A GRADUATE COURSE

IN

CONVEX OPTIMIZATION

A GRADUATE COURSE

IN

CONVEX OPTIMIZATION

DDA6110 Notebook

Lecturer

Prof. Yin Zhang

Institute for Data and Decision Analytics

Tex Written By

Mr. Jie Wang

The Chinese University of Hongkong, Shenzhen

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Contents

Acknowledgments

This book is taken notes from the DDA6001 in spring semester, 2020. These lecture notes were taken and compiled in LaTeX by Jie Wang, an undergraduate student in Spring 2020. Prof. Masakiyo Miyazawa has not edited this document. Students taking this course may use the notes as part of their reading and reference materials. This version of the lecture notes were revised and extended for many times, but may still contain many mistakes and typos, including English grammatical and spelling errors, in the notes. It would be greatly appreciated if those students, who will use the notes as their reading or reference material, tell any mistakes and typos to Jie Wang for improving this notebook.

Notations and Conventions

sup	least upper bound
inf	greatest lower bound
$\overline{E}$	closure of E
$f \circ g$	composition
$\limsup(\liminf)$	upper (lower) limit
$L(\mathcal{P}, f), U(\mathcal{P}, f)$	Riemann sums
$\mathcal{R}[a, b]$	classes of Riemann integrable functions on $[a, b]$
$\underline{\int_a^b} f(x)\,\mathrm{d}x, \overline{\int_a^b} f(x)\,\mathrm{d}x$	Riemann integrals
$\langle \boldsymbol{x}, \boldsymbol{y} \rangle$	inner product
$\omega(f; E)$	oscillation of f over set E
$\\| \cdot \\|$	norm
∇f	gradient
$\frac{\partial f}{\partial x_i}, f_{x_i}, f_i, \partial_i f, D_i f$	partial derivatives
$D_{\boldsymbol{v}} f$	directional derivative at direction $\boldsymbol{v}$
$\frac{\partial(y_1, \ldots, y_m)}{\partial(x_1, \ldots, x_n)}$	Jacobian
$\mathbb{S}^n$	set of real symmetric $n \times n$ matrices
$\succ$ $(\succeq)$	positive (semi)-definite
$\mathcal{C}^m$	classes of m-th order continuously differentiable functions
$\mathcal{C}(E; \mathbb{R}^m)$	set of $\mathcal{C}^1$ mapping from E to $\mathbb{R}^m$
$(\mathcal{H}, d)$	metric space

Chapter 1

Week1

1.1. Monday

1.1.1. Introduction to Convex Optimization

The basic optimization model is as follows:

$$\begin{aligned} \min \quad & f(x) : \mathbb{R}^n \to \mathbb{R} \\ \text{s.t.} \quad & x \in X \subseteq \mathbb{R}^n \end{aligned} \tag{1.1}$$

For instance, the constraint X can be union of some inequality constraints:

$$X = \left\{ x \in \mathbb{R}^n \middle| f_i(x) \le b_i, \quad i = 1, \dots, m \right\}$$

We only consider **convex** problems in this course, which means that in (1.1):

- the objective function f is convex;
- the constraint set X is a convex set.

The goal of optimization is to find $x^* \in X$ such that

$$f(x^*) \le f(x), \quad \forall x \in X.$$

or determine whether such an x^* exists or not.

■ **Example 1.1** The least squares problem is a convex problem:

$$\begin{array}{ll} \min & \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2 \\ \text{s.t.} & x \in \mathbb{R}^n \end{array}$$

The optimal solution can be compactly written as

$$\boldsymbol{x}^* = \boldsymbol{A}^{*}\boldsymbol{b}$$

where $\boldsymbol{A}^{*}$ denotes the pseudo inverse of $\boldsymbol{A}$. ■

■ **Example 1.2** The linear programming is also a convex problem:

$$\begin{array}{ll} \min & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} \le b_i, \quad i = 1,2,\dots,m. \end{array}$$

The linear programming is a special case of the conic programming. We will cover conic programming in this course. ■

Definition 1.1 [Line Segments] The line segment between $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$ is defined as

$$\{\theta\boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \mid \theta \in [0,1]\} = \{\boldsymbol{x}_2 + \theta(\boldsymbol{x}_1 - \boldsymbol{x}_2) \mid \theta \in [0,1]\}$$

