

MATH466/MATH766

Math of machine learning

02/03-02/05 Lecture 7-8 Convex Optimization

References:

- Ch1 of Convex Optimization by Stephen Boyd and Lieven Vandenberghe
- <https://sites.math.washington.edu/~burke/crs/408f/notes/nlp/unoc.pdf>

Todays contents:

- Optimization overview
- Existence of minimizer
- Convexity and global minimizer, unique minimizer
- First and second order conditions of convex functions

Important concepts:

- coercivity
- convexity

Recommend reading:

- Ch 1.1, 1.3, 1.4 of Convex Optimization by Stephen Boyd and Lieven Vandenberghe
-

Warm up.

If $x_m < a$ for $m=1, 2, \dots$, then $\lim_{m \rightarrow +\infty} x_m \quad a$

$\sim x_m \leq a \quad \dots$, then $\lim_{m \rightarrow +\infty} x_m \quad a$.

D. Optimization Overview

- A general problem formulation

$$\min_x f_0(x) \quad (*)$$

subject to $f_i(x) \leq b_i, i=1, \dots, m$

optimization variable : $x = (x_1, \dots, x_n)$

objective function : $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$

constraint functions : $f_i : \mathbb{R}^n \rightarrow \mathbb{R}, i=1, \dots, m$

bounds : $b_i \in \mathbb{R}, i=1, \dots, m$

x^* is a global optimizer to $(*)$ if

① $f_i(x^*) \leq b_i$ (feasible)

and ② , $f_0(\bar{x}) \geq f_0(x^*)$ (global optimal)

x^* is a local optimizer to $(*)$ if

① $f_i(x^*) \leq b_i$ (feasible)

and ② there exists $r > 0$

s.t.

$f_0(\bar{x}) \geq f_0(x^*)$ (local optimal)

① linear programming

When f_0 and $f_i (i=1, \dots, m)$ are linear

"Solving (most) linear programs is a mature technology."

② convex programming

When f_0 and $f_i (i=1, \dots, m)$ are convex

$$\text{local} \quad \xleftarrow{\text{cvx}} \quad \text{global} \quad \xrightarrow{\text{cvx}} \quad \text{uniqueness}$$

③ nonlinear programming

When f_0 or $f_i (i=1, \dots, m)$ is not linear.

but not known to be convex.

This problem is not easy. In general

- local optimization : fast
 { give up seeking the optimal x , ...
 global optimization : seeking global optimal
 sacrifice efficiency

techniques
in cvx opt
can help

1. Existence of minimizer

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

* 1.1 When C is empty, solution doesn't exist

1.2 Weierstrass Extreme Value Theorem

- Continuous function : if $\lim_{m \rightarrow +\infty} x_m = x$, then $\lim_{m \rightarrow +\infty} f(x_m) = f(x)$
- Compact set (by Heine-Borel Thm) : $S \subseteq \mathbb{R}^n$ is compact iff
 - S is bounded : $\exists r > 0$ s.t. $\forall x \in S$, $\|x\| < r$
 - S is closed : for any x , if there exist $\{x_m\} \subseteq S$, $\lim_{m \rightarrow +\infty} x_m = x$ then $x \in S$

e.g. Compact :

non-compact :

• Weierstrass Extreme Value Theorem

If $f: S \rightarrow \mathbb{R}$, S is compact and f is continuous on S ,
then there exists $x^* \in S$ s.t. $f(x^*) \leq f(x)$ holds for any $x \in S$

i.e. cts func on compact set attains its extreme value on the set.

- Rigorous proof of this thm can be found in Real Analysis or advanced (proof-based) calculus textbook.
- Continuity is necessary.
o.w. counter e.g.
- compactness is necessary
o.w. counter e.g.

$$f: [-\frac{\pi}{2}, \frac{\pi}{2}] \rightarrow \mathbb{R} \quad x \mapsto \begin{cases} 0 & \text{if } x = \pm \frac{\pi}{2} \\ \tan x & \text{o.w.} \end{cases}$$

$$(-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow \mathbb{R}, \quad x \mapsto \tan x$$

$$[1, +\infty) \rightarrow \mathbb{R}, \quad x \mapsto \frac{1}{x}$$

Recall: Many questions we saw in Supervised Learning are defined on \mathbb{R} , which is not compact.

