MSBD 5007 HW2

RONG Shuo

March 30, 2025

Question1

Determine the convexity of the following functions, where $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{X} \in \mathbb{S}^n_{++}$(the set of symmetric positive definite matrices). Justify your answer.

(a) $f(\boldsymbol{x}) = \log(e^{x_1} + e^{x_2} + \cdots + e^{x_n})$.

(b) $f(\boldsymbol{X}) = \text{logdet}(\boldsymbol{X})$.

Answer

(a)

$$\nabla f(\boldsymbol{x}) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n})$$

$$\begin{aligned}
\frac{\partial f}{\partial x_i} &= \frac{1}{e^{x_1} + e^{x_1} + \cdots + e^{x_1}} \times \frac{\partial}{\partial x_i}(e^{x_1} + e^{x_2} + \cdots + e^{x_n}) \\
&= \frac{e^{x_i}}{e^{x_1} + e^{x_2} + \cdots + e^{x_n}}
\end{aligned}$$

Therefore,

$$\nabla f(\boldsymbol{x}) = (\frac{e^{x_1}}{\sum_{i=1}^n e^{x_i}}, \frac{e^{x_2}}{\sum_{i=1}^n e^{x_i}}, \cdots, \frac{e^{x_n}}{\sum_{i=n}^n e^{x_i}})$$

We know,

$$\begin{aligned}
\boldsymbol{H}_{ij} &= \frac{\partial^2 f}{\partial x_i \partial y_i} \\
\boldsymbol{H}_{ii} &= \frac{\partial}{\partial x_i}(\frac{e^{x_i}}{S}) = \frac{e^{x_i}(S - e^{x_i})}{S^2} \text{ if } i = j \\
\boldsymbol{H}_{ij} &= -\frac{e^{x_i} e^{x_j}}{S^2} \text{ if } i \neq j \\
\nabla^2 f(\boldsymbol{x}) &= \frac{1}{S}\text{diag}(\boldsymbol{e}) - \frac{1}{S^2}\boldsymbol{e}\boldsymbol{e}^T
\end{aligned}$$

where $\boldsymbol{e} = (e^{x_1}, e^{x_2}, \cdots, e^{x_n})^T$, and $S = \sum_{i=1}^n e^{x_i}$

$$\begin{aligned}
&\boldsymbol{z}\nabla^2 f(\boldsymbol{x})\boldsymbol{z}^T = \frac{1}{S}\sum_{i=1}^n e^{x_i} z_i^2 - \frac{1}{S^2}(\sum_{i=1}^n e^{x_i} z_i)^2 \\
&(\sum_{i=1}^n e^{x_i} z_i)^2 \leq (\sum_{i=1}^n e^{x_i} z_i^2)(\sum_{i=1}^n e^{x_i}) \\
&\frac{1}{S}\sum_{i=1}^n e^{x_i} z_i^2 \geq \frac{1}{S^2}(\sum_{i=1}^n e^{x_i} z_i)^2 \\
&\boldsymbol{z}\nabla^2 f(\boldsymbol{x})\boldsymbol{z}^T \geq 0
\end{aligned}$$

Thus, we prove that Hessian is positive semi-definite, so we can conclude $f(\boldsymbol{x})$ is convex.

(b)

Let $\boldsymbol{X} \in \mathbb{S}^n_{++}$ and $\boldsymbol{V} \in \mathbb{S}^n$, we define $g(t) = \text{logdet}(\boldsymbol{X} + t\boldsymbol{V})$, where $\boldsymbol{X} + t\boldsymbol{V}$ is symmetric, positive and definite.

