Convex Optimization

Pere Ara
(Universitat Autònoma de Barcelona)

2023-2024

Convex functions

Definition

$f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$

for all $x, y \in \mathbb{R}^n$ and all $t \in [0, 1]$.

Definition

$f : \mathbb{R}^n \to \mathbb{R}$ is **strictly convex** if

$$f((1-t)x + ty) < (1-t)f(x) + tf(y)$$

for all $x \neq y$ in $\mathbb{R}^n$ and all $t \in (0, 1)$.

Definition

$f : \mathbb{R}^n \to \mathbb{R}$ is **concave** if $-f$ is convex. This is to say that

$$f((1-t)x + ty) \geq (1-t)f(x) + tf(y)$$

for all $x, y \in \mathbb{R}^n$ and all $t \in [0, 1]$.

Analogously, f is **strictly concave** if $-f$ is strictly convex.

Observe that $(1-t)x + ty = x + t(y-x)$, hence for $t \in \mathbb{R}$, $(1-t)x + ty$ describes the line determined by x and y, and for $t \in [0,1]$, it describes the segment between x and y.

Exercise

Show that $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its restriction to lines is convex, that is, for each distinct $x, y \in \mathbb{R}^n$, we have that $g\colon \mathbb{R} \to \mathbb{R}$ given by

$$g(t) = f((1-t)x + ty) = f(x + t(y-x))$$

for $t \in \mathbb{R}$, is a convex function in the variable t.

Convexity and derivatives

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Then f is convex if and only if

$$f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$$

for all $x, y \in \mathbb{R}^n$.
Moreover, f is strictly convex if and only if

$$f(y) > f(x) + \nabla f(x) \cdot (y - x)$$

for all $x, y \in \mathbb{R}^n$ with $x \neq y$.

Observe that for $n = 1$ this means that the function lies above the tangent line at any point.

Theorem

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a convex differentiable function. Then $x_0 \in \mathbb{R}^n$ is a global minimum for f if and only if $\nabla f(x_0) = 0$.

Proof.

We already know that if x_0 is a global minimum then $\nabla f(x_0) = 0$.

We now show the converse. Suppose that $\nabla f(x_0) = 0$.
By the Theorem above we have

$$f(y) \geq f(x_0) + \nabla f(x_0) \cdot (y - x_0) = f(x_0).$$

Hence x_0 is a global minimum. □

Theorem

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable. Then we have

1. *f is convex if and only if $H(f)(x) = \nabla^2 f(x)$ is positive semidefinite for all $x \in \mathbb{R}^n$.*
2. *If $H(f)(x) = \nabla^2 f(x)$ is positive definite for all $x \in \mathbb{R}^n$, then f is strictly convex.*

Global minima of convex functions

In general a convex function can have more than one global minima, or have no global minima. However we have:

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex function. Then f has at most one global minimum $x^* \in \mathbb{R}^n$.

Proof.

Suppose that $x \neq y$ are two global minima for f, so that

$$f(x) = f(y) = f_{\min} \quad \text{and} \quad f(z) \geq f_{\min} \ \forall z \in \mathbb{R}^n.$$

For $t \in (0,1)$ we have

$$f_{\min} \leq f((1-t)x + ty) < (1-t)f(x) + tf(y) = f_{\min},$$

which is a contradiction. Hence there is at most one global minima for f. $\square$

By introducing a technical definition, we can assure that the function has exactly one global minimum.

Definition

Let $\mu > 0$ be a positive real number. We say that $f : \mathbb{R}^n \to \mathbb{R}$ is μ-strongly convex if the function

$$g(x) = f(x) - \frac{\mu}{2}\|x\|^2$$

is convex.

Theorem

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable and $\mu > 0$. Then f is μ-strongly convex if and only if $\nabla^2 f(x) - \mu I$ is positive semidefinite for all $x \in \mathbb{R}^n$.

Theorem

If f is μ-strongly convex for some $\mu > 0$, then f has exactly one global minimum.

The reason is that given any $x \in \mathbb{R}^n$, one can show that there is a big ball around x such that $f(y) \geq f(x)$ for all y not in this ball. Hence the global minimum is the global minimum of f on this big ball. (We are assuming f is differentiable, in particular continuous.)

We have

$$\mu\text{-convex} \implies \text{strictly convex} \implies \text{convex}$$

Convex sets

Definition

A subset S of $\mathbb{R}^n$ is *convex* if for all $x, y \in S$ and all $t \in [0, 1]$ we have

$$(1-t)x + ty \in S.$$

A nice relation between convex functions and convex sets is givenm by the epigraph of a function.

Given $f : \mathbb{R}^n \to \mathbb{R}$, the *epigraph* of f is:

$$E = \{(x, s) \in \mathbb{R}^n \times \mathbb{R} : s \geq f(x)\}.$$

Then we have

$$f \text{ is a convex function} \iff E \text{ is a convex set}$$

Given a convex function $f\colon \mathbb{R}^n \to \mathbb{R}$ and $s \in \mathbb{R}$, we have that

$$S = \{x \in \mathbb{R}^n : f(x) \leq s\}$$

is a convex set.

