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 $\mathcal{M}$

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

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

The augmented Lagrangian function is given by:

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

There are two penalty algorithms formally listed below:

Quadratic Penalty (Courant, 1943).

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

Augmented Lagrangian Multiplier Method.

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

where $\boldsymbol{c}$ 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 $\boldsymbol{c}$? Let's raise an example first.

■ **Example 10.3** Given the problem

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

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

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

Applying the optimality condition, we derive

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

The second order necessary condition should be

$$\nabla^2_{\boldsymbol{xx}} L_c = \begin{pmatrix} c-1 & 0 \\ 0 & c+1 \end{pmatrix} \succeq 0$$

Therefore, from this example we can see that if $\boldsymbol{c}$ 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

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

with associated local minimum $\boldsymbol{x}^k$. Suppose $\{\boldsymbol{c}^k\}$ is monotone increaseing to infinite, and $\{\boldsymbol{\lambda}^k\}$ is bounded, then every limit point in the sequence $\{\boldsymbol{x}^k\}$ 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.,

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

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

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

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

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

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

$$f(\bar{\boldsymbol{x}}) + (\bar{\boldsymbol{\lambda}})^{\mathrm{T}} h(\bar{\boldsymbol{x}}) + \limsup_{k\to\infty} \frac{\boldsymbol{c}^k}{2} \|h(\boldsymbol{x}^k)\|^2 \leq f^* \tag{10.15}$$

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

$$h(\bar{\boldsymbol{x}}) = 0 \tag{10.16}$$

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