MSBD 5007 HW2

RONG Shuo

March 30, 2025

Question1

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

- (a) $f(\mathbf{x}) = \log(e^{x_1} + e^{x_2} + \dots + e^{x_n}).$
- (b) $f(\boldsymbol{X}) = \operatorname{logdet}(\boldsymbol{X})$.

Answer

(a)

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

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

$$= \frac{e^{x_i}}{e^{x_1} + e^{x_2} + \cdots + e^{x_n}}$$

Therefore,

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

We know,

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

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

$$z\nabla^{2} f(x)z^{T} = \frac{1}{S} \sum_{i=1}^{n} e^{x_{i}} z_{i}^{2} - \frac{1}{S^{2}} (\sum_{i=1}^{n} e^{x_{i}} z_{i})^{2}$$
$$(\sum_{i=1}^{n} e^{x_{i}} z_{i})^{2} \leq (\sum_{i=1}^{n} e^{x_{i}} z_{i}^{2}) (\sum_{i=1}^{n} e^{x_{i}})$$
$$\frac{1}{S} \sum_{i=1}^{n} e^{x_{i}} z_{i}^{2} \geq \frac{1}{S^{2}} (\sum_{i=1}^{n} e^{x_{i}} z_{i})^{2}$$
$$z\nabla^{2} f(x)z^{T} \geq 0$$

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

(b)

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

$$g(t) = \operatorname{logdet}(\boldsymbol{X} + t\boldsymbol{V})$$

$$= \operatorname{logdet}(\boldsymbol{X}^{\frac{1}{2}}(\boldsymbol{I} + t\boldsymbol{X}^{-\frac{1}{2}}\boldsymbol{V}\boldsymbol{X}^{-\frac{1}{2}})\boldsymbol{X}^{\frac{1}{2}})$$

$$= \operatorname{logdet}\boldsymbol{X} + \operatorname{logdet}(\boldsymbol{I} + t\boldsymbol{\Lambda})$$

$$= \operatorname{logdet}\boldsymbol{X} + \sum_{i=1}^{n} \operatorname{log}(1 + t\lambda_{i})$$

where λ_i is the eigenvalues of Λ

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

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

Question2

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

- (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 $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{A} \mathbf{r}_k$

$$egin{aligned} oldsymbol{r}_0 &= oldsymbol{b} - oldsymbol{A} oldsymbol{x}_0 &= egin{bmatrix} 2 \ -4 \end{bmatrix} \ lpha_0 &= rac{5}{14} \ oldsymbol{x}_1 &= egin{bmatrix} 0 \ 0 \end{bmatrix} + rac{5}{14} egin{bmatrix} 2 \ 4 \end{bmatrix} = egin{bmatrix} rac{5}{7} \ -rac{10}{7} \end{bmatrix} \ oldsymbol{r}_1 &= egin{bmatrix} 0 \ 0 \end{bmatrix} - rac{5}{14} egin{bmatrix} 2 \ 4 \end{bmatrix} = egin{bmatrix} 2 \ 1 \end{bmatrix} \end{aligned}$$

next iteration:

$$\alpha_1 = \frac{5}{16}$$

$$x_2 = \begin{bmatrix} \frac{5}{7} \\ -\frac{10}{7} \end{bmatrix} + \frac{5}{16} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{75}{56} \\ -\frac{125}{112} \end{bmatrix}$$

$$r_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \frac{5}{16} \begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} \frac{7}{16} \\ -\frac{14}{16} \end{bmatrix}$$

In conclusion, $\boldsymbol{x}_1 = \begin{bmatrix} \frac{5}{7} \\ -\frac{1}{10} \end{bmatrix} \boldsymbol{x}_2 = \begin{bmatrix} \frac{75}{56} \\ -\frac{125}{112} \end{bmatrix}$

