CMSE890 Homework#1

Haiyang Yu

March 25, 2020

1.

If $\boldsymbol{x}$ is optimal, for all feasible $\boldsymbol{y}$, $\boldsymbol{y}_k = \frac{k}{k+1}\boldsymbol{x} + \frac{1}{k+1}\boldsymbol{y}$, we have $f(\boldsymbol{y}_k) - f(\boldsymbol{x}) \geq 0, \forall k$. So

$$\frac{f(\boldsymbol{y}_k) - f(\boldsymbol{x})}{\frac{1}{k+1}} = \frac{f(\boldsymbol{x} + \frac{1}{k+1}(\boldsymbol{y} - \boldsymbol{x})) - f(\boldsymbol{x})}{\frac{1}{k+1}} \geq 0$$

Let $k \to \infty$, we get $\nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}) \geq 0$

2.

Because $\boldsymbol{x}^{\mathrm{T}}W\boldsymbol{x}$ is a scalar. $(\boldsymbol{x}^{\mathrm{T}}W\boldsymbol{x})^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}}W\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\frac{W+W^{\mathrm{T}}}{2}\boldsymbol{x}$. So we always assume W is symmetric. Symmetric positive definite matrix has the real and positive eigenvalue and eigenvector is orthogonal, which means $P^{-1} = P^{\mathrm{T}}$.

Suppose

$$\begin{aligned}
W_1 &= P^{-1}\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)P \\
&= (P^{-1}\mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_n})P)^{\mathrm{T}}(P^{-1}\mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_n})P) \\
&\triangleq S^{\mathrm{T}}S
\end{aligned}$$

In the same way, we define $W_2 = R^{\mathrm{T}}R$.

$$\begin{aligned}
||\boldsymbol{x}||_{W_1} &= \sqrt{\boldsymbol{x}^{\mathrm{T}}W_1x} = \sqrt{\boldsymbol{x}^{\mathrm{T}}S^{\mathrm{T}}S\boldsymbol{x}} \\
&= ||S\boldsymbol{x}||_2 = ||SR^{-1}R\boldsymbol{x}||_2 \\
&\leq ||SR^{-1}||_2 \cdot ||R\boldsymbol{x}||_2 = ||SR^{-1}||_2 \cdot ||x||_{W_2}
\end{aligned}$$

In the same way, we get $||\boldsymbol{x}||_{W_2} \leq ||RS^{-1}||_2 \cdot ||x||_{W_1}$, so $m = ||SR^{-1}||_2^{-1}$ and $M = ||RS^{-1}||_2$.

3.

Suppose $f(\boldsymbol{x}) = ||\boldsymbol{x}||_1 + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{b},\ \boldsymbol{b} = (b_1, b_2, \cdots, b_n)^{\mathrm{T}}$.

Case 1: $|b_i| \leq 1, \forall i \in \{1, 2, \cdots, n\}$.

Then

$$\begin{aligned}
f(\boldsymbol{x}) &= \sum_{i=1}^{n}(|x_i| + b_i x_i) \\
&\geq \sum_{i=1}^{n}(|x_i| - |b_i||x_i|) \\
&= \sum_{i=1}^{n}(1 - |b_i|)|x_i| \\
&\geq 0
\end{aligned}$$

However, $f(\boldsymbol{x}^{\boldsymbol{*}}) = 0$ when $\boldsymbol{x}^{\boldsymbol{*}} = \boldsymbol{0}$, So

$$\min_{||\boldsymbol{x}||_2 \le 1} f(\boldsymbol{x}) = 0$$

$$\arg\min_{||\boldsymbol{x}||_2 \le 1} f(\boldsymbol{x}) = \boldsymbol{0}$$

Case 2: $\exists i \in \{1, 2, \cdots, n\}$ that makes $|b_i| > 1$.

