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 $\mathcal{M}$

12.3. Wednesday

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

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

In our project we need to slove it very accurately.

Trust Region sub-problem.

$$\begin{aligned} \min \quad & \tfrac{1}{2}\boldsymbol{p}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{p} + \boldsymbol{g}^{\mathrm{T}}\boldsymbol{p} \\ \text{such that} \quad & \|\boldsymbol{p}\| \le \Delta \end{aligned}$$

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

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

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

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

$$\|\boldsymbol{p}\| \le \Delta \tag{12.8d}$$

The second order necessary condition is

$$\boldsymbol{v}^{\mathrm{T}}(\boldsymbol{B} + \lambda \boldsymbol{I})\boldsymbol{v} \ge 0, \forall \boldsymbol{v} \perp \boldsymbol{p}^*$$

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

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

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

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

since

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

and

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

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

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

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

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

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

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

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

which implies that $\boldsymbol{B} + \lambda \boldsymbol{I} \succeq 0$. ∎

Read chapter 4 for numerical optimization.

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

12.3.1. Gradient Projection

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

We have

$$\boldsymbol{x}^* = \mathrm{Proj}_X(\boldsymbol{x}^* - \alpha \nabla f(\boldsymbol{x}^*)), \qquad \forall \alpha > 0$$

Thus we have the iteration

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

At x^r, do the proximal problem:

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