Linear and Convex Optimization Homework 01

Qiu Yihang, 2021/09/25-09/30

1. ***Solution:***

(a) $f(\boldsymbol{x})$ is coercive. The proof is as follows.

Let $M = \max\{|x_1, x_2|\}$.

$$f(\boldsymbol{x}) = 2x_1^2 + x_1x_2 + x_2^2 - 3x_1 - 5x_2 \geq \frac{3}{2}x_1^2 + \frac{1}{2}x_2^2 - 3x_1 - 5x_2$$

$$\geq \frac{1}{2}(x_1^2 + x_2^2) - 3x_1 - 5x_2 \geq \frac{1}{2}\|\boldsymbol{x}\| - 8\sqrt{\|\boldsymbol{x}\|}$$

$$(\because (3x_1 + 5x_2)^2 = 9x_1^2 + 30x_1x_2 + 25x_2^2 \leq 64M^2 \leq 64\|\boldsymbol{x}\|)$$

Meanwhile, when $\|\boldsymbol{x}\| \to \infty$, $\|\boldsymbol{x}\| \gg 16\sqrt{\|\boldsymbol{x}\|}$.

Thus, when $\|\boldsymbol{x}\| \to \infty$, $f(\boldsymbol{x}) \to \infty$, **i.e.** $f(\boldsymbol{x})$ is coercive. ■

(b) $\nabla f(\boldsymbol{x}) = (4x_1 + x_2 - 3 \quad 2x_2 + x_1 - 5)$, $\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$

By the first-order necessary condition of local minimum, we can find all stationary points.

When $\nabla f(\boldsymbol{x}) = \boldsymbol{o}$, $\begin{cases} 4x_1 + x_2 - 3 = 0 \\ x_1 + 2x_2 - 5 = 0 \end{cases} \Rightarrow \begin{cases} x_1 = \frac{1}{7} \\ x_2 = \frac{17}{7} \end{cases}$.

Let $\boldsymbol{x}^* = \left(\frac{1}{7}, \frac{17}{7}\right)$. $f(\boldsymbol{x}^*) = -\frac{44}{7}$.

Now we prove $\boldsymbol{x}^*$ is a local minimum.

$4 > 0, \begin{vmatrix} 4 & 1 \\ 1 & 2 \end{vmatrix} > 0 \Rightarrow \nabla^2 f(\boldsymbol{x}) \succ \boldsymbol{O}$, **i.e.** $\nabla^2 f(\boldsymbol{x})$ is positive definite.

Since $\nabla f(\boldsymbol{x}^*) = \boldsymbol{o}$ and $\nabla^2 f(\boldsymbol{x}) \succ \boldsymbol{O}$, $\boldsymbol{x}^*$ is a local minimum.

Moreover, $\boldsymbol{x}^*$ is the only local minimum, **i.e.** the global minimum.

Thus, the minimum of $f(\boldsymbol{x})$ over $\mathbb{R}^2$ is $-\frac{44}{7}$. ■

2. (a) ***Solution:***

f does not have a global minimum. The proof is as follows.

For any $\boldsymbol{w}$, there exist two cases:

1) $\exists k \in \{1,2,\dots,m\}$ s.t. $y_k\boldsymbol{x}_k^T\boldsymbol{w} \leq 0$.

$$\sum_{i=1}^{m} \log\left(1 + e^{-y_i\boldsymbol{x}_i^T\boldsymbol{w}}\right) > \log\left(1 + e^{-y_k\boldsymbol{x}_k^T\boldsymbol{w}}\right) > 0$$

Since $y_i\boldsymbol{x}_i^T\boldsymbol{w}_0 > 0, \forall i \in \{1,2,\dots,m\}$ and $\log(1 + e^{-z}) \to 0$ as $z \to +\infty$, we can always find an $M > 0$ large enough **s.t.**

$$\sum_{i=1}^{m} \log\left(1 + e^{-M(y_i\boldsymbol{x}_i^T\boldsymbol{w}_0)}\right) < \log\left(1 + e^{-y_k\boldsymbol{x}_k^T\boldsymbol{w}}\right)$$

Let $\boldsymbol{w}^* = M\boldsymbol{w_0} = M\left(w_{0_{i,j}}\right)_{n\times 1}$.

