Problem Solutions

Fu Qiang

Problem1. Let f be a continuously differentiable function on $\mathbb{R}^n$. Suppose there exists a positive constant L such that ∇f is L-Lipschitz continuous, namely

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x-y\|_2 \quad \text{for all } x, y \in \mathbb{R}^n.$$

(a) Prove that

$$\inf_{y\in\mathbb{R}^n} f(y) \leq f(x) - \frac{1}{2L}\|\nabla f(x)\|_2^2 \quad \text{for all } x \in \mathbb{R}^n.$$

(b) If in addition f is convex, prove that

$$f(x) - f(y) - [\nabla f(x)]^T(x-y) \leq -\frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2.$$

Solution.

(a)

Lemma: f is a continuously differentiable function on $\mathbb{R}^n$. Then for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \frac{L}{2}\|\mathbf{y}-\mathbf{x}\|_2^2$$

Proof: Let $g(t) = f(\mathbf{x} + t(\mathbf{y}-\mathbf{x}))$. Then $g'(t) = \langle \nabla f(\mathbf{x} + t(\mathbf{y}-\mathbf{x})), \mathbf{y}-\mathbf{x}\rangle$. By the Newton - Leibniz formula, we can obtain

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y}-\mathbf{x})), \mathbf{y}-\mathbf{x}\rangle dt$$

The above equation is equivalent to

$$f(\mathbf{y}) - f(\mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle + \int_0^1 \langle \nabla f(\mathbf{x} + t(\mathbf{y}-\mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle dt$$

From the Lipschitz continuous and Cauchy - Schwarz inequality, we can get

$$
\begin{aligned}
|f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle| &= \left| \int_0^1 \langle \nabla f(\mathbf{x}+t(\mathbf{y}-\mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle dt \right| \\
&\leq \int_0^1 \left| \langle \nabla f(\mathbf{x}+t(\mathbf{y}-\mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle \right| dt \\
&\leq \int_0^1 \|\nabla f(\mathbf{x}+t(\mathbf{y}-\mathbf{x})) - \nabla f(\mathbf{x})\|_2 \cdot \|\mathbf{y}-\mathbf{x}\|_2 dt \\
&\leq \int_0^1 tL\|\mathbf{y}-\mathbf{x}\|_2^2 dt \\
&= \frac{L}{2}\|\mathbf{y}-\mathbf{x}\|_2^2
\end{aligned}
$$

Therefore

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \frac{L}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 \qquad \square$$

Fix $\mathbf{x} \in \mathbb{R}^n$. Define the function

$$q(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \frac{L}{2}\|\mathbf{y}-\mathbf{x}\|_2^2$$

From the Lemma, $f(\mathbf{y}) \leq q(\mathbf{y})$ holds for all $\mathbf{y} \in \mathbb{R}^n$. Therefore,

$$\inf_{\mathbf{y}\in\mathbb{R}^n} f(\mathbf{y}) \leq \inf_{\mathbf{y}\in\mathbb{R}^n} q(\mathbf{y})$$

Let $\mathbf{d} = \mathbf{y} - \mathbf{x}$. Then

$$q(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T\mathbf{d} + \frac{L}{2}\|\mathbf{d}\|_2^2$$

The gradient of this quadratic function with respect to $\mathbf{d}$ is $\nabla_q = \nabla f(\mathbf{x}) + L\mathbf{d}$. Set this gradient to zero:

$$\nabla f(\mathbf{x}) + L\mathbf{d} = 0 \implies \mathbf{d} = -\frac{1}{L}\nabla f(\mathbf{x})$$

Therefore, the minimum point is

$$\mathbf{y}^* = \mathbf{x} + \mathbf{d} = \mathbf{x} - \frac{1}{L}\nabla f(\mathbf{x})$$

