

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	Duality Theory	83
10.1.2	Penalty Algorithms	86
10.2	Wednesday	89
10.2.1	Introduction to penalty algorithms	89
10.2.2	Convergence Analysis	90

11	Week11	91
11.1	Monday	91
11.1.1	Augmented Lagrangian	91
11.1.2	ADMM	94
11.2	Monday Tutorial	94
11.2.1	Implicit Function Theorem	96
11.2.2	Trust Region	96

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}

10.2. Wednesday

In this lecture we will discuss the main penalty algorithms formally.

10.2.1. Introduction to penalty algorithms

Given the equality constraint problem

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ & h(\mathbf{x}) = 0 \\ & \mathbf{x} \in X \subseteq \mathbb{R}^n \end{aligned} \tag{10.10}$$

The augmented Lagrangian function is given by:

$$L_c(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T h(\mathbf{x}) + \frac{c}{2} \|h(\mathbf{x})\|^2 \tag{10.11}$$

There are two penalty algorithms formally listed below:

Quadratic Penalty (Courant, 1943).

$$\begin{aligned} \mathbf{x}^r &\leftarrow \arg \min_{\mathbf{x} \in X} L_{c^r}(\mathbf{x}, \boldsymbol{\lambda}^r, c^r) \\ \text{increase } c^{r+1} &> c^r \rightarrow \infty, \quad \|\boldsymbol{\lambda}^r\| < +\infty \end{aligned}$$

Augmented Lagrangian Multiplier Method.

$$\begin{aligned} \mathbf{x}^r &\leftarrow \arg \min_{\mathbf{x} \in X} L_c(\mathbf{x}, \boldsymbol{\lambda}^r) \\ \boldsymbol{\lambda}^{r+1} &\leftarrow \boldsymbol{\lambda}^r + c h(\mathbf{x}) \end{aligned}$$

where c is sufficiently large in general, but not goes to infinite.

- The reason why the λ converges to the optimal Lagrange multiplier will be discussed later.
- Why do we need to take sufficiently large c ? Let's raise an example first.

■ **Example 10.3** Given the problem

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

with optimal solution $\mathbf{x}^* = (1, 0)$; $\lambda^* = 1$. The augmented Lagrangian function is given by:

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

Applying the optimality condition, we derive

$$\mathbf{x}^*(c, \lambda) = \begin{pmatrix} \frac{c-\lambda}{c-1} \\ 0 \end{pmatrix}$$

The second order necessary condition should be

$$\nabla_{\mathbf{xx}}^2 L_c = \begin{pmatrix} c-1 & 0 \\ 0 & c+1 \end{pmatrix} \succeq 0$$

Therefore, from this example we can see that if c is not large enough, the second order optimality condition will be violated. Later we will give a formal proof for this statement. ■

10.2.2. Convergence Analysis

Theorem 10.2 Given the sequence of optimization problem

$$\min \quad L_{c^k}(\mathbf{x}, \boldsymbol{\lambda}^k), \mathbf{x} \in X \tag{10.12}$$

with associated local minimum \mathbf{x}^k . Suppose $\{c^k\}$ is monotone increasing to infinite, and $\{\boldsymbol{\lambda}^k\}$ is bounded, then every limit point in the sequence $\{\mathbf{x}^k\}$ is the global minimum for original problem (10.10)

Proof. 1. Firstly note that the optimal value f^* for (10.10) is the infimum of the

optimal value for (10.12), i.e.,

$$\begin{aligned}
f^* &= \inf_{h(\mathbf{x})=0, \mathbf{x} \in X} f(\mathbf{x}) \\
&= \inf_{h(\mathbf{x})=0, \mathbf{x} \in X} \left\{ f(\mathbf{x}) + (\boldsymbol{\lambda}^k)^\top h(\mathbf{x}) + \frac{\mathbf{c}^k}{2} \|h(\mathbf{x})\|^2 \right\} \\
&= \inf_{h(\mathbf{x})=0, \mathbf{x} \in X} L_{\mathbf{c}^k}(\mathbf{x}, \boldsymbol{\lambda}^k)
\end{aligned}$$

We study the lower bound for $L_{\mathbf{c}^k}(\mathbf{x}, \boldsymbol{\lambda}^k)$. By definition,

$$L_{\mathbf{c}^k}(\mathbf{x}, \boldsymbol{\lambda}^k) \geq L_{\mathbf{c}^k}(\mathbf{x}^k, \boldsymbol{\lambda}^k) \quad (10.13)$$

Taking infimum over \mathbf{x} both sides, we obtain:

$$f^* \geq L_{\mathbf{c}^k}(\mathbf{x}^k, \boldsymbol{\lambda}^k) \quad (10.14)$$

2. Then we show that the limit point $\bar{\mathbf{x}}$ is such that $f^* \geq f(\bar{\mathbf{x}})$ and $\bar{\mathbf{x}}$ is feasible by using (10.14). Suppose $\{\mathbf{x}^k, \boldsymbol{\lambda}^k\} \rightarrow \{\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}}\}$, and by taking limsup both sides for (10.14), we obtain:

$$f(\bar{\mathbf{x}}) + (\bar{\boldsymbol{\lambda}})^\top h(\bar{\mathbf{x}}) + \limsup_{k \rightarrow \infty} \frac{\mathbf{c}^k}{2} \|h(\mathbf{x}^k)\|^2 \leq f^* \quad (10.15)$$

Since $\mathbf{c}^k \rightarrow \infty$ and $\|h(\mathbf{x})\|^2 \geq 0$, and the LHS of (10.15) is bounded above, we derive $h(\mathbf{x}^k) \rightarrow 0$, and in particular,

$$h(\bar{\mathbf{x}}) = 0 \quad (10.16)$$

Combining (10.15) and (10.16), we conclude that $f^* \geq f(\bar{\mathbf{x}})$ and $\bar{\mathbf{x}}$ is feasible, i.e., $\bar{\mathbf{x}}$ is global minimum. ■