■

Definition 1.2 [Convex Set in $\mathbb{R}^n$] The set $X \subseteq \mathbb{R}^n$ is convex if for any $x_1, x_2 \in X$,

$$\theta x_1 + (1-\theta)x_2 \in X, \qquad \forall \theta \in [0,1].$$

■

Here are two examples for convex sets over matrices:

$$\mathcal{S}^n = \{\boldsymbol{W} \in \mathbb{R}^n \times \mathbb{R}^n \mid \boldsymbol{W} = \boldsymbol{W}^{\mathrm{T}}\}, \quad \mathcal{S}_+^n = \{\boldsymbol{W} \in \mathcal{S}^n \mid \boldsymbol{W} \succeq 0\},$$

R Sometimes the convex set could be a collection of matrices, but we can vectorize a matrix into a vector.

Definition 1.3 [Convex Function] A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if

- $\text{dorm}(f)$ is a convex set, and
- $\forall x, y \in \text{dorm}(f)$, it holds that for $\theta \in [0,1]$,

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y),$$

which means that the secant line betweeen any two points is above the function.

Proposition 1.1 $f(x) = \max_i f_i(x)$ is convex if $f_i(x)$ is convex for each i.

Proof. For any $\theta \in [0,1]$, the following inequality holds:

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \max_i \ f_i(\theta x + (1-\theta)y) \\ &\le \max_i \ \theta f_i(x) + (1-\theta) f_i(y) \\ &\le \max_i \ \theta f_i(x) + \max_i \ (1-\theta) f_i(y) \\ &= \theta f(x) + (1-\theta) f(y) \end{aligned}$$

∎

Example 1.3 Define the function

$$\begin{aligned} & f : \mathbb{S}^n \to \mathbb{R} \\ \text{with} \quad & f(\boldsymbol{X}) = \lambda_{\max}(\boldsymbol{X}) \triangleq \max_{\|\boldsymbol{v}\|=1} \ \boldsymbol{v}^{\mathrm{T}} \boldsymbol{X} \boldsymbol{v} \end{aligned}$$

This function is convex since f can be written as the maximization of a collection of affines (in terms of $\boldsymbol{X}$):

$$f(\boldsymbol{X}) = \max_{\|\boldsymbol{v}\|=1} f_{\boldsymbol{v}}(\boldsymbol{X}), \qquad \text{with } f_{\boldsymbol{v}}(\boldsymbol{X}) = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}.$$

1.2. Wednesday

Convex Programming (CP):

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & x \in X \subseteq \mathbb{R}^n \end{aligned}$$

with f and X being convex. The standard constraint set X can be written as:

$$X = \left\{ x \middle| f_i(x) \le b_i, \ \ i = 1,2,3,\dots,m, \quad Ax = c \right\}$$

with $A \in \mathbb{R}^{p \times n}$, $p < n$, and f_i's are convex.

Proposition 1.2 The constraint set

$$X = \left\{ x \middle| f_i(x) \le b_i, \ \ i = 1,\dots,m \right\}$$

is **convex** if all f_i's are convex.

Proof. For any $x_1, x_2 \in X$ and $\theta \in [0,1]$, verify that

$$\begin{aligned} f_i(\theta x_1 + (1-\theta)x_2) &\le \theta f_i(x_1) + (1-\theta) f_i(x_2) \\ &\le \theta b_i + (1-\theta) b_i = b_i \end{aligned}$$

where the first inequality is because of the convexity, and the second inequality is because that $x_i \in X, i = 1,2$. As a result, $\theta x_1 + (1-\theta)x_2 \in X$, i.e., X is a convex set. ■

Proposition 1.3 If $\{X_i\}_{i \in \mathcal{I}}$ is a collection of convex sets, then the union $X = \bigcap_{i \in \mathcal{I}}$ is convex.