In this case, we suppose that $|b_1| \ge |b_2| \ge \cdots \ge |b_k| > 1 \ge |b_{k+1}| \ge \cdots \ge |b_n|$.(In other words, we sort $|b_i|$)

Then, for $||\boldsymbol{x}||_2 \le 1$,

$$\begin{aligned}
f(\boldsymbol{x}) &= \sum_{i=1}^{n} (|x_i| + b_i x_i) \\
&\ge \sum_{i=1}^{n} (|x_i| - |b_i||x_i|) \\
&= \sum_{i=1}^{k} (1 - |b_i|)|x_i| + \sum_{i=k+1}^{n} (1 - |b_i|)|x_i| \\
&\ge \sum_{i=1}^{k} (1 - |b_i|)|x_i| \\
&= (\boldsymbol{b}^*)^{\mathrm{T}} \mathrm{sgn}(\boldsymbol{x})
\end{aligned}$$

where $\boldsymbol{b}^* = (1 - |b_1|, 1 - |b_2|, \cdots, 1 - |b_k|, 0, \cdots, 0)^{\mathrm{T}}$ and $\mathrm{sgn}(\boldsymbol{x}) = (|x_1|, |x_2|, \cdots, |x_n|)^{\mathrm{T}}$. We have $\{\mathrm{sgn}(\boldsymbol{x}) : ||\boldsymbol{x}||_2 \le 1\} \subset \{\boldsymbol{x} : ||\boldsymbol{x}||_2 \le 1\}$.

So

$$\begin{aligned}
f(\boldsymbol{x}) &\ge (\boldsymbol{b}^*)^{\mathrm{T}} \mathrm{sgn}(\boldsymbol{x}) \\
&\ge \inf_{||\boldsymbol{x}||_2 \le 1} (\boldsymbol{b}^*)^{\mathrm{T}} \mathrm{sgn}(\boldsymbol{x}) \\
&\ge \inf_{||\boldsymbol{x}||_2 \le 1} (\boldsymbol{b}^*)^{\mathrm{T}} \boldsymbol{x} \\
&= (\boldsymbol{b}^*)^{\mathrm{T}} \left(-\frac{\boldsymbol{b}^*}{||\boldsymbol{b}^*||_2} \right) \\
&= -||\boldsymbol{b}^*||_2 \\
&= -\sqrt{\sum_{i=1}^{k} (1 - |b_i|)^2}
\end{aligned}$$

However, let $\boldsymbol{x}^* = \left(\frac{(1-|b_1|)\mathrm{sgn}(b_1)}{\sqrt{\sum_{i=1}^{k}(1-|b_i|)^2}}, \frac{(1-|b_2|)\mathrm{sgn}(b_2)}{\sqrt{\sum_{i=1}^{k}(1-|b_i|)^2}}, \cdots, \frac{(1-|b_k|)\mathrm{sgn}(b_k)}{\sqrt{\sum_{i=1}^{k}(1-|b_i|)^2}}, 0, \cdots, 0 \right)^{\mathrm{T}}$

Then

$$\begin{aligned}
f(\boldsymbol{x}^*) &= \sum_{i=1}^{k} \left(\frac{|\mathrm{sgn}(b_i)(1 - |b_i|)|}{\sqrt{\sum_{j=1}^{k}(1 - |b_j|)^2}} + \frac{b_i \mathrm{sgn}(b_i)(1 - |b_i|)}{\sqrt{\sum_{j=1}^{k}(1 - |b_j|)^2}} \right) \\
&= \sum_{i=1}^{k} \left(\frac{(|b_i| - 1)}{\sqrt{\sum_{j=1}^{k}(1 - |b_j|)^2}} + \frac{|b_i|(1 - |b_i|)}{\sqrt{\sum_{j=1}^{k}(1 - |b_j|)^2}} \right) \\
&= \sum_{i=1}^{k} \frac{-(1 - |b_i|)^2}{\sqrt{\sum_{j=1}^{k}(1 - |b_j|)^2}} \\
&= -\sqrt{\sum_{i=1}^{k} (1 - |b_i|)^2}
\end{aligned}$$