$$\begin{aligned}
g(t) &= \text{logdet}(\boldsymbol{X} + t\boldsymbol{V}) \\
&= \text{logdet}(\boldsymbol{X}^{\frac{1}{2}}(\boldsymbol{I} + t\boldsymbol{X}^{-\frac{1}{2}}\boldsymbol{V}\boldsymbol{X}^{-\frac{1}{2}})\boldsymbol{X}^{\frac{1}{2}}) \\
&= \text{logdet}\boldsymbol{X} + \text{logdet}(\boldsymbol{I} + t\boldsymbol{\Lambda}) \\
&= \text{logdet}\boldsymbol{X} + \sum_{i=1}^{n} \log(1 + t\lambda_i)
\end{aligned}$$

where λ_i is the eigenvalues of Λ

$$g''(t) = \sum_{i=1}^{n} \frac{-\lambda_i^2}{(1 + t\lambda_i)^2} \leq 0$$

Thus $g''(t) \leq 0$ for all t where $\boldsymbol{X} + t\boldsymbol{V}$ is symmetric, positive and definite, so $g(t)$ is concave. Since $g(t)$ is concave for any direction, $f(X) = \text{logdet}\boldsymbol{X}$ is concave on $\mathbb{S}^n_{++}$.

Question2

Consider the linear system $\boldsymbol{Ax} = \boldsymbol{b}$, with $\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$ and $\boldsymbol{b} \in \begin{bmatrix} 2 \\ -4 \end{bmatrix}$, and the initial guess $\boldsymbol{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

(a) Present the first two update iterations using the **steepest descent algorithm**.
(b) Present the first two updated iterations using the **conjugate gradient algorithm**.

Answer

(a)

We know if we want use steepest descent algorithm to solve the linear system, we need:

$$\alpha_k = \frac{\boldsymbol{r}_k^T \boldsymbol{r}_k}{\boldsymbol{r}_k^T \boldsymbol{A} \boldsymbol{r}_k}$$

where $\boldsymbol{r}_{k+1} = \boldsymbol{r}_k - \alpha_k \boldsymbol{A} \boldsymbol{r}_k$

$$\begin{aligned}
\boldsymbol{r}_0 &= \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_0 = \begin{bmatrix} 2 \\ -4 \end{bmatrix} \\
\alpha_0 &= \frac{5}{14} \\
\boldsymbol{x}_1 &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \frac{5}{14}\begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{5}{7} \\ -\frac{10}{7} \end{bmatrix} \\
\boldsymbol{r}_1 &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \frac{5}{14}\begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}
\end{aligned}$$

next iteration:

$$\begin{aligned}
\alpha_1 &= \frac{5}{16} \\
\boldsymbol{x}_2 &= \begin{bmatrix} \frac{5}{7} \\ -\frac{10}{7} \end{bmatrix} + \frac{5}{16}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{75}{56} \\ -\frac{125}{112} \end{bmatrix} \\
\boldsymbol{r}_2 &= \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \frac{5}{16}\begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} \frac{7}{16} \\ -\frac{14}{16} \end{bmatrix}
\end{aligned}$$

In conclusion, $\boldsymbol{x}_1 = \begin{bmatrix} \frac{5}{7} \\ -\frac{10}{7} \end{bmatrix}$ $\boldsymbol{x}_2 = \begin{bmatrix} \frac{75}{56} \\ -\frac{125}{112} \end{bmatrix}$

(b)

$$\begin{aligned}
\boldsymbol{r}_0 &= \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_0 = \begin{bmatrix} 2 \\ -4 \end{bmatrix} \\
\alpha_0 &= \frac{\boldsymbol{r}_0^T \boldsymbol{r}_0}{\boldsymbol{r}_0^T \boldsymbol{A} \boldsymbol{r}_0} = \frac{5}{14} \\
\boldsymbol{x}_1 &= \begin{bmatrix} \frac{5}{7} \\ -\frac{10}{7} \end{bmatrix}
\end{aligned}$$

next iteration,

$$\begin{aligned}
\boldsymbol{r}_1 &= \boldsymbol{r}_0 - \alpha_0 \boldsymbol{A}\boldsymbol{r}_0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \\
\beta_0 &= \frac{\boldsymbol{r}_1^T \boldsymbol{r}_1}{\boldsymbol{r}_0^T \boldsymbol{r}_0} = \frac{5}{20} = \frac{1}{4} \\
\boldsymbol{p}_1 &= \boldsymbol{r}_1 + \beta_0 \boldsymbol{p}_0 = \begin{bmatrix} \frac{5}{2} \\ 0 \end{bmatrix} \\
\alpha_1 &= \frac{\boldsymbol{r}_1^T \boldsymbol{r}_1}{\boldsymbol{p}_1^T \boldsymbol{A} \boldsymbol{p}_1} = \frac{2}{5} \\
\boldsymbol{x}_2 &= \boldsymbol{x}_1 + \alpha_1 \boldsymbol{p}_1 = \begin{bmatrix} \frac{12}{7} \\ -\frac{10}{7} \end{bmatrix}
\end{aligned}$$