(b)

$$egin{aligned} oldsymbol{r}_0 &= oldsymbol{b} - oldsymbol{A} oldsymbol{x}_0 &= egin{bmatrix} oldsymbol{r}_0^T oldsymbol{r}_0 \ oldsymbol{r}_0^T oldsymbol{A} oldsymbol{r}_0 &= rac{5}{14} \ oldsymbol{x}_1 &= egin{bmatrix} rac{5}{7} \ -rac{10}{7} \end{bmatrix} \end{aligned}$$

next iteration,

$$egin{aligned} oldsymbol{r}_1 &= oldsymbol{r}_0 - lpha_0 oldsymbol{A} oldsymbol{r}_0 &= egin{bmatrix} 2 \ 1 \end{bmatrix} \ eta_0 &= oldsymbol{r}_1^T oldsymbol{r}_1 = rac{5}{20} = rac{1}{4} \ oldsymbol{p}_1 &= oldsymbol{r}_1 + eta_0 oldsymbol{p}_0 &= egin{bmatrix} rac{5}{2} \ 0 \end{bmatrix} \ lpha_1 &= oldsymbol{r}_1^T oldsymbol{r}_1 &= rac{2}{5} \ oldsymbol{x}_2 &= oldsymbol{x}_1 + lpha_1 oldsymbol{p}_1 &= egin{bmatrix} rac{12}{70} \ -rac{10}{7} \end{bmatrix} \end{aligned}$$

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 $0 < m \le M < \infty$.

Prove that:

- (a) The function f has a unique global minimizer x^* .
- (b) For all $x \in \mathbb{R}^n$, the following inequality holds:

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

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 $\nabla^2 f(x) \ge m > 0$, we know

$$\mathbf{v}\nabla^2 f(\mathbf{x})\mathbf{v} \ge m\|\mathbf{v}\|^2 > 0$$
 for any non-zero vector \mathbf{v}
 $\nabla^2 f(\mathbf{x}) \succ 0$

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

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2} (\boldsymbol{y} - \boldsymbol{x})^T \nabla^2 f(\boldsymbol{z}) (\boldsymbol{y} - \boldsymbol{x})$$
$$(\boldsymbol{y} - \boldsymbol{x})^T \nabla^2 f(\boldsymbol{z}) (\boldsymbol{y} - \boldsymbol{x}) \ge m \|\boldsymbol{y} - \boldsymbol{x}\|^2$$
$$f(\boldsymbol{y}) \ge f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2} m \|\boldsymbol{y} - \boldsymbol{x}\|^2$$
$$f(\boldsymbol{y}) \ge f(\boldsymbol{0}) + \nabla f(\boldsymbol{0})^T \boldsymbol{y} + \frac{m}{2} \|\boldsymbol{y}\|^2$$

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

(b)

We know

$$f(y) \ge f(x) + \nabla f(x)^{T} (y - x) + \frac{1}{2} m \|y - x\|^{2}$$

$$f(x^{*}) \ge f(x) + \nabla f(x)^{T} (x^{*} - x) + \frac{1}{2} m \|x^{*} - x\|^{2}$$

$$f(x) - f(x^{*}) \le \nabla f(x)^{T} (x - x^{*}) - \frac{1}{2} m \|x - x^{*}\|^{2}$$

According to Cauchy-Schwarz inequality, we know:

$$\nabla f(x)^{T} \|x - x^{*}\| \le \|\nabla f(x)\| \|x - x^{*}\|$$

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

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

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

Therefore, $f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \le \frac{1}{2m} \|\nabla f(\boldsymbol{x})\|^2$ And we know:

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

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

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:

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

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

$$\min_{oldsymbol{x} \in \mathbb{R}^n} \left\{
abla f(oldsymbol{x}^{(k)}) + \langle f(oldsymbol{x}^{(k)}), oldsymbol{x} - oldsymbol{x}^{(k)}
angle + rac{1}{2lpha_k} \|oldsymbol{W}^{rac{1}{2}}(oldsymbol{x} - oldsymbol{x}^{(k)})\|_2^2
ight\}$$

Answer

(a)

We know the CS inequality, that:

$$|\langle oldsymbol{u}, oldsymbol{v}
angle| \leq \|oldsymbol{u}\|_{oldsymbol{W}^{rac{1}{2}}}^* \cdot \|oldsymbol{v}\|_{oldsymbol{W}^{rac{1}{2}}}$$

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

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

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

that subject to $\|\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{d}\|_{2} \leq \alpha_{k} \|\boldsymbol{W}^{-\frac{1}{2}}\nabla f(\boldsymbol{x}^{(k)})\|_{2}$

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

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:

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

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.

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

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