How to fix this?

1.3. Coercive functions

- Def. (coercive) A function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is coercive if

for every sequence $\{x_m\} \subseteq \mathbb{R}^n$, such that $\lim_{m \rightarrow +\infty} \|x_m\| = +\infty$

it holds that $\lim_{m \rightarrow +\infty} f(x_m) = +\infty$

- Prop. A continuous function is coercive iff

for any $\alpha \in \mathbb{R}$, the set $S_\alpha := \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$ is compact

Pf. " \Rightarrow ": "closedness" (taking limit preserves sign)

for any $\alpha \in \mathbb{R}$, if $\{x_m\} \subseteq S_\alpha$, $\lim_{m \rightarrow +\infty} x_m = x$

then by continuity $f(x) = \lim_{m \rightarrow +\infty} f(x_m) \leq \alpha$, $x \in S_\alpha$

"boundedness" (by contradiction)

if there exists $\alpha_0 \in \mathbb{R}$ s.t. S_{α_0} is not bounded

then there exists a sequence $\{x_m\}$

s.t. $\lim_{m \rightarrow +\infty} \|x_m\| = +\infty$ and $\lim_{m \rightarrow +\infty} f(x_m) \leq \alpha_0$

Contradiction to f being coercive.

" \Leftarrow " left as exercise

- Coercivity + Continuity \Rightarrow Existence of global optimizer

Prop. Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be continuous and coercive,

then there exists $x^* \in \mathbb{R}^n$ s.t. $f(x^*) \leq f(x)$ for any $x \in \mathbb{R}^n$.

Pf left as exercise (use the above prop. and W. extreme value thm)

2. Global minimizer and Uniqueness from Convexity

Q: local information $\xrightarrow{?}$ global minimizer

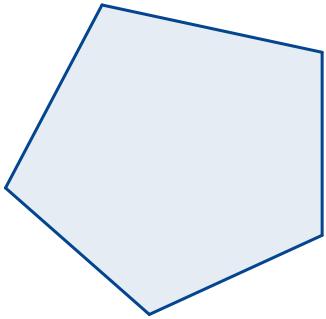
2.1 Convex set and convex function

- Def. (convex set) A set C is convex if

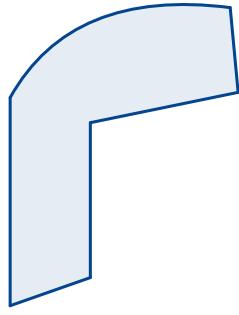
for any $x_1, x_2 \in C$, and any $\theta \in [0, 1]$, $\theta x_1 + (1-\theta)x_2 \in C$

The line segment between any two points in C lies in C .

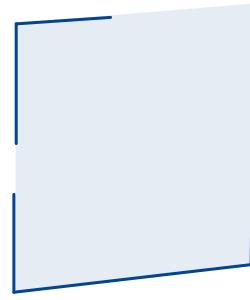
e.g.



cvx



non-cvx



non-cvx

- Def. (convex function) A function f is convex if

$\text{dom}(f)$ is a convex set

and for any $x, y \in \text{dom}(f)$, $\theta \in (0, 1)$,

it holds that $f((1-\theta)x + \theta y) \leq (1-\theta)f(x) + \theta f(y)$

A function f is **strictly convex** if

$\text{dom}(f)$ is a convex set

and for any $x, y \in \text{dom}(f)$, $\theta \in (0, 1)$, $x \neq y$

it holds that $f((1-\theta)x + \theta y) < (1-\theta)f(x) + \theta f(y)$

Examples

- $f: \mathbb{R} \rightarrow \mathbb{R}, x \mapsto e^{ax}, a \in \mathbb{R}$

$$f: \mathbb{R}_{++} := \{x \in \mathbb{R} \mid x > 0\} \rightarrow \mathbb{R}, x \mapsto -\log x$$