Therefore,

$$\sum_{i=1}^{m} \log\left(1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}\right) > \sum_{i=1}^{m} \log\left(1 + e^{-M(y_i \boldsymbol{x}_i^T \boldsymbol{w}_0)}\right) = \sum_{i=1}^{m} \log\left(1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}^*}\right)$$

i.e. $f(\boldsymbol{w}) > f(\boldsymbol{w}^*)$.

2) $\forall i \in \{1,2,\dots,m\},\ y_i \boldsymbol{x}_i^T \boldsymbol{w} > 0$.

We can always find an $M > 0$ large enough **s.t.**

$$\boldsymbol{x}_i^T \boldsymbol{w} < M \boldsymbol{x}_i^T \boldsymbol{w_0}, \forall i \in \{1,2,\dots,m\}$$

Let $\boldsymbol{w}^* = M\boldsymbol{w_0} = M\left(w_{0_{i,j}}\right)_{n\times 1}$.

Since $\log(1 + e^{-z}) \to 0$ as $z \to +\infty$,

$$\log\left(1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}\right) > \log\left(1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}^*}\right)$$

$$\sum_{i=1}^{m} \log\left(1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}\right) > \sum_{i=1}^{m} \log\left(1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}^*}\right)$$

i.e. $f(\boldsymbol{w}) > f(\boldsymbol{w}^*)$.

Thus, for any $\boldsymbol{w}$ exists a $\boldsymbol{w}^*$ s.t. $f(\boldsymbol{w}) > f(\boldsymbol{w}^*)$.

In other words, f does not have a global minimum. ■

(b) i) ***Proof:***

To prove $f(\boldsymbol{w}) \geq h(\boldsymbol{w})$, just need to prove

$$\sum_{i=1}^{m} \log(1 + e^{-z_i}) \geq \max_{1\leq i\leq m} -z_i \ \ (\exists i \ \mathbf{s.t.}\, z_i < 0)$$

Now we prove the inequality above holds.

$$\sum_{i=1}^{m} \log(1 + e^{-z_i}) = \log\left(\prod_{i=1}^{m} 1 + e^{-z_i}\right) \geq \log\left(\left(\max_{1\leq i\leq m} 1 + e^{-z_i}\right)\cdot 1 \cdot \dots \cdot 1\right)$$

$$= \log\left(\max_{1\leq i\leq m} 1 + e^{-z_i}\right) \geq \log\left(\max_{1\leq i\leq m} e^{-z_i}\right) = \max_{1\leq i\leq m} \log(e^{-z_i}) = \max_{1\leq i\leq m} -z_i$$

Qed. ■

ii) ***Proof:***

Considering $\|\boldsymbol{w}\| = 1 \leq 1$, S is bounded. On the other hand, S is closed (since for any $x \in S^C, \|\boldsymbol{x}\| \neq 1,\ \exists \varepsilon > 0\ \mathbf{s.t.} \forall \boldsymbol{y} \in B(\boldsymbol{x}, \varepsilon), \|\boldsymbol{y}\| \neq 1, \mathbf{i.e.} B(\boldsymbol{x}, \varepsilon) \subset S^C$).

Therefore, S is a compact set.

From the assumption given, since $y_i \boldsymbol{x}_i^T \boldsymbol{w} = y_i \sum_{k=1}^{n} (x_i)_k w_k$ is continuous, we know $h(\boldsymbol{w}) = \max_{1\leq i\leq m} -y_i \boldsymbol{x}_i^T \boldsymbol{w}$ is continuous.

Thus, by **Extreme Value Theorem**, $h(\boldsymbol{w})$ has a global minimum $\boldsymbol{w}_0$ on S.

Meanwhile, $\forall \boldsymbol{w},\ \exists i_0 = 1,2,\dots,m$ **s.t.** $y_{i_0}\boldsymbol{x}_{i_0}^T\boldsymbol{w} < \boldsymbol{0}$.

Therefore, $\forall \boldsymbol{w},\ h(\boldsymbol{w}) = \max\limits_{1\le i\le m} -y_i\boldsymbol{x}_i^T\boldsymbol{w} > -y_{i_0}\boldsymbol{x}_{i_0}^T\boldsymbol{w} > \boldsymbol{0}$.

Thus, $C \triangleq h(\boldsymbol{w_0}) > 0$. *Qed.* ■

<u>* In fact, we can prove the assumption given in the problem.</u>

Lemma. When $f_i(x)$ is continuous, $i = 1,2,\dots,N$, $g_N(x) = \max\limits_{1\le i\le N} f_i(x)$ is continuous.

