Introduction to Optimization Theory
Homework Assignment 5

Chen Zhiyang, 2017011377

June 2019

Ex. 9.5

Proof
Since $\nabla^2 f(\boldsymbol{x}) \leq M\boldsymbol{I}$, we have

$$f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + (\nabla f(\boldsymbol{x}))^T(\boldsymbol{y}-\boldsymbol{x}) + \frac{M}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2.$$

Plug in $\boldsymbol{y} = \boldsymbol{x} + t\Delta\boldsymbol{x}$, we have

$$f(\boldsymbol{x}+t\Delta\boldsymbol{x}) \leq f(\boldsymbol{x}) + t(\nabla f(\boldsymbol{x}))^T\Delta\boldsymbol{x} + \frac{Mt^2}{2}\|\Delta\boldsymbol{x}\|_2^2.$$

The stopping criterion $f(\boldsymbol{x}+\Delta\boldsymbol{x}) \leq f(\boldsymbol{x}) + \alpha t\nabla f(\boldsymbol{x})\Delta\boldsymbol{x}$ is satisfied if

$$t(\alpha-1)(\nabla f(\boldsymbol{x}))^T\Delta\boldsymbol{x} \geq \frac{Mt^2}{2}\|\Delta\boldsymbol{x}\|_2^2,$$

which means

$$t \leq -2(1-\alpha)\frac{\nabla f(\boldsymbol{x})^T\Delta\boldsymbol{x}}{M\|\Delta\boldsymbol{x}\|_2^2}.$$

Since $0 < \alpha < 0.5$, we have

$$0 < t \leq -\frac{\nabla f(\boldsymbol{x})^T\Delta\boldsymbol{x}}{M\|\Delta\boldsymbol{x}\|_2^2} = t_{\max}.$$

From $\beta^s \leq \min(t_{\max}, 1)$, the number of iterations is upper bounded by $s \leq \log_\beta \min(t_{\max}, 1)$.

Ex. 9.10

(a)

$$f'(x) = \frac{e^{2x} - 1}{e^{2x} + 1}, f''(x) = \frac{4e^{2x}}{(e^{2x} + 1)^2}.$$

Initially, $x^{(0)} = 1$.
First iteration:

$$x^{(1)} = x^{(0)} - \frac{f'(x^{(0)})}{f''(x^{(0)})} = -0.813.$$

Second iteration:

$$x^{(2)} = x^{(1)} - \frac{f'(x^{(1)})}{f''(x^{(1)})} = 0.409.$$

(The algorithm converges.)
Initially, $x^{(0)} = 1.1$.
First iteration:

$$x^{(1)} = x^{(0)} - \frac{f'(x^{(0)})}{f''(x^{(0)})} = -1.129.$$

Second iteration:

$$x^{(2)} = x^{(1)} - \frac{f'(x^{(1)})}{f''(x^{(1)})} = 1.234.$$

(The algorithm fails to converge.)

(b)

$$f'(x) = -\frac{1}{x} + 1, f''(x) = \frac{1}{x^2}.$$

Initially, $x^{(0)} = 3$.
First iteration:

$$x^{(1)} = x^{(0)} - \frac{f'(x^{(0)})}{f''(x^{(0)})} = -3.$$

Note that $x^{(1)} \notin \mathbf{dom}\, f$.

Ex. 10.1

(a)

Proof

Label non-singularity of the KKT matrix and the four statements as (0), (1), (2), (3) and (4), respectively.

(0) $\Rightarrow$ (1):

Suppose not, *i.e.*, there exists an $\boldsymbol{x}$ s.t. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{P}\boldsymbol{x} = \boldsymbol{0}, \boldsymbol{x} \neq \boldsymbol{0}$, and the KKT matrix is non-singular. However, note that

$$\begin{bmatrix} \boldsymbol{P} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{0} \end{bmatrix} = \boldsymbol{0},$$

which means the KKT matrix is singular. This is a contradiction.