$$f: \mathbb{R}_{++} \rightarrow \mathbb{R}, x \mapsto x \log x$$

- $f: \mathbb{R}_{++} \rightarrow \mathbb{R}, x \mapsto x^a, a \geq 1 \text{ or } a \leq 0$

$$f: \mathbb{R} \rightarrow \mathbb{R}, x \mapsto |x|^p, p \geq 1$$

- $f: \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \|x\| \text{ where } \|\cdot\| \text{ is a norm on } \mathbb{R}^n$

$$f: \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \max\{x_1, \dots, x_n\}$$

$$f: \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \log(e^{x_1} + e^{x_2} + \dots + e^{x_n}) \quad \text{differentiable approximation of } \max\{x_1, \dots, x_n\}.$$

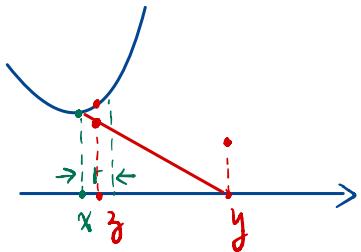
$$f: \mathbb{R} \times \mathbb{R}_{++} \rightarrow \mathbb{R}, (x, y) \mapsto x^2/y$$

Consider

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

2.2. local + convex \rightarrow global

Thm 1. Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be convex. If x^* is a local minimizer of f then x^* is a global minimizer of f .



Pf by contradiction.

Assume there exists y s.t. $f(x^*) > f(y)$

Let $z := x + t(y-x) = (1-t)x + ty, t \in (0,1)$

then $\|z-x\| = t\|y-x\|, f(z) \leq (1-t)f(x) + tf(y) < f(x)$

Contradiction to x^* being a local minimizer \square

2.3. strictly convex \rightarrow uniqueness

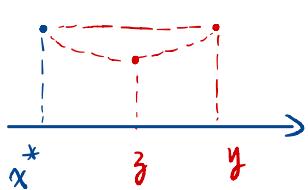
Thm 2. Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be strictly convex. If x^* is a global minimizer of f then x^* is the unique minimizer of f .

Pf by contradiction

Assume there exists $y \neq x^*$ s.t. $f(y) = f(x^*)$

For any $z = (1-t)x^* + t(y), t \in (0,1)$

$$f(z) < (1-t)f(x^*) + tf(y) = f(x^*)$$



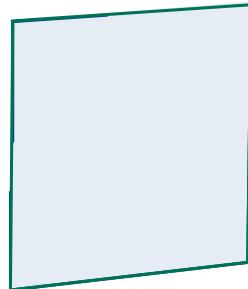
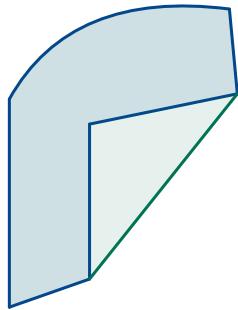
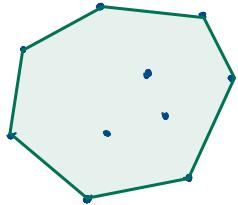
Contradiction to x^* being a global minimizer. \square

3. More about convexity

3.1. convex hull

$$\text{conv } C := \{ \theta_1 x_1 + \dots + \theta_k x_k \mid x_i \in C, \theta_i \geq 0, i=1, \dots, k, \theta_1 + \dots + \theta_k = 1 \}$$

e.g.



convex hull $\text{conv } C$ is always convex.