Question3

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable functions. Suppose that for every $\boldsymbol{x} \in \mathbb{R}^n$, the eigenvalues of the Hessian matrix $\nabla^2 f(\boldsymbol{X})$ lie uniformly in the interval $[m, M]$ with $0 < m \leq M < \infty$.

Prove that:

(a) The function f has a unique global minimizer $\boldsymbol{x}^*$.

(b) For all $\boldsymbol{x} \in \mathbb{R}^n$, the following inequality holds:

$$\frac{1}{2M} \|\nabla f(x)\|^2 \leq f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \leq \frac{1}{2m} \|\nabla f(\boldsymbol{x})\|^2$$

Answer

(a)

We know the second order sufficient condition:

$$\nabla^2 f(\boldsymbol{x}) \succ 0 \implies f \text{ strictly convex}$$

Since all eigenvalues of $\nabla^2 f(\boldsymbol{x}) \geq m > 0$, we know

$$\begin{aligned}
&\boldsymbol{v} \nabla^2 f(\boldsymbol{x}) \boldsymbol{v} \geq m \|\boldsymbol{v}\|^2 > 0 \text{ for any non-zero vector } \boldsymbol{v} \\
&\nabla^2 f(\boldsymbol{x}) \succ 0
\end{aligned}$$

Therefore, f is strictly convex. And we know the Theorem that for an optimization problem, where $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex on Ω and Ω is a convex set. Then the optimal solution must be unique.

To show f is coercive, we know

$$\begin{aligned}
f(\boldsymbol{y}) &= f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2} (\boldsymbol{y} - \boldsymbol{x})^T \nabla^2 f(\boldsymbol{z}) (\boldsymbol{y} - \boldsymbol{x}) \\
(\boldsymbol{y} - \boldsymbol{x})^T \nabla^2 f(\boldsymbol{z}) (\boldsymbol{y} - \boldsymbol{x}) &\geq m \|\boldsymbol{y} - \boldsymbol{x}\|^2 \\
f(\boldsymbol{y}) &\geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2} m \|\boldsymbol{y} - \boldsymbol{x}\|^2 \\
f(\boldsymbol{y}) &\geq f(\boldsymbol{0}) + \nabla f(\boldsymbol{0})^T \boldsymbol{y} + \frac{m}{2} \|\boldsymbol{y}\|^2
\end{aligned}$$

Therefore, as $\|\boldsymbol{y}\| \to \infty$, $f(\boldsymbol{y}) \to \infty$. Hence, f is coercive. In conclusion, there exists a optimal solution(coercive), and it's unique(strictly convex).

(b)

We know

$$
\begin{aligned}
f(\boldsymbol{y}) &\geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}) + \frac{1}{2}m\|\boldsymbol{y}-\boldsymbol{x}\|^2 \\
f(\boldsymbol{x}^*) &\geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{x}^*-\boldsymbol{x}) + \frac{1}{2}m\|\boldsymbol{x}^*-\boldsymbol{x}\|^2 \\
f(\boldsymbol{x}) - f(\boldsymbol{x}^*) &\leq \nabla f(\boldsymbol{x})^T(\boldsymbol{x}-\boldsymbol{x}^*) - \frac{1}{2}m\|\boldsymbol{x}-\boldsymbol{x}^*\|^2
\end{aligned}
$$

