\langle y - x_+, x - x_+ \rangle \leq 0, \qquad \forall\, y \in C.$$

(Also called the Bourbaki–Cheney–Goldstein inequality.)





Projection theorem

Theorem (Projection theorem).

Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then, $x_+ = \Pi_C(x)$ if and only if $x_+ \in C$ and

$$\langle y - x_+, x - x_+ \rangle \leq 0, \qquad \forall\, y \in C.$$

Proof. $(\Rightarrow)$ Assume $x_+ = \operatorname{argmin}_{z \in C} \|z - x\|^2$ and let $y \in C$. Then,

$$\|y - x\|^2 \geq \|x_+ - x\|^2, \qquad \forall\, y \in C$$

and $\theta(y - x_+) + x_+ \in C$ for $\theta \in (0, 1]$. So

$$\|\theta(y - x_+) + x_+ - x\|^2 \geq \|x_+ - x\|^2.$$

Reorganizing the terms, we get

$$\theta^2 \|y - x_+\|^2 + 2\theta\langle y - x_+, x_+ - x \rangle \geq 0.$$

Dividing by θ and letting $\theta \to 0$, we conclude

$$\langle y - x_+, x_+ - x \rangle \geq 0.$$

Projection theorem

($\Leftarrow$) Conversely, if $x_+ \in C$ and

$$\langle y - x_+, x_+ - x \rangle \geq 0, \qquad \forall\, y \in C,$$

then

$$\|(y - x) - (x_+ - x)\|^2 + 2\langle y - x_+, x_+ - x \rangle \geq 0, \qquad \forall\, y \in C.$$

Expanding the squares, we get

$$\|y - x\|^2 + \|x_+ - x\|^2 - 2\langle x_+ - x, x_+ - x \rangle \geq 0, \qquad \forall\, y \in C,$$

and we conclude

$$\|y - x\|^2 \geq \|x_+ - x\|^2, \qquad \forall\, y \in C,$$

i.e., $x_+ = \Pi_C(x)$. $\square$

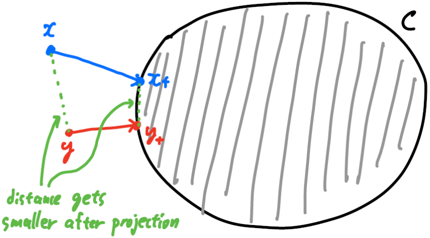
$$\|y - x_+\|^2 + 2\langle y - x_+, x_+ - x\rangle \geq 0, \qquad \forall\, y \in C,$$

Projection is nonexpansive

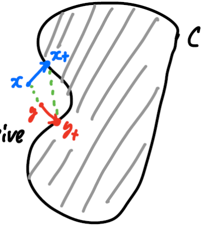
Theorem.
Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then $\Pi_C : \mathbb{R}^n \to \mathbb{R}^n$ is a nonexpansive operator.

In other words, if $x_+ = \Pi_C(x)$ and $y_+ = \Pi_C(y)$, then

$$\|x_+ - y_+\| \leq \|x - y\|.$$



If C is non-convex
Π_C may not be nonexpansive



Projection is nonexpansive

Theorem.

Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then $\Pi_C \colon \mathbb{R}^n \to \mathbb{R}^n$ is a nonexpansive operator.

Proof. Let $x, y \in \mathbb{R}^n$, $x_+ = \Pi_C(x)$, and $y_+ = \Pi_C(y)$. By the projection theorem,

$$\langle y_+ - x_+, x - x_+ \rangle \le 0$$
$$\langle x_+ - y_+, y - y_+ \rangle \le 0.$$

Summing these two inequalities, we get

$$\langle x_+ - y_+, x_+ - y_+ \rangle \le \langle x_+ - y_+, x - y \rangle.$$

Using Cauchy–Schwartz, we get

$$\|x_+ - y_+\|^2 \le \langle x_+ - y_+, x - y \rangle \le \|x_+ - y_+\| \|x - y\|.$$

Dividing by $\|x_+ - y_+\|$ (when nonzero), we conclude the statement. $\square$