

**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 |

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 12

Week12

12.1. Monday

12.1.1. Comments on Final Project

$$\begin{aligned} \min \quad & f(x, y) = \frac{1}{2}(\mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{y}^T \mathbf{A} \mathbf{y}) - \mathbf{b}^T \mathbf{x} - \mathbf{c}^T \mathbf{y} := q_b(\mathbf{x}) + q_c(\mathbf{y}) \\ \text{such that} \quad & \frac{1}{2}(\mathbf{x}^T \mathbf{x} - 1) = 0 \\ & \frac{1}{2}(\mathbf{y}^T \mathbf{y} - 1) = 0 \\ & \mathbf{x}^T \mathbf{y} = 0 \end{aligned} \tag{12.1}$$

The Lagrangian function is given by:

$$L(\mathbf{x}, \mathbf{y}, \lambda) = q_b(\mathbf{x}) + q_c(\mathbf{y}) + \frac{\lambda_1}{2}(\mathbf{x}^T \mathbf{x} - 1) + \frac{\lambda_2}{2}(\mathbf{y}^T \mathbf{y} - 1) + \lambda_3 \mathbf{x}^T \mathbf{y} \tag{12.2}$$

Stationarity:

$$\mathbf{A} \mathbf{x} - \mathbf{b} + \lambda_1 \mathbf{x} + \lambda_3 \mathbf{y} = 0$$

$$\mathbf{A} \mathbf{y} - \mathbf{c} + \lambda_2 \mathbf{y} + \lambda_3 \mathbf{x} = 0$$

$$\|h(\mathbf{x})\|^2 = 0$$

The stopping criteria is given by:

$$\max \left\{ \frac{\|\nabla_{\mathbf{x}} L\|}{\|\mathbf{b}\| + 1}, \frac{\|\nabla_{\mathbf{y}} L\|}{\|\mathbf{c}\| + 1}, \|h(\mathbf{x})\| \right\} \leq \text{tol}$$

The problem(12.1) admits its global minimum since the constraint set is compact.

Three methods are suggested: ADMM (you used this); ALMM; quadratic penalty method.

Sometimes we have the second-kind Lagrangian function

$$\hat{L}(\mathbf{x}, \mathbf{y}, \lambda) = q_b(\mathbf{x}) + q_c(\mathbf{y}) + \lambda_3 \mathbf{x}^T \mathbf{y} + \frac{h}{2} (\mathbf{x}^T \mathbf{y})^2$$

We do the minimization

$$\begin{aligned} \min \quad & \hat{L}(\mathbf{x}, \mathbf{y}, \lambda) \\ & \mathbf{x} \in X = \{\mathbf{x} \mid \mathbf{x}^T \mathbf{x} = 1\} \\ & \mathbf{y} \in Y = \{\mathbf{y} \mid \mathbf{y}^T \mathbf{y} = 1\} \end{aligned} \tag{12.3}$$

The update rule is therefore given by ($\tau = 1.618$)

$$\lambda_3 = \lambda_3 + \tau \rho (\mathbf{x}^T \mathbf{y})$$

Develop a solver for the **trust region** sub-problem

$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{p}^T \mathbf{B} \mathbf{p} + \mathbf{g}^T \mathbf{p} \\ & \mathbf{p}^T \mathbf{p} = \Delta^* \end{aligned} \tag{12.4}$$

How to get the **global minimum** for the non-convex problem (12.4)?

12.1.2. Trust Region Method

Our \mathbf{B} has the form

$$\mathbf{A} + \lambda_1 \mathbf{I} + \rho \mathbf{y} \mathbf{y}^T$$

with sparse \mathbf{A} and dense rank 1 matrix \mathbf{y} . Apply the Sherman-Morrison Formula; apply conjugate gradient (command: pcg)

Unconstraint method. Our goal is to minimize $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$. To choose the step-size, we apply the trust region method. First approximate $f(x)$ with quadratic problem:

$$\begin{aligned} f(x^* + p) &\approx f(x^*) + \nabla^T f(x^*)p + \frac{1}{2}p^T \nabla^2 f(x^*)p \\ &\approx f(x^*) + \nabla^T f(x^*)p + \frac{1}{2}p^T \underbrace{\mathbf{B}}_{\text{Approximate Hessian}} p := m(p) \end{aligned}$$

It suffices to choose the step size p to minimize the quadratic function above. The constraint is that the $\|p\|$ should be small enough. Thus it suffices to solve

$$\begin{aligned} \min \quad & \frac{1}{2}\mathbf{p}^T \mathbf{B} \mathbf{p} + \mathbf{g}^T \mathbf{p} \\ & \|\mathbf{p}\| \leq \Delta \end{aligned} \tag{12.5}$$

Given Δ ,

1. Solve the TR subproblem to get p
2. $\rho := \frac{f(x) - f(x+p)}{m(0) - m(p)}$.
 - (a) If $\rho < \frac{1}{4}$, decrease the trust region Δ to $\frac{1}{4}\Delta$.
 - (b) If $\rho > \frac{3}{4}$, increase Δ into 2Δ
 - (c) Else, keep Δ .
3. (a) If $\rho \geq \frac{1}{4}$ and $\rho > \eta$, then $x \leftarrow x + p$
 - (b) Else, x keeps unchanged.