According to Cauchy-Schwarz inequality, we know:

$$
\begin{aligned}
\nabla f(\boldsymbol{x})^T\|\boldsymbol{x}-\boldsymbol{x}^*\| &\leq \|\nabla f(\boldsymbol{x})\|\|\boldsymbol{x}-\boldsymbol{x}^*\| \\
f(\boldsymbol{x}) - f(\boldsymbol{x}^*) &\leq \|\nabla f(\boldsymbol{x})\|\|\boldsymbol{x}-\boldsymbol{x}^*\| - \frac{1}{2}m\|\boldsymbol{x}-\boldsymbol{x}^*\|^2
\end{aligned}
$$

We know if $\|\boldsymbol{x}-\boldsymbol{x}^*\| = \frac{1}{m}\|\nabla f(\boldsymbol{x})\|$, we get the maximum.

$$
f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \leq \|\nabla f(\boldsymbol{x})\|\frac{1}{m}\|\nabla f(\boldsymbol{x})\| - \frac{1}{2}m(\frac{1}{m}\|\nabla f(\boldsymbol{x})\|)^2 = \frac{1}{2m}\|\nabla f(\boldsymbol{x})\|^2
$$

Therefore, $f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \leq \frac{1}{2m}\|\nabla f(\boldsymbol{x})\|^2$

And we know:

$$
\begin{aligned}
& f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}) + \frac{M}{2}\|\boldsymbol{y}-\boldsymbol{x}\|^2 \\
& \boldsymbol{y} = \boldsymbol{x} - \frac{1}{M}\nabla f(\boldsymbol{x}) \\
& f(\boldsymbol{x} - \frac{1}{M}\nabla f(\boldsymbol{x})) \leq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(-\frac{1}{M}\nabla f(\boldsymbol{x})) + \frac{M}{2}\|\frac{1}{M}\nabla f(\boldsymbol{x})\|^2 \\
& f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(-\frac{1}{M}\nabla f(\boldsymbol{x})) + \frac{M}{2}\|\frac{1}{M}\nabla f(\boldsymbol{x})\|^2 = f(\boldsymbol{x}) - \frac{1}{2M}\|\nabla f(\boldsymbol{x})\|^2 \\
& f(\boldsymbol{x}^*) \leq f(\boldsymbol{x} - \frac{1}{M}\nabla f(\boldsymbol{x})) \leq f(\boldsymbol{x}) - \frac{1}{2M}\|\nabla f(\boldsymbol{x})\|^2
\end{aligned}
$$

Therefore $f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \geq \frac{1}{2M}\|\nabla f(\boldsymbol{x})\|^2$

Question4

Consider the optimization problem $\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$, where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function. To develop a weighted gradient descent method, let $\boldsymbol{W} \in \mathbb{R}^{n\times n}$ be a symmetric positive definite (SPD) matrix. Denote by $\boldsymbol{W}^{\frac{1}{2}}$ the unique SPD square root of $\boldsymbol{W}$ (i.e., $(\boldsymbol{W}^{\frac{1}{2}})^2 = \boldsymbol{W}$) and by $\boldsymbol{W}^{-\frac{1}{2}}$ its inverse. Given the current iterate $\boldsymbol{x}^{(k)}$, define the next iterate $\boldsymbol{x}^{(k+1)}$ as the solution of the following constrained optimization problem:

$$
\begin{gathered}
\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}^{(k)}) + \langle \nabla f(\boldsymbol{x}^{(k)}), \boldsymbol{x}-\boldsymbol{x}^{(k)}\rangle \\
\text{subject to } \|\boldsymbol{W}^{\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{x}^(k))\|_2 \leq \alpha_k\|\boldsymbol{W}^{-\frac{1}{2}}\nabla f(\boldsymbol{x}^(k))\|_2
\end{gathered}
$$

where $\alpha_k > 0$ is a step-size parameter.

