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 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 Constrainted 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

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 7

Week7

7.1. Monday

7.1.1. Announcement

- The new homework is assigned. You are required to write a code for linear programming. The detail will be announced in tutorial.

- The mid-term will be some written problem and computation porblem. We will cover contents in Chapter 1,2,3,4 in the textbook.

7.1.2. Recap about linear programming

The standard form of linear programming is

$$\begin{aligned} \min \quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{such that} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq 0 \end{aligned}$$

Define the Lagrange function

$$\begin{aligned} L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) &= \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{y}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) - \boldsymbol{z}^{\mathrm{T}}\boldsymbol{x} \\ &= (\boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{z})^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} \end{aligned}$$

From this Lagrange function we define a dual function

$$Q(\boldsymbol{y},\boldsymbol{z}) = \inf_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) = \begin{cases} \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}, & \text{if } \boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{z} = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

For any feasible $\boldsymbol{x}$ and $\boldsymbol{z} \geq 0$, we always have

$$Q(\boldsymbol{y},\boldsymbol{x}) \leq \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{z}^{\mathrm{T}}\boldsymbol{x} \leq \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$$

Moreover,

$$\max_{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}+\boldsymbol{z}=\boldsymbol{c},\boldsymbol{z}\geq 0} \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} = \sup_{\boldsymbol{z}\geq 0} Q(\boldsymbol{y},\boldsymbol{z}) \leq \min_{\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geq 0} \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$$

- The weak duality says that the optimal dual objective is always no more than the optimal primal objective.

- The strong duality says that the optimal dual objective is equal to the optimal primal objective.

Theorem 7.1 The LP optimality condition is given by:

1. Primal-Feasibility:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq 0$$

2. Dual-Feasibility:

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{z} = \boldsymbol{c}, \boldsymbol{z} \geq 0$$

3. Complementarity/Strong-Duality:

$$\boldsymbol{x} \circ \boldsymbol{z} = 0$$

R Hence, the LP is essentially solving the linear systems. The Assignment 5 aims to solve the linear programming problem via primal-dual algorithm. Furthermore, it is called the primal-dual interior-point method (pdipm). The idea of interior-point method is to set the barrier:

$$\boldsymbol{x} \circ \boldsymbol{z} = u * \mathbf{1},$$

and with $u \to 0$, $\boldsymbol{x} \circ \boldsymbol{z} = 0$.

Something about A5. During the implementation, you need to solve the system

$$\underbrace{(A(Z^{-1}X)A^{\mathrm{T}})}_{\boldsymbol{B}} dy = rh_s$$

Given that the A is sparse, we call the command chol to decompose $\boldsymbol{M} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$ with $\boldsymbol{R}$ to be upper triangular. Then we obtain the solution

$$dy = R/(R'/rhs)$$

However, it is still time-consuming. We can accelerate it by re-arrange the order of the linear system by the command symamd. The detailed handle discussing the advantage for re-ordering is attached in assignment.

Incremental Problem. When solving the least squares problem

$$\min \frac{1}{2} \sum_{i=1}^{m} (\boldsymbol{a}_i^{\mathrm{T}} \boldsymbol{x} - \boldsymbol{b}_i)^2 = \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2,$$

the optimal analytic solution is given by $\boldsymbol{x}^* = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$. Let's try the numerical method:

$$\boldsymbol{x}^{k+1} \leftarrow \boldsymbol{x}^k - \alpha^k \boldsymbol{g}^1 - \alpha^k g^2 - \cdots - \alpha^k \boldsymbol{g}^m,$$

where $\alpha^k = \frac{\theta}{k}$. Prof.YZ's α is greater than 10. Although we have the analytic solution, it may not necessarily better than the numerical one since it will perturbed by the noise.

7.1.3. Optimization over convex set

For the optimization problem

$$\begin{aligned} \min \quad & f(x) \\ \text{such that} \quad & x \in X \end{aligned} \tag{7.1}$$

with X to be convex, the necessary conditon for optimality is

$$\nabla^{\mathrm{T}} f(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \quad \forall x \in X$$

If f is convex, then it is indeed the necessary and sufficient condition.

Example. For $X = \{x \mid x \geq 0\}$, we find

$$(\nabla f(\boldsymbol{x}^*))_i = \begin{cases} \dfrac{\partial f(\boldsymbol{x}^*)}{\partial x_i} \geq 0, & x_i^* = 0 \\ \dfrac{\partial f(\boldsymbol{x}^*)}{\partial x_i} = 0, & x_i^* > 0 \end{cases}$$

Project. Given $z \in \mathbb{R}^n$, the least square problem

$$\begin{aligned} \min \quad & \|\boldsymbol{x} - \boldsymbol{z}\|_2 \\ & \boldsymbol{x} \in X \end{aligned}$$

is essentially the projection problem:

$$[\boldsymbol{z}]^+ = \boldsymbol{x}^* = \arg\min_{\boldsymbol{x} \in X} \|\boldsymbol{x} - \boldsymbol{z}\|$$

Proposition 7.1 — Non-expansive.

$$\|[\boldsymbol{z}_1]^+ - [\boldsymbol{z}_2]^+\| \leq \|\boldsymbol{z}_1 - \boldsymbol{z}_2\|$$

with X to be a convex set.

Projection-based optimality condition. Applying projection property, we derive alternative optimality condition for (7.1):

Proposition 7.2 $\boldsymbol{x}^*$ is a stationary point iff

$$\boldsymbol{x}^* = [\boldsymbol{x}^* - \alpha \nabla f(\boldsymbol{x}^*)]+,$$

for $\forall \alpha > 0$