

A GRADUATE COURSE
IN
OPTIMIZATION

A GRADUATE COURSE
IN
OPTIMIZATION
CIE6010 Notebook

Prof. Yin Zhang

The Chinese University of Hong Kong, Shenzhen



香港中文大學(深圳)

The Chinese University of Hong Kong, Shenzhen

Contents

Acknowledgments	ix
Notations	xi
1 Week1	1
1.1 Monday	1
1.1.1 Introduction to Optimizaiton	1
1.2 Wednesday	2
1.2.1 Reviewing for Linear Algebra	2
1.2.2 Reviewing for Calculus	2
1.2.3 Introduction to Optimization	3
2 Week2	7
2.1 Monday	7
2.1.1 Reviewing and Announments	7
2.1.2 Quadratic Function Case Study	8
2.2 Wednesday	11
2.2.1 Convex Analysis	11
3 Week3	17
3.1 Wednesday	17
3.1.1 Convex Analysis	17
3.1.2 Iterative Method	18
3.2 Thursday	22
3.2.1 Announcement	22
3.2.2 Sparse Large Scale Optimization	22

4	Week4	27
4.1	Wednesday	27
4.1.1	Comments for MATLAB Project	27
4.1.2	Local Convergence Rate	28
4.1.3	Newton's Method	29
4.1.4	Tutorial: Introduction to Convexity	30
5	Week5	33
5.1	Monday	33
5.1.1	Review	33
5.1.2	Existence of solution to Quadratic Programming	36
5.2	Wednesday	39
5.2.1	Comments about Newton's Method	39
5.2.2	Constant Step-Size Analysis	40
6	Week6	45
6.1	Monday	45
6.1.1	Announcement	45
6.1.2	Introduction to Quasi-Newton Method	45
6.1.3	Constrained Optimization Problem	46
6.1.4	Announcement on Assignment	47
6.1.5	Introduction to Stochastic optimization	49
6.2	Tutorial: Monday	49
6.2.1	LP Problem	49
6.2.2	Gauss-Newton Method	50
6.2.3	Introduction to KKT and CQ	51
6.3	Wednesday	52
6.3.1	Review	52
6.3.2	Dual-Primal of LP	53

7	Week7	57
7.1	Monday	57
7.1.1	Announcement	57
7.1.2	Recap about linear programming	57
7.1.3	Optimization over convex set	60
7.2	Wednesday	62
7.2.1	Motivation	62
7.2.2	Convex Projections	63
7.2.3	Feasible direction method	65
8	Week8	69
8.1	Monday	69
8.1.1	Constraint optimization	70
8.1.2	Inequality Constraint Problem	71
8.2	Monday Tutorial: Review for CIE6010	71
9	Week9	79
9.1	Monday	79
9.1.1	Reviewing for KKT	79
9.2	Monday Tutorial: Reviewing for Mid-term	82
10	Week10	83
10.1	Monday	83
10.1.1	Penalty Algorithms	86

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	i th entry of column vector \mathbf{x}
a_{ij}	(i, j) th entry of matrix \mathbf{A}
\mathbf{a}_i	i th column of matrix \mathbf{A}
\mathbf{a}_i^T	i th row of matrix \mathbf{A}
\mathbb{S}^n	set of all $n \times n$ real symmetric matrices, i.e., $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $a_{ij} = a_{ji}$ for all i, j
\mathbb{H}^n	set of all $n \times n$ complex Hermitian matrices, i.e., $\mathbf{A} \in \mathbb{C}^{n \times n}$ and $\bar{a}_{ij} = a_{ji}$ for all i, j
\mathbf{A}^T	transpose of \mathbf{A} , i.e, $\mathbf{B} = \mathbf{A}^T$ means $b_{ji} = a_{ij}$ for all i, j
\mathbf{A}^H	Hermitian transpose of \mathbf{A} , i.e, $\mathbf{B} = \mathbf{A}^H$ means $b_{ji} = \bar{a}_{ij}$ for all i, j
$\text{trace}(\mathbf{A})$	sum of diagonal entries of square matrix \mathbf{A}
$\mathbf{1}$	A vector with all 1 entries
$\mathbf{0}$	either a vector of all zeros, or a matrix of all zeros
\mathbf{e}_i	a unit vector with the nonzero element at the i th entry
$\mathcal{C}(\mathbf{A})$	the column space of \mathbf{A}
$\mathcal{R}(\mathbf{A})$	the row space of \mathbf{A}
$\mathcal{N}(\mathbf{A})$	the null space of \mathbf{A}
$\text{Proj}_{\mathcal{M}}(\mathbf{A})$	the projection of \mathbf{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 $\mathbf{a}_1, \dots, \mathbf{a}_r \in \mathbb{R}^n$, and

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

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

$$\mathbf{c}^T \mathbf{y} \leq 0, \quad \forall \mathbf{y} \text{ such that } \mathbf{A}^T \mathbf{y} \leq 0, \quad (10.1a)$$

if and only if

$$\mathbf{c} = \mathbf{A} \mathbf{u}, \forall \mathbf{u} \geq 0, \mathbf{u} \in \mathbb{R}^r \quad (10.1b)$$

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

Proof. To show the converse, we have

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

with $\mathbf{u} \geq 0, \mathbf{A}^T \mathbf{y} \leq 0$, and therefore $\mathbf{c}^T \mathbf{y} \leq 0$. ■

Proof. ■

Convex Program.

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

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

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

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

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

Proposition 10.1 — **Weak Duality.**

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

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

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

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

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

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

■ **Example 10.1**

$$\begin{aligned} p^* = \min \quad & \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{c}^T \mathbf{x}, \quad \mathbf{Q} \succ 0 \\ \text{such that} \quad & \mathbf{Ax} \leq \mathbf{b} \end{aligned} \quad (10.2)$$

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

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

The optimality condition implies that

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

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

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

Thus we derive the dual program:

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

where $\mathbf{P} := \mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T$. ■

10.1.1. Penalty Algorithms

Logarithm Penalty. Consider the inequality constraint problem

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

The Barrier problem is given by:

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

As $\mu \rightarrow 0$, $x(\mu)$ 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

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

The quadratic penalty algorithm aims to solve

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

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

■ Example 10.2

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

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

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

which follows that

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

We can apply two algorithm to converge to optimal solution:

1. Quadratic Penalty Method: As $c \rightarrow \infty$ with λ bounded, we derive $x_1(\lambda, c) \rightarrow 1$.
2. Lagrangian Multiplier Method: We set $\nabla L(\mathbf{x}, \lambda) = 0$ to obtain an appropriate $\lambda^* = -1$. As $\lambda \rightarrow \lambda^*$, we obtain $x_1(\lambda, c) \rightarrow 1$ 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.,

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

