A GRADUATE COURSE

IN

OPTIMIZATION

A GRADUATE COURSE

IN

OPTIMIZATION

CIE6010 Notebook

Prof. Yin Zhang

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 from the CIE6010 in fall semester, 2018.

CUHK(SZ)

Notations and Conventions

X	Set
$\inf X \subseteq \mathbb{R}$	Infimum over the set X
$\mathbb{R}^{m\times n}$	set of all $m \times n$ real-valued matrices
$\mathbb{C}^{m\times n}$	set of all $m \times n$ complex-valued matrices
x_i	ith entry of column vector $\boldsymbol{x}$
a_{ij}	(i,j)th entry of matrix $\boldsymbol{A}$
$\boldsymbol{a}_i$	ith column of matrix $\boldsymbol{A}$
$\boldsymbol{a}_i^{\mathrm{T}}$	ith row of matrix $\boldsymbol{A}$
$\mathbb{S}^n$	set of all $n \times n$ real symmetric matrices, i.e., $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $a_{ij} = a_{ji}$ for all i,j
$\mathbb{H}^n$	set of all $n \times n$ complex Hermitian matrices, i.e., $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ and $\bar{a}_{ij} = a_{ji}$ for all i,j
$\boldsymbol{A}^{\mathrm{T}}$	transpose of $\boldsymbol{A}$, i.e, $\boldsymbol{B} = \boldsymbol{A}^{\mathrm{T}}$ means $b_{ji} = a_{ij}$ for all i,j
$\boldsymbol{A}^{\mathrm{H}}$	Hermitian transpose of $\boldsymbol{A}$, i.e, $\boldsymbol{B} = \boldsymbol{A}^{\mathrm{H}}$ means $b_{ji} = \bar{a}_{ij}$ for all i,j
$\mathrm{trace}(\boldsymbol{A})$	sum of diagonal entries of square matrix $\boldsymbol{A}$
$\mathbf{1}$	A vector with all 1 entries
$\mathbf{0}$	either a vector of all zeros, or a matrix of all zeros
$\boldsymbol{e}_i$	a unit vector with the nonzero element at the ith entry
$\mathcal{C}(\boldsymbol{A})$	the column space of $\boldsymbol{A}$
$\mathcal{R}(\boldsymbol{A})$	the row space of $\boldsymbol{A}$
$\mathcal{N}(\boldsymbol{A})$	the null space of $\boldsymbol{A}$
$\mathrm{Proj}_{\mathcal{M}}(\boldsymbol{A})$	the projection of $\boldsymbol{A}$ onto the set $\mathcal{M}$

Chapter 10

Week10

10.1. Monday

Announcement. No assignment in this week, so you may take a break. However, in next week new assignments and projects will be updated, which requires you to apply penalty algorithms.

Theorem 10.1 — Farka's Lemma. Let $\boldsymbol{a}_1, \dots, \boldsymbol{a}_r \in \mathbb{R}^n$, and

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \boldsymbol{a}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_r^{\mathrm{T}} \end{pmatrix},$$

then for any $\boldsymbol{c} \in \mathbb{R}^n$,

$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{y} \le 0, \quad \forall \boldsymbol{y} \text{ such that } \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \le 0, \tag{10.1a}$$

if and only if

$$\boldsymbol{c} = \boldsymbol{A}\boldsymbol{u}, \forall \boldsymbol{u} \ge 0, \boldsymbol{u} \in \mathbb{R}^r \tag{10.1b}$$

R The interpretation is that the vector $\boldsymbol{c}$ has more than 90 degrees angle with all vectors $\boldsymbol{a}_i$ in the polar cone, if and only if $\boldsymbol{c}$ is in the polar cone.

Proof. To show the converse, we have

$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y},$$

with $\boldsymbol{u} \geq 0, \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \leq 0$, and thereofre $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{y} \leq 0$. ■

Proof. ■

Convex Program.

$$\begin{aligned} \min \quad & f(x) \\ \text{such that} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & g(\boldsymbol{x}) \leq 0 \end{aligned}$$

with f, g to be convex. The Lagrangian function is given by:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) + \boldsymbol{\mu}^{\mathrm{T}} g(\boldsymbol{x}),$$

with $\boldsymbol{\mu} \geq 0$. This function is convex in $\boldsymbol{x}$. Therefore the dual function is given by:

$$Q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x}, \boldsymbol{\mu} \geq 0} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$

Proposition 10.1 — **Weak Deality.**

$$Q(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq f(\boldsymbol{x})$$

for dual feasible $\boldsymbol{\lambda}, \boldsymbol{\mu}$ and primal feasible $\boldsymbol{x}$.