Proof. **BASE STEP.** $N = 2$. $g_N(x) = \frac{f_1(x)+f_2(x)+|f_1(x)-f_2(x)|}{2}$.

Since $f_1(x), f_2(x)$ are both continuous, $|f_1(x) - f_2(x)|$ is also continuous. Therefore, $g_N(x)$ is continuous.

INDUCTIVE STEP.

Suppose when $N = k$, $g_N(x)$ is continuous. Now we prove $g_{k+1}(x)$ is also continuous.

$$g_{k+1}(x) = \max\{\, g_k(x), f_{k+1}(x)\,\} = \frac{g_k(x)+f_{k+1}(x)+|g_k(x)-f_{k+1}(x)|}{2}.$$

Since $g_k(x), f_{k+1}(x)$ are continuous, similarly to the proof of base step, we can prove that $g_{k+1}(x)$ is also continuous, **i.e.** when $N = k+1$, $g_N(x)$ is also continuous.

Thus, $g_N(x)$ is continuous. □

iii) ***Proof:***

Let $\boldsymbol{w}^{(1)} = \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$. Then $\|\boldsymbol{w}^{(1)}\| = \boldsymbol{1}$. From ii) we know $h(\boldsymbol{w}^{(1)}) \ge C$.

$$\forall \boldsymbol{w}, h(\boldsymbol{w}) = \max_{1\le i\le m} -y_i\boldsymbol{x}_i^T\boldsymbol{w} = \max_{1\le i\le m} y_i \sum_{k=1}^{n}(x_i)_k w_k = \|w\| \max_{1\le i\le m} y_i \sum_{k=1}^{n}\frac{(x_i)_k w_k}{\|\boldsymbol{w}\|}$$

$$= \|\boldsymbol{w}\| \max_{1\le i\le m} -\frac{y_i\boldsymbol{x}_i^T\boldsymbol{w}}{\|\boldsymbol{w}\|} = \|\boldsymbol{w}\| \max_{1\le i\le m} -y_i\boldsymbol{x}_i^T\boldsymbol{w}^{(1)} \ge \|\boldsymbol{w}\| h(\boldsymbol{w}^{(1)}) \ge C\|\boldsymbol{w}\|$$

Qed. ■

iv) ***Proof:***

From i), ii) and iii) we know that $f(\boldsymbol{w}) \ge h(\boldsymbol{w}) \ge C\|\boldsymbol{w}\|$ (where $C > 0$).

Thus, $f(\boldsymbol{w}) \to \infty$ as $\|\boldsymbol{w}\| \to \infty$, **i.e.** $f(\boldsymbol{w})$ is coercive.

Meanwhile, we know $f(\boldsymbol{w})$ is continuous since $\sum_{i=1}^{m}\log(1+e^{-z_i})$ is continuous and $-y_i\boldsymbol{x}_i^T\boldsymbol{w}$ is continuous.

Therefore, the global minimum of $f(\boldsymbol{w})$ exists. *Qed.* ■

(c) ***Solution:***

$$f(\boldsymbol{w}) = \sum_{i=1}^{m} \log\left(1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}\right) = \sum_{i=1}^{m} \log\left(1 + e^{-y_i \sum_{j=1}^{n} (x_i)_j w_j}\right)$$

$$\frac{\partial f}{\partial w_k} = \sum_{i=1}^{m} \frac{\partial}{\partial w_k} \log\left(1 + e^{-y_i \sum_{j=1}^{n} (x_i)_j w_j}\right) = \sum_{i=1}^{m} \frac{-y_i (x_i)_k e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}}{1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}}$$

Thus,

$$f'(\boldsymbol{w}) = \left(\frac{\partial f}{\partial w_1}, \frac{\partial f}{\partial w_2}, \dots, \frac{\partial f}{\partial w_n}\right)^T = \sum_{i=1}^{m} \frac{-y_i e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}}{1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}} ((x_i)_1, (x_i)_2, \dots, (x_i)_n)^T = \sum_{i=1}^{m} \frac{-y_i e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}}{1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}} \boldsymbol{x}_i$$

i.e.,

$$\nabla f(\boldsymbol{w}) = f'(\boldsymbol{w})^T = \left(\sum_{i=1}^{m} \frac{-y_i e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}}{1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}} \boldsymbol{x}_i\right)^T = \sum_{i=1}^{m} \frac{-y_i e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}}{1 + e^{-y_i \boldsymbol{x}_i^T \boldsymbol{w}}} \boldsymbol{x}_i^T$$

■

3. (a) ***Proof:***

By the assumption given in the problem, we know

$$h(x + \Delta x) = h(x) + h'(x)\Delta x + \frac{1}{2} h''(x + t\Delta x)(\Delta x)^2$$

for some $t \in (0,1)$.