Answer the following questions:

(a) Derive an explicit formula for $\boldsymbol{x}^{(k+1)}$

(b) Prove that $\boldsymbol{x}^{(k+1)}$ is equivalently the unique minimizer of the unconstrained quadratic problem:

$$
\min_{\boldsymbol{x}\in\mathbb{R}^n}\left\{\nabla f(\boldsymbol{x}^{(k)}) + \langle f(\boldsymbol{x}^{(k)}), \boldsymbol{x}-\boldsymbol{x}^{(k)}\rangle + \frac{1}{2\alpha_k}\|\boldsymbol{W}^{\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{x}^{(k)})\|_2^2\right\}
$$

Answer

(a)

We know the CS inequality, that:

$$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \le \|\boldsymbol{u}\|^*_{\boldsymbol{W}^{\frac{1}{2}}} \cdot \|\boldsymbol{v}\|_{\boldsymbol{W}^{\frac{1}{2}}}$$

where $\|\boldsymbol{u}\|^*_{\boldsymbol{W}^{\frac{1}{2}}} = \|\boldsymbol{W}^{-\frac{1}{2}}\boldsymbol{u}\|_2$

Let $\boldsymbol{d} = \boldsymbol{x} - \boldsymbol{x}^{(k)}$, the problem is equivalent to:

$$\min_{\boldsymbol{d}} \langle \nabla f(\boldsymbol{x}^{(k)}), \boldsymbol{d} \rangle$$

that subject to $\|\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{d}\|_2 \le \alpha_k \|\boldsymbol{W}^{-\frac{1}{2}}\nabla f(\boldsymbol{x}^{(k)})\|_2$

Using this inequality, we can get:

$$|\langle \nabla f(\boldsymbol{x}^{(k)}), \boldsymbol{d} \rangle| \le \|\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{d}\|_2 \alpha_k \|\boldsymbol{W}^{-\frac{1}{2}}\nabla f(\boldsymbol{x}^{(k)})\|_2$$

We need $\boldsymbol{u}$ and $\boldsymbol{v}$ to be linear dependent to get the minimum. s.t.

$$\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{d} = k\boldsymbol{W}^{-\frac{1}{2}}\nabla f(\boldsymbol{x}^{(k)})$$

We can derive $\boldsymbol{d} = k\boldsymbol{W}^{-1}\nabla f(\boldsymbol{x}^{(k)})$

And we need:

$$\|\boldsymbol{W}^{\frac{1}{2}}k\boldsymbol{W}^{-1}\nabla f(\boldsymbol{x}^{(k)})\|_2 = \alpha_k \|\boldsymbol{W}^{-\frac{1}{2}}\nabla f(\boldsymbol{x}^{(k)})\|_2$$

So we get $k = \alpha_k$. Therefore, we get the answer:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k \boldsymbol{W}^{-1}\nabla f(\boldsymbol{x}^{(k)})$$

(b)

We know the problem can be treat as:

$$\min_{\boldsymbol{d}}\{\langle \nabla f(\boldsymbol{x}^{(k)}), \boldsymbol{d} \rangle + \frac{1}{2\alpha_k}\|\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{d}\|_2^2\}$$
$$\min_{\boldsymbol{d}}\{\nabla f(\boldsymbol{x}^{(k)})^T\boldsymbol{d} + \frac{1}{2\alpha_k}\boldsymbol{d}^T\boldsymbol{W}\boldsymbol{d}\}$$

Let $\Phi(\boldsymbol{d}) = \nabla f(\boldsymbol{x}^{(k)})^T\boldsymbol{d} + \frac{1}{2\alpha_k}\boldsymbol{d}^T\boldsymbol{W}\boldsymbol{d}$

It's easy to find the Hessian matrix $\boldsymbol{H} = \frac{1}{\alpha_k}\boldsymbol{W}$, and since $\alpha_k > 0$ and $\boldsymbol{W}$ is SPD, so we know $\Phi(\boldsymbol{d})$ is strictly convex.

$$\nabla\Phi(\boldsymbol{d}) = \nabla f(\boldsymbol{x}^{(k)}) + \frac{1}{\alpha_k}\boldsymbol{W}\boldsymbol{d} = 0$$
$$\boldsymbol{d} = -\alpha_k \boldsymbol{W}^{-1}\nabla f(\boldsymbol{x}^{(k)})$$

Therefore, $\boldsymbol{x}^{(k+1)}$ is equivalently the unique minimizer of this unconstrained quadratic problem.