As an application one can see that any ball $B(a, R)$ in $\mathbb{R}^n$ is a convex set.

Constrained optimization for convex functions

Theorem

If $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function and $S \subset \mathbb{R}^n$ is a convex set, then the set of global minima of $\min_{x \in S} f(x)$ is a (possibly empty) convex set.

Proof.

Let x, y be two global minima of f on S, so that

$$f(x) = f(y) = f_{\min}, \quad \text{and} \quad f(z) \geq f_{\min} \ \forall z \in S.$$

For all $t \in [0, 1]$, $(1-t)x + ty \in S$ and

$$f_{\min} \leq f((1-t)x + ty) \leq (1-t)f(x) + tf(y) = f_{\min}.$$

Hence $f((1-t)x + ty) = f_{\min}$ and so $(1-t)x + ty$ is a global minimum of f on S.

$\square$

Theorem

If $S \subseteq \mathbb{R}^n$ is a convex set and $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex then there exists at most one global minimum $x^* \in S$ for $\min_{x \in S} f(x)$.

Proof.

Let x, y be two distinct global minima of f on S, so

$$f(x) = f(y) = f_{\min} \quad \text{and} \quad f(z) \geq f_{\min} \ \forall z \in S.$$

Now for all $t \in (0, 1)$ we have

$$f_{\min} \leq f((1-t)x + ty) < (1-t)f(x) + tf(y) = f_{\min}$$

and we get the contradiction $f_{\min} < f_{\min}$. $\square$

Theorem

Let S be a non-empty closed and convex subset of $\mathbb{R}^n$. If $f : \mathbb{R}^n \to \mathbb{R}$ is a μ-strongly convex function for some $\mu > 0$, then there is exactly one global minimum $x^* \in S$ for $min_{x \in S} f(x)$.

REASON: We already know there is at most one.

For a fixed $x \in S$ there is a big ball B centered at x such that $f(y) \geq f(x)$ for all y not in B.

Now the global minimum of f on S will be the global minimum of f on the closed and bounded set $S \cap B$.

Constrained optimization and Lagrange multipliers. Duality.

We consider the problem

$$\min_x f(x) \quad \text{subject to} \quad g_i(x) \leq 0, \; i = 1, \ldots, m.$$

where $f : \mathbb{R}^D \to \mathbb{R}$, $g_i : \mathbb{R}^D \to \mathbb{R}$.

Consider the Lagrangian

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) = f(x) + \lambda^T g(x), \quad \lambda \geq 0.$$

Consider the primal

$$\mathcal{L}_P(x) = \sup_{\lambda \in \mathbb{R}_+^m} \mathcal{L}(x, \lambda) \in \mathbb{R} \cup \{+\infty\}.$$

We have

$$\min_{x \in \mathbb{R}^D} \mathcal{L}_P(x) = \min_{x \in S} f(x).$$

The primal problem is:

$$\min_{x \in \mathbb{R}^D} \mathcal{L}_P(x).$$

The dual problem is:

$$\max_{\lambda \in \mathbb{R}_+^m} \mathcal{L}_D(\lambda)$$

where

$$\mathcal{L}_D(\lambda) = \inf_{x \in \mathbb{R}^D} \mathcal{L}(x, \lambda).$$

Hence the dual problem interchanges the max and the min.

Weak duality

Theorem (Weak duality)

For all $x \in \mathbb{R}^D$ and all $\lambda \in \mathbb{R}^m_+$ we have

$$\mathcal{L}_D(\lambda) \leq \mathcal{L}_P(x).$$

In particular we have

$$d^* := [max_{\lambda \in \mathbb{R}^m_+} \mathcal{L}_D(\lambda)] \leq [min_{x \in \mathbb{R}^D} \mathcal{L}_P(x)] = [min_{x \in S} f(x)] =: p^*,$$

where d^* is the optimal value for the dual problem, and p^* is the optimal value for the primal problem.

Observe that:

- $\min_{x \in \mathbb{R}^D} \mathcal{L}(x, \lambda)$ is an unconstrained problem, for any given $\lambda \in \mathbb{R}^m_+$.
- $\mathcal{L}(x, \lambda)$ is affine with respect to λ.

It follows that

$$\mathcal{L}_D(\lambda) = \min_{x \in \mathbb{R}^D} \mathcal{L}(x, \lambda) \quad \text{is always concave}$$

It follows that $\max_{\lambda \in \mathbb{R}^m_+} \mathcal{L}_D(\lambda)$ is a concave optimization problem.

Strong duality for convex functions

Theorem (Strong duality)

Let $f : \mathbb{R}^D \to \mathbb{R}$ be a convex function and let S be a convex set given by

$$S = \{x \in \mathbb{R}^D : g_i(x) \leq 0\}$$

for convex functions $g_1, \ldots, g_m$. Then we have strong duality, that is, the optimal value d^* for the dual problem equals the optimal value p^* for the primal problem, that is,

$$d^* := [max_{\lambda \in \mathbb{R}^m_+} \mathcal{L}_D(\lambda)] = [min_{x \in \mathbb{R}^D} \mathcal{L}_P(x)] = [min_{x \in S} f(x)] =: p^*.$$