(1) $\Rightarrow$ (0):

Suppose the KKT matrix is singular, *i.e.*,

$$\begin{bmatrix} \boldsymbol{P} & \boldsymbol{A}^T \\ \boldsymbol{A} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} = \boldsymbol{0}, \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \neq \boldsymbol{0}.$$

$$\boldsymbol{P}\boldsymbol{x} + \boldsymbol{A}^T\boldsymbol{y} = \boldsymbol{0} \Rightarrow \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} + (\boldsymbol{A}\boldsymbol{x})^T\boldsymbol{y} = \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} = \boldsymbol{0} \Rightarrow \boldsymbol{x} = \boldsymbol{0}, \boldsymbol{y} \neq \boldsymbol{0} \Rightarrow \boldsymbol{A}^T\boldsymbol{y} = \boldsymbol{0},$$

which contradicts rank $\boldsymbol{A} = p$.

(1) $\Rightarrow$ (2):

Suppose $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} = \boldsymbol{0}, \boldsymbol{x} \neq \boldsymbol{0}$. Since $\boldsymbol{P} \geq 0$, we have $\boldsymbol{P}\boldsymbol{x} = \boldsymbol{0}$, which means $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{P}) \cap \mathcal{N}(\boldsymbol{A})$. Contradiction.

(2) $\Rightarrow$ (4):

Let $\boldsymbol{Q} = \boldsymbol{I}$, obviously we have $\boldsymbol{P} + \boldsymbol{A}^T\boldsymbol{A} > 0$.

(4) $\Rightarrow$ (1):

Suppose $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{P}) \cap \mathcal{N}(\boldsymbol{A}), \boldsymbol{x} \neq \boldsymbol{0}$, we have

$$\boldsymbol{P}\boldsymbol{x} + \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \Rightarrow \boldsymbol{x}^T(\boldsymbol{P} + \boldsymbol{A}^T\boldsymbol{Q}\boldsymbol{A})\boldsymbol{x} = \boldsymbol{0},$$

which is a contradiction.

(2) $\Leftrightarrow$ (3):

If $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}, \boldsymbol{x} \neq \boldsymbol{0}, \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} > 0$, since $\boldsymbol{x} \in \mathcal{R}(\boldsymbol{F})$, $\boldsymbol{x} = \boldsymbol{F}\boldsymbol{y}$ for some $\boldsymbol{y} \neq \boldsymbol{0}$. We have

$$\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} = \boldsymbol{y}^T(\boldsymbol{F}^T\boldsymbol{P}\boldsymbol{F})\boldsymbol{y} > 0 \Leftrightarrow \boldsymbol{F}^T\boldsymbol{P}\boldsymbol{F} > 0.$$

The above reductions form a strongly-connected directed graph, *i.e.*, the statements are equivalent.

Ex. 10.9

(a)

It is obvious that f is a convex function. Suppose the dual optimal cost is p^*. If p^* is not a feasible cost, either

$$\lim_{\boldsymbol{x}\to\infty} f(\boldsymbol{x}) = p^*,$$

or

$$\lim_{\boldsymbol{x}\to\boldsymbol{x}^*} f(\boldsymbol{x}) = p^*,$$

where $\boldsymbol{x}^*$ has zero components.
Note that $\lim_{x\to\infty} x \log x = \infty$, the first situation is impossible.
As for the second situation, let $\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^*$, $g(t) = \sum(x_i^* + t\Delta x_i)\log(x_i^* + t\Delta x_i), t > 0$. We have

$$g'(t) = \sum \Delta x_i (1 + \log(x_i^* + t\Delta x_i)).$$

If $x_i^* = 0$, $\Delta x_i > 0$, which means $\lim_{t\to 0} g(t) = -\infty$. This is also impossible.

Ex. 11.4

Let ϕ be the barrier function of **Ex. 11.1**. We have

$$\nabla^2(tf_0 + \phi) = \nabla^2(tf_0 + \tilde{\phi}) + \frac{2}{R^2 - \|\boldsymbol{x}\|_2^2}\boldsymbol{I} + \frac{4}{(R^2 - \|\boldsymbol{x}\|_2^2)^2}\boldsymbol{x}\boldsymbol{x}^T \geq \frac{2}{R^2}\boldsymbol{I}.$$

We can take $m = \frac{2}{R^2}$.