3.2 Operations that preserve convexity of a set

- $S_1, S_2 \text{ conv} \Rightarrow S_1 \cap S_2 \text{ conv}$ How about $S_1 \cup S_2$?
- $S \subseteq \mathbb{R}^n \text{ conv}, f: \mathbb{R}^n \rightarrow \mathbb{R}^m, x \mapsto Ax+b, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$
 $\Rightarrow f(S) = \{f(x) \mid x \in S\} \subseteq \mathbb{R}^m \text{ conv}$

3.3 Operations that preserve convexity of a function

- $f_1, \dots, f_m: \mathbb{R}^n \rightarrow \mathbb{R}$ convex, $w_1, \dots, w_m \geq 0$
 $\Rightarrow f := \sum_{i=1}^m w_i f_i$ is convex
- $f: \mathbb{R}^n \rightarrow \mathbb{R}, A \in \mathbb{R}^{n \times m}, b \in \mathbb{R}^m, f$ is convex
 $\Rightarrow g: \mathbb{R}^m \rightarrow \mathbb{R}, x \mapsto f(Ax+b)$ is convex
- $f_1, \dots, f_m: \mathbb{R}^n \rightarrow \mathbb{R}$ convex
 $\Rightarrow f := \max \{f_1, \dots, f_m\}$ is convex
- $h: \mathbb{R} \rightarrow \mathbb{R}, g: \mathbb{R}^n \rightarrow \mathbb{R}, f = h \circ g$
 g convex, h convex and nondecreasing $\Rightarrow f$ convex

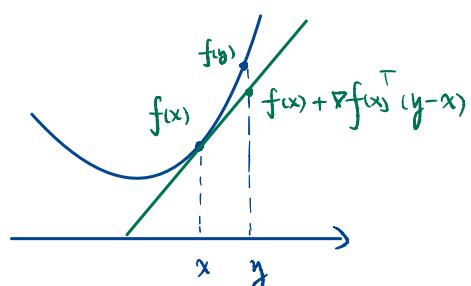
3.4. First and Second order conditions of convex functions

- Prop. 1. Suppose f is differentiable

Then f is wx iff $\text{dom } f$ is convex and for all $x, y \in \text{dom } f$

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) \text{ holds}$$

first-order Taylor approximation



For wx func. the first-order Taylor approximation
(local information)

is a global underestimator of the function.

Pf for $f: \mathbb{R} \rightarrow \mathbb{R}$:

" \Rightarrow " f is $\text{wx} \Rightarrow \text{dom } f$ is wx , thus for any $x, y \in \mathbb{R}, 0 < t \leq 1$

$$x+t(y-x) \in \text{dom } f$$

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

Therefore $f(y) \geq f(x) + \frac{f(x+t(y-x))-f(x)}{t}$ for any t

Taking $t \rightarrow 0$ yields $f(y) \geq f(x) + f'(x)(y-x)$.

" \Leftarrow " For any $x \neq y, 0 \leq \theta \leq 1, z = \theta x + (1-\theta)y$

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

$$\Rightarrow \theta f(x) + (1-\theta)f(y) \geq f(z) + f'(z)[\theta(x-z) + (1-\theta)(y-z)]$$

\square

Proof for high-dimension left as exercise

(hint: choose a direction and convert to 1D)

Prop 2. If f is twice-differentiable then f is cvx iff
 $\text{dom } f$ is cvx and $\nabla^2 f(x) \succeq 0$ for all $x \in \text{dom}(f)$

Pf for $f: \mathbb{R} \rightarrow \mathbb{R}$:

" \Rightarrow " By prop. 1, for any $x, y \in \text{dom}(f)$, $x \neq y$

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

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

$$\Rightarrow f'(x)(y-x) \leq f(y) - f(x) \leq f'(y)(y-x)$$

$$\Rightarrow \frac{f'(y) - f'(x)}{y-x} \geq 0 \quad \xrightarrow{\lim y \rightarrow x} f''(x) \geq 0.$$

" \Leftarrow " By mean value version of Taylor's theorem

$$f(y) = f(x) + f'(x)(y-x) + \frac{1}{2} f''(z)(y-x)^2, \text{ for some } z \in [x,y]$$

$$\geq f(x) + f'(x)(y-x) \text{ since } f''(z) \geq 0$$

- f is strictly convex

Three main approaches to prove the convexity of a function