Substitute into $q(y)$ to get the minimum value:

$$q(\mathbf{y}^*) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T\left(-\frac{1}{L}\nabla f(\mathbf{x})\right) + \frac{L}{2}\left\|-\frac{1}{L}\nabla f(\mathbf{x})\right\|_2^2 = f(\mathbf{x}) - \frac{1}{2L}\|\nabla f(\mathbf{x})\|_2^2$$

That is,

$$\inf_{\mathbf{y}\in\mathbb{R}^n} q(\mathbf{y}) = q(\mathbf{y}^*) = f(\mathbf{x}) - \frac{1}{2L}\|\nabla f(\mathbf{x})\|_2^2$$

From $f(\mathbf{y}) \leq q(\mathbf{y})$ and $\inf_{\mathbf{y}} f(\mathbf{y}) \leq \inf_{\mathbf{y}} q(\mathbf{y})$, we get

$$\inf_{\mathbf{y}\in\mathbb{R}^n} f(\mathbf{y}) \leq f(\mathbf{x}) - \frac{1}{2L}\|\nabla f(\mathbf{x})\|_2^2$$

This inequality holds for any $\mathbf{x} \in \mathbb{R}^n$.

(b) Fix $\mathbf{x} \in \mathbb{R}^n$. Define the function

$$p(\mathbf{y}) = f(\mathbf{y}) - \nabla f(\mathbf{x})^T \mathbf{y}$$

Since f is a convex function, after subtracting a linear term, p is still a convex function. Because ∇f is L-Lipschitz continuous, ∇p is also L-Lipschitz continuous.
At the point $\mathbf{y} = \mathbf{x}$, calculate the gradient:

$$\nabla p(\mathbf{x}) = 0$$

Since p is a convex function and $\nabla p(\mathbf{x}) = 0$, p attains the global minimum at $\mathbf{x}$, that is

$$\inf_{\mathbf{z}\in\mathbb{R}^n} p(\mathbf{z}) = p(\mathbf{x})$$

Problem (a) shows that

$$\inf_{\mathbf{z}\in\mathbb{R}^n} p(\mathbf{z}) \leq p(\mathbf{y}) - \frac{1}{2L}\|\nabla p(\mathbf{y})\|_2^2, \quad \forall \mathbf{y} \in \mathbb{R}^n$$

Substitute $\inf_{\mathbf{z}\in\mathbb{R}^n} p(\mathbf{z}) = p(\mathbf{x})$, we get

$$p(\mathbf{x}) \leq p(\mathbf{y}) - \frac{1}{2L}\|\nabla p(\mathbf{y})\|_2^2$$

Equivalently,

$$p(\mathbf{x}) - g(\mathbf{y}) \leq -\frac{1}{2L}\|\nabla g(\mathbf{y})\|_2^2$$

Then, we inspect $p(\mathbf{x}) - p(\mathbf{y})$ and $\|\nabla p(\mathbf{y})\|_2^2$

$$\begin{aligned} p(\mathbf{x}) - p(\mathbf{y}) &= \left[f(\mathbf{x}) - \nabla f(\mathbf{x})^T x\right] - \left[f(\mathbf{y}) - \nabla f(\mathbf{x})^T \mathbf{y}\right] \\ &= f(\mathbf{x}) - f(\mathbf{y}) - \nabla f(\mathbf{x})^T(\mathbf{x} - \mathbf{y}) \end{aligned}$$

$$\|\nabla p(\mathbf{y})\|_2^2 = \|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\|_2^2 = \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2$$

Summarizing

$$f(\mathbf{x}) - f(\mathbf{y}) - \nabla f(\mathbf{x})^T(\mathbf{x} - \mathbf{y}) \leq -\frac{1}{2L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2$$

Problem2.

Given a symmetric matrix $A \in \mathbb{R}^{n\times n}$ and a vector $b \in \mathbb{R}^n$, define

$$q(x) = \frac{1}{2}x^T Ax - b^T x, \quad x \in \mathbb{R}^n.$$

Prove that the following statements are equivalent.

