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

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 12

Week12

12.1. Monday

12.1.1. Comments on Final Project

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

The Lagrangian function is given by:

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

Stationarity:

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} + \lambda_1\boldsymbol{x} + \lambda_3\boldsymbol{y} &= 0 \\
\boldsymbol{A}\boldsymbol{y} - \boldsymbol{c} + \lambda_2\boldsymbol{y} + \lambda_3\boldsymbol{x} &= 0 \\
\|h(\boldsymbol{x})\|^2 &= 0
\end{aligned}$$

The stopping criteria is given by:

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

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

Three methods are suggests: ADMM (yz used this); ALMM; quadratic penalty method.

Sometimes we have the second-kind Lagrangian function

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

We do the minimization

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

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

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

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

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

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

12.1.2. Trust Region Method

Our $\boldsymbol{B}$ has the form

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

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

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

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

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

$$\begin{array}{ll} \min & \frac{1}{2}\boldsymbol{p}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{p}+\boldsymbol{g}^{\mathrm{T}}\boldsymbol{p} \\ & \|\boldsymbol{p}\|\le\Delta \end{array} \tag{12.5}$$

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 $\rho\ge\frac{1}{4}$ and $\rho>\eta$, then $x\leftarrow x+p$
   (b) Else, x keeps unchanged.