We are curious on the tightest lower bound on LHS, thus maximizing the dual function to obtain the dual program:

$$\begin{aligned} \max \quad & Q(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{such that} \quad & \boldsymbol{\mu} \geq 0 \end{aligned}$$

Proposition 10.2 — **Strong Duality.** For convex programming, we have

$$d^* = p^*,$$

with d^*, p^* to be the optimal value from dual and primal problems, respectively.

QC: one of the followings is satisfied

- $g_i(\boldsymbol{x})$ are linear
- $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, g(\boldsymbol{x}) \leq 0$
- Regularity

Under QC, the primal and dual could attain optimality togethoer iff

- $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, g(\boldsymbol{x}) \leq 0$
- $\boldsymbol{\mu} \geq 0$
- $\boldsymbol{\mu} \circ g(\boldsymbol{x}) = \boldsymbol{0}$

We have derive the dual formula for linear programming, but how about the quadratic programming?

■ **Example 10.1**

$$\begin{aligned} p^* = \min \quad & \tfrac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, \quad \boldsymbol{Q} \succ 0 \\ \text{such that} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \end{aligned} \tag{10.2}$$

The Lagrangian function $L(\boldsymbol{x},\boldsymbol{\mu}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$, and therefore

$$Q(\boldsymbol{\mu}) = \min_{\boldsymbol{x},\boldsymbol{\mu}\geq 0} L(\boldsymbol{x},\boldsymbol{\mu}) \tag{10.3}$$

The optimality condition implies that

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\mu}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\mu} = 0 \implies \boldsymbol{x} = -\boldsymbol{Q}^{-1}(\boldsymbol{c} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\mu})$$

Thus substituting optimal $\boldsymbol{x}$ into (10.3), we derive

$$Q(\boldsymbol{\mu}) = -\frac{1}{2}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\mu} - \boldsymbol{t}^{\mathrm{T}}\boldsymbol{\mu} + \text{constant}$$

Thus we derive the dual program:

$$\begin{aligned} d^* = \min \quad & \tfrac{1}{2}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\mu} + \boldsymbol{t}^{\mathrm{T}}\boldsymbol{\mu}, \\ \text{such that} \quad & \boldsymbol{\mu} \geq 0 \end{aligned} \tag{10.4}$$

where $P := AQ^{-1}A^{T}$. ■

10.1.1. Penalty Algorithms

Logarithm Penalty. Consider the inequality constraint problem

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ & g_i(\boldsymbol{x}) \leq 0 \end{aligned}$$

The Barrier problem is given by:

$$\min f(x) - \mu \sum_i \log(-g_i(\boldsymbol{x})), \quad \mu > 0$$

As $\mu \to 0$, $x(\mu)$ converges to the optimal solution. We pick big μ at first and obtain a good initial guess, and then we continue to decrease μ.

Quadratic Penalty. For the constraint problem

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ & h(\boldsymbol{x}) = 0 \\ & \boldsymbol{x} \in X \end{aligned}$$

The quadratic penalty algorithm aims to solve

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) + \lambda^{T} h(\boldsymbol{x}) + \tfrac{c}{2} \|h(\boldsymbol{x})\|_2^2 \\ & \boldsymbol{x} \in X, \end{aligned}$$

where λ is **bounded**. Conversely, as $c \to \infty$, $\boldsymbol{x}(c)$ converges to the optimal solution. We pick small c at first and obtain a good initial guess, and then we continus to increase c.

■ **Example 10.2**

$$\begin{aligned} \min \quad & \tfrac{1}{2}(x_1^2 + x_2^2) \\ & x_1 = 1 \end{aligned}$$

The quadratic penalty function is

$$L_c(x) = \frac{1}{2}(x_1^2 + x_2^2) + \lambda(x_1 - 1) + \frac{c}{2}(x_1 - 1)^2$$

and therefore

$$\nabla_{\boldsymbol{x}} L_c(\boldsymbol{x}) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c \begin{pmatrix} x_1 - 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which follows that

$$x_1(\lambda, c) = \tfrac{c-\lambda}{c+1}, \quad x_2(\lambda, c) = 0$$

We can apply two algorithm to converge to optimal solution:

1. Quadratic Pendlty Method: As $c \to \infty$ with λ bounded, we derive $x_1(\lambda, c) \to 1$.
2. Lagrangian Multiplier Method: We set $\nabla L(\boldsymbol{x}, \lambda) = 0$ to obtain an appropriate $\lambda^* = -1$. As $\lambda \to \lambda^*$, we obtain $x_1(\lambda, c) \to 1$ for $c > 1$ (the key for this kind of algorithm is to choose big c).

Such an algorithm can also be applied for the non-convex problem, e.g.,

$$\begin{aligned} \min \quad & \tfrac{1}{2}(-x_1^2 + x_2^2) \\ & x_1 = 1 \end{aligned}$$