Proof. Consider any $x_1, x_2 \in X$, i.e., $x_1, x_2 \in X_i, \forall i \in \mathcal{I}$, then by the convexity of X_is, the line segment

$$\theta x_1 + (1-\theta)x_2 \in X_i, \forall i \in \mathcal{I},$$

i.e., $\theta x_1 + (1-\theta)x_2 \in X$. ■

Definition 1.4 [Epigraph] Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, the epigraph of f is a set defined as

$$\text{epi}(f) = \left\{ (x,t) \middle| f(x) \le t, x \in \text{dorm}(f) \right\} \in \mathbb{R}^{n+1}.$$

Proposition 1.4 The function f is convex if and only if the epigraph epi(f) is convex.

Proof. We only talk about the proof on the forward direction. Consider any $(x_1,t_1),(x_2,t_2) \in$ epi(f), then we show that $\theta(x_1,t_1) + (1-\theta)(x_2,t_2) \in$ epi(f) for any $\theta \in [0,1]$:

$$\begin{aligned} f(\theta x_1 + (1-\theta)x_2) &\le \theta f(x_1) + (1-\theta) f(x_2) \\ &\le \theta t_1 + (1-\theta)t_2, \end{aligned}$$

which proves the desired result. ■

Let's start to introduce some useful convex sets.

Definition 1.5 [Combination of two points] Consider the combination of two points x_1, x_2:

$$\theta_1 x_1 + \theta_2 x_2,$$

- When $\theta_1, \theta_2 \in \mathbb{R}$, it is called an affine combination; the resulted space is an **affine space**.
- When $\theta_1, \theta_2 \in \mathbb{R}_+$, it is called a non-negative combination; the resulted space is a **convex cone**.
- When $\theta_2 = 1 - \theta_1, \theta_1 \in [0,1]$, it is called a convex combination; the resulted space is a **convex set**.

Definition 1.6 [Affine Set] The standard form of an affine space is $\{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$.

Definition 1.7 [Cone] A set is called a cone with a vertex at the origin if for any $\boldsymbol{x} \in X$, $a\boldsymbol{x} \in X$ for any $a \geq 0$.

The standard form of a convex conic programming is the following:

$$\begin{aligned} \min \quad & \langle \boldsymbol{C}, \boldsymbol{X} \rangle \\ \text{s.t.} \quad & \langle \boldsymbol{A}_i, \boldsymbol{X} \rangle = \boldsymbol{B}_i, \quad i = 1, 2, \ldots, m \\ & \boldsymbol{X} \in \mathcal{K} \end{aligned}$$

where $\mathcal{K}$ is a convex cone.

One special form of conic programming is the linear programming:

$$\begin{aligned} \min \quad & \langle \boldsymbol{c}, \boldsymbol{x} \rangle \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

Another is the second-order cone programming:

$$\mathcal{K} = \{(\boldsymbol{x}, t) \mid \|\boldsymbol{x}\|_2 \leq t\}$$

The semidefinite programming also belongs to the case of conic programming:

$$\mathcal{K} = \mathcal{S}_+^n \triangleq \{\boldsymbol{X} \in \mathbb{S}^n \mid \boldsymbol{X} \succeq 0, \forall \boldsymbol{v} \in \mathbb{R}^n\}$$

Convex hull of a set S:

It is the smallest convex set containing S, called $\text{conv}(S)$.

Definition 1.8 [Polyhedron]

- A hyperplane in $\mathbb{R}^n$ can be written as the form $\{\boldsymbol{x} \mid \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} = b\}$.
- Half space: $\{\boldsymbol{x} \mid \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \leq \boldsymbol{b}\}$.
- Polyhedron: a intersection of finite hyperplanes and half spaces.

- Ellipsoid:

$$\left\{\boldsymbol{x}\,\middle|\,(\boldsymbol{x}-\boldsymbol{x}_c)^{\mathrm{T}}\boldsymbol{P}^{-1}(\boldsymbol{x}-\boldsymbol{x}_c)\leq 1,\ \boldsymbol{P}\succ 0\right\}=\{\boldsymbol{x}\mid \|\boldsymbol{P}^{-1/2}(\boldsymbol{x}-\boldsymbol{x}_c)\|_2^2\leq 1\}$$

$$=\{\boldsymbol{x}\mid \boldsymbol{x}=\boldsymbol{x}_c+\boldsymbol{A}\boldsymbol{u},\|\boldsymbol{u}\|_2\leq 1\}$$