So

$$\min_{||\boldsymbol{x}||_2\leq 1} f(\boldsymbol{x}) = -\sqrt{\sum_{i=1}^{k}(1-|b_i|)^2}$$

$$\arg\min_{||\boldsymbol{x}||_2\leq 1} f(\boldsymbol{x}) = \left(\frac{(1-|b_1|)\mathrm{sgn}(b_1)}{\sqrt{\sum_{i=1}^{k}(1-|b_i|)^2}}, \frac{(1-|b_2|)\mathrm{sgn}(b_2)}{\sqrt{\sum_{i=1}^{k}(1-|b_i|)^2}}, \cdots, \frac{(1-|b_k|)\mathrm{sgn}(b_k)}{\sqrt{\sum_{i=1}^{k}(1-|b_i|)^2}}, 0, \cdots, 0\right)^{\mathrm{T}}$$

4.

Let $p = 1/\theta$, $q = 1/(1-\theta)$. Then $1/p + 1/q = 1$

$$\begin{aligned}
f(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}) &= \sum_{i=1}^{n}\log\left(1 + \mathrm{e}^{\boldsymbol{a}_i^{\mathrm{T}}(\theta\boldsymbol{x}+(1-\theta)\boldsymbol{y})}\right) \\
&= \sum_{i=1}^{n}\log\left(1^{\theta}\cdot 1^{1-\theta} + \mathrm{e}^{\theta\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}}\cdot\mathrm{e}^{(1-\theta)\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}}\right) \\
&\leq \sum_{i=1}^{n}\log\left(\left((1^{\theta})^{p} + (\mathrm{e}^{\theta\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}})^{p}\right)^{1/p}\cdot\left((1^{1-\theta})^{q} + (\mathrm{e}^{(1-\theta)\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}})^{q}\right)^{1/q}\right) \\
&= \frac{1}{p}\sum_{i=1}^{n}\log(1+\mathrm{e}^{\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}}) + \frac{1}{q}\sum_{i=1}^{n}\log(1+\mathrm{e}^{\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{y}}) \\
&= \theta f(\boldsymbol{x}) + (1-\theta)f(\boldsymbol{y})
\end{aligned}$$

so f is convex.

5.

(1)

$$\begin{aligned}
(F(\boldsymbol{x}) - F(\boldsymbol{y}))^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y}) \geq 0 &\Leftrightarrow (\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}A^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y}) \geq 0 \\
&\Leftrightarrow A \text{ is positive semi-definite.}
\end{aligned}$$

(2)

$$\begin{aligned}
(F(\boldsymbol{x}) - F(\boldsymbol{y}))^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y}) > 0 &\Leftrightarrow (\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}A^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y}) > 0 \\
&\Leftrightarrow A \text{ is positive definite.}
\end{aligned}$$

(3)

$$\begin{aligned}
(F(\boldsymbol{x}) - F(\boldsymbol{y}))^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y}) \geq m||\boldsymbol{x}-\boldsymbol{y}||^2 &\Leftrightarrow (\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}(A-mI)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y}) \geq 0 \\
&\Leftrightarrow A - mI \text{ is positive semi-definite.}
\end{aligned}$$

(4)

$$\begin{aligned}
||F(\boldsymbol{x}) - F(\boldsymbol{y})|| \leq L||\boldsymbol{x}-\boldsymbol{y}|| &\Leftrightarrow (\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}A^{\mathrm{T}}A(\boldsymbol{x}-\boldsymbol{y}) \leq L^2(\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y}) \\
&\Leftrightarrow L^2 - A^{\mathrm{T}}A \text{ is positive semi-definite.}
\end{aligned}$$

(5)

$$\begin{aligned}
(F(\boldsymbol{x}) - F(\boldsymbol{y}))^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y}) \geq \frac{1}{L}||F(\boldsymbol{x}) - F(\boldsymbol{y})||^2 &\Leftrightarrow (\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}A^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y}) \geq \frac{1}{L}(\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}A^{\mathrm{T}}A(\boldsymbol{x}-\boldsymbol{y}) \\
&\Leftrightarrow A^{\mathrm{T}} - \frac{1}{L}A^{\mathrm{T}}A \text{ is positive semi-definite.}
\end{aligned}$$