Let $\boldsymbol{d_0} = \frac{\boldsymbol{d}}{\|\boldsymbol{d}\|}, l = \|\boldsymbol{d}\|$. Let $g(l) = f(\boldsymbol{x} + l\boldsymbol{d_0})$, which is obviously a univariant function. Also, $g(0) = f(\boldsymbol{x})$.

By **Chain Rule**, $g'(l) = f'(\boldsymbol{x} + l\boldsymbol{d_0}) \frac{\partial(\boldsymbol{x} + l\boldsymbol{d_0})}{\partial l} = \nabla f(\boldsymbol{x})^T \boldsymbol{d_0}$,

$$g''(l) = (\nabla f(\boldsymbol{x})^T \boldsymbol{d_0})' = \boldsymbol{d_0^T} \nabla^2 f(\boldsymbol{x}) \boldsymbol{d_0}.$$

Thus, $g(l) = g(0) + g'(0)l + \frac{1}{2} g''(tl) l^2$ for some $t \in (0,1)$,

i.e. $f(\boldsymbol{x} + l\boldsymbol{d_0}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{d_0} l) + \frac{1}{2} \boldsymbol{d_0^T} \nabla^2 f(\boldsymbol{x} + tl\boldsymbol{d_0}) \boldsymbol{d_0} l^2$

$= f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{d} + \frac{1}{2} \boldsymbol{d}^T \nabla^2 f(\boldsymbol{x} + t\boldsymbol{d}) \boldsymbol{d}$ for some $t \in (0,1)$.

Qed. ■

(b) ***Proof:***

Let $\boldsymbol{h}(t) = \nabla f(\boldsymbol{x} + t\boldsymbol{d})$. By **Chain Rule**, we have $\boldsymbol{h}'(t) = \nabla^2 f(\boldsymbol{x} + t\boldsymbol{d})\boldsymbol{d}$.

By **Newton-Leibniz Formula**,

$$\int_0^1 \nabla^2 f(\boldsymbol{x} + t\boldsymbol{d})\boldsymbol{d} \, \mathrm{d}t = \boldsymbol{h}(1) - \boldsymbol{h}(0) = \nabla f(\boldsymbol{x} + \boldsymbol{d}) - \nabla f(\boldsymbol{x})$$

i.e. $\nabla f(\boldsymbol{x} + \boldsymbol{d}) = \nabla f(\boldsymbol{x}) + \int_0^1 \nabla^2 f(\boldsymbol{x} + t\boldsymbol{d})\boldsymbol{d} \, \mathrm{d}t$. *Qed.* ■

4. ***Solution:***

(1) $A = \begin{pmatrix} 6 & 2 & 0 \\ 2 & 5 & -2 \\ 0 & -2 & 4 \end{pmatrix}$ is **positive definite**. ■

(Since $6 > 0$, $\begin{vmatrix} 6 & 2 \\ 2 & 5 \end{vmatrix} = 26 > 0$, $\begin{vmatrix} 6 & 2 & 0 \\ 2 & 5 & -2 \\ 0 & -2 & 4 \end{vmatrix} = 80 > 0$)

(2) $B = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 2 & 1 \\ 0 & 1 & -3 \end{pmatrix}$ is **indefinite**. ■

(Since $1 > 0, \begin{vmatrix} 1 & 2 \\ 2 & 2 \end{vmatrix} = -2 < 0$)

(3) $C = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$ is **positive semidefinite**. ■

(Since $\mathbf{Def}\left(C_{\{i\}}\right) = 2 \geq 0, \forall i \in \{1,2,3\}$,

$\mathbf{Def}\left(C_{\{i,j\}}\right) = \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 5 \geq 0, \forall i < j \ \ \mathbf{s.t.} \{i, j\} \in \{1,2,3\}$,

$\mathbf{Def}(C) = \begin{vmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{vmatrix} = 0 \geq 0$)