

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	93
11.1	Monday	93
11.1.1	Equality Constraint Problem	93
11.1.2	ADMM	96
11.2	Wednesday	97
11.2.1	Comments on Assignment 6	97
11.2.2	Inequality Constraint Problem	98
11.2.3	Non-smooth unconstraint problem	99
12	Week12	101
12.1	Monday	101
12.1.1	Comments on Final Project	101
12.1.2	Trust Region Method	102
12.2	Monday Tutorial	103
12.2.1	Sub-gradient	104
12.3	Wednesday	107
12.3.1	Gradient Projection	109

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}

12.3. Wednesday

Given the unconstrained problem $\min f(\mathbf{x})$, at \mathbf{x}^k ,

$$\begin{aligned} \min \quad & m_k(\mathbf{p}) = f(\mathbf{x}^k) + \nabla^T f(\mathbf{x}^k) \mathbf{p} + \frac{1}{2} \mathbf{p}^T \mathbf{B}_k \mathbf{p} \\ \text{such that} \quad & \|\mathbf{p}\| \leq \Delta \end{aligned}$$

In our project we need to solve it very accurately.

Trust Region sub-problem.

$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{p}^T \mathbf{B} \mathbf{p} + \mathbf{g}^T \mathbf{p} \\ \text{such that} \quad & \|\mathbf{p}\| \leq \Delta \end{aligned}$$

Necessary condition: \mathbf{p}^* is a local minimum implies that (KKT condition)

$$(\mathbf{B} + \lambda \mathbf{I}) \mathbf{p}^* = -\mathbf{g} \tag{12.8a}$$

$$\lambda \geq 0 \tag{12.8b}$$

$$\lambda(\Delta - \|\mathbf{p}\|) = 0 \tag{12.8c}$$

$$\|\mathbf{p}\| \leq \Delta \tag{12.8d}$$

The second order necessary condition is

$$\mathbf{v}^T (\mathbf{B} + \lambda \mathbf{I}) \mathbf{v} \geq 0, \forall \mathbf{v} \perp \mathbf{p}^*$$

In addition, we replace the 2nd order condition with a little bit stronger condition.

Proposition 12.2 $\mathbf{B} + \lambda \mathbf{I} \succeq 0$ together with KKT condition are iff for \mathbf{p}^* to be a global minimum.

Proof. Assume all the condition holds. Then (12.8a) implies that \mathbf{p}^* is a global minimum of

$$\hat{m}(\mathbf{p}) = m(\mathbf{p}) + \frac{\lambda}{2} \mathbf{p}^T \mathbf{p}$$

since

$$\nabla^2 \hat{m}(\mathbf{p}) = \mathbf{B} + \lambda \mathbf{I} \succeq 0$$

and

$$\nabla \hat{m}(\mathbf{p}) = (\mathbf{B} + \lambda \mathbf{I})\mathbf{p}^* + \mathbf{g} = 0$$

which implies that $\hat{m}(\mathbf{p}^*) \leq \hat{m}(\mathbf{p})$, i.e.,

$$\begin{aligned} m(\mathbf{p}^*) &\leq m(\mathbf{p}) + \frac{\lambda}{2}(\mathbf{p}^\top \mathbf{p} - (\mathbf{p}^*)^\top (\mathbf{p}^*)) \\ &= m(\mathbf{p}) + \frac{\lambda}{2}(\mathbf{p}^\top \mathbf{p} - \Delta^2 + \Delta^2 - (\mathbf{p}^*)^\top (\mathbf{p}^*)) \\ &= m(\mathbf{p}) + \frac{\lambda}{2}(\mathbf{p}^\top \mathbf{p} - \Delta^2) \\ &= m(\mathbf{p}), \quad \forall \text{feasible } \mathbf{p}. \end{aligned}$$

For the reverse direction, assume \mathbf{p}^* is global minimum. If $\|\mathbf{p}^*\| < \Delta$, then $\lambda = 0$. In this case, \mathbf{p}^* is the global minimum for $m(\mathbf{p})$ without constraint, then the quadratic $m(\mathbf{p})$ must be convex, which implies $\mathbf{B} \succeq 0$.

If $\|\mathbf{p}^*\| = \Delta$, it suffices to show

$$\frac{1}{2}(\mathbf{p} - \mathbf{p}^*)^\top (\mathbf{B} + \lambda \mathbf{I})(\mathbf{p} - \mathbf{p}^*) \geq 0,$$

for \mathbf{p}, \mathbf{p}^* on the ball. Check that

$$\begin{aligned} \frac{1}{2}(\mathbf{p} - \mathbf{p}^*)^\top (\mathbf{B} + \lambda \mathbf{I})(\mathbf{p} - \mathbf{p}^*) &= \frac{1}{2}\mathbf{p}^\top (\mathbf{B} + \lambda \mathbf{I})\mathbf{p} + \frac{1}{2}(\mathbf{p}^*)^\top (\mathbf{B} + \lambda \mathbf{I})\mathbf{p} - \mathbf{p}^\top (\mathbf{B} + \lambda \mathbf{I})\mathbf{p}^* \\ &= \frac{1}{2}\mathbf{p}^\top (\mathbf{B} + \lambda \mathbf{I})\mathbf{p} + \frac{1}{2}(\mathbf{p}^*)^\top (\mathbf{B} + \lambda \mathbf{I})\mathbf{p} + \mathbf{p}^\top \mathbf{g} \\ &= \frac{1}{2}\mathbf{p}^\top (\mathbf{B} + \lambda \mathbf{I})\mathbf{p} + \frac{1}{2}(\mathbf{p}^*)^\top (\mathbf{B} + \lambda \mathbf{I})\mathbf{p} + \mathbf{p}^\top \mathbf{g} \\ &\quad - (\mathbf{p}^*)^\top (\mathbf{B} + \lambda \mathbf{I})\mathbf{p} - (\mathbf{p}^*)^\top \mathbf{g} \\ &= \hat{m}(\mathbf{p}) - \hat{m}(\mathbf{p}^*) \\ &= m(\mathbf{p}) - m(\mathbf{p}^*) + \frac{\lambda}{2}(\mathbf{p}^\top \mathbf{p} - (\mathbf{p}^*)^\top \mathbf{p}^*) \\ &= m(\mathbf{p}) - m(\mathbf{p}^*) \geq 0, \end{aligned}$$

which implies that $\mathbf{B} + \lambda \mathbf{I} \succeq 0$. ■

Read chapter 4 for numerical optimization.

You may use the command *fmincon* to implement, but you can barely pass.

12.3.1. Gradient Projection

$$\begin{array}{ll} \min & f(x) \\ \text{such that} & x \in X \end{array}$$

We have

$$\mathbf{x}^* = \text{Proj}_X(\mathbf{x}^* - \alpha \nabla f(\mathbf{x}^*)), \quad \forall \alpha > 0$$

Thus we have the iteration

$$x^{r+1} = \text{Proj}_X(x^r - \alpha_r \nabla f(x^r))$$

At x^r , do the proximal problem:

$$x^{r+1} = \arg \min_{x \in X} f(x) + \frac{1}{2c^k} \|x - x^k\|^2$$