(a) q is bounded from below.

(b) $A \succeq 0$ and $b \in \mathrm{range}(A)$.

(c) q has a local minimum.

(d) q has a global minimum.

Solution.

(a) ⇒ (b)

Assume that A is not positive semi - definite. Then there exists a vector $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} \neq \mathbf{0}$, such that $\mathbf{v}^T A\mathbf{v} < 0$. Consider $\mathbf{x}_t = t\mathbf{v}$ where $t \in \mathbb{R}$. Substitute it into q:

$$q(\mathbf{x}_t) = q(t\mathbf{v}) = \frac{1}{2}(t\mathbf{v})^T A(t\mathbf{v}) - B^T(t\mathbf{v}) = \frac{1}{2}t^2(\mathbf{v}^T A\mathbf{v}) - t(B^T\mathbf{v})$$

Since $\mathbf{v}^T A\mathbf{v} < 0$, when $t \to \infty$, the quadratic term $\frac{1}{2}t^2(\mathbf{v}^T A\mathbf{v}) \to -\infty$, and another term $-t(B^T\mathbf{v})$, so $q(t\mathbf{v}) \to -\infty$. This contradicts the fact that q is bounded below. Therefore, A must be positive semi - definite, namely, $A \succeq 0$.

Since A is symmetric, there is an orthogonal decomposition $\mathbb{R}^n = \mathrm{range}(A) \oplus \ker(A)$. Let $B = B_r + B_n$, where $B_r \in \mathrm{range}(A)$, $B_n \in \ker(A)$, and $B_r^T B_n = 0$. Assume $B_n \neq \mathbf{0}$. Consider $\mathbf{x}_t = tB_n$ where $t \in \mathbb{R}$. Substitute it into q:

$$q(\mathbf{x}_t) = q(tB_n) = \frac{1}{2}(tB_n)^T A(tB_n) - B^T(tB_n) = \frac{1}{2}t^2(B_n^T AB_n) - t(B^T B_n)$$

Because $B_n \in \ker(A)$, we have $AB_n = \mathbf{0}$, so $B_n^T AB_n = 0$. Further:

$$B^T B_n = (B_r + B_n)^T B_n = B_r^T B_n + B_n^T B_n = 0 + \|B_n\|^2 > 0 \quad (\text{since } B_n \neq \mathbf{0})$$

Then:

$$q(tB_n) = -t\|B_n\|^2$$

When $t \to \infty$, $q(tB_n) \to -\infty$, which contradicts the fact that q is bounded below. Therefore, $B_n = \mathbf{0}$, that is, $B \in \mathrm{range}(A)$.

(b) $\Rightarrow$ (c)

Assume that $A \succeq 0$ and $\mathbf{b} \in \text{range}(A)$. Then there exists $\mathbf{x}^* \in \mathbb{R}^n$ such that $A\mathbf{x}^* = \mathbf{b}$. Calculate the gradient:

$$\nabla q(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$$

At $\mathbf{x}^*$:

$$\nabla q(\mathbf{x}^*) = A\mathbf{x}^* - \mathbf{b} = \mathbf{0}$$

$$\nabla^2 q(\mathbf{x}) = A \succeq 0$$

For any direction $\mathbf{d} \in \mathbb{R}^n$ and sufficiently small $t > 0$, at $\mathbf{x}^*$, we have:

$$q(\mathbf{x}^* + t\mathbf{d}) = q(\mathbf{x}^*) + \underbrace{t(\nabla q(\mathbf{x}^*)^T \mathbf{d})}_{=0} + \frac{t^2}{2}\mathbf{d}^T A\mathbf{d} + O(t^3)$$

Since $A \succeq 0$, we have $\mathbf{d}^T A\mathbf{d} \geq 0$. Thus:

$$q(\mathbf{x}^* + t\mathbf{d}) - q(\mathbf{x}^*) = \frac{t^2}{2}\mathbf{d}^T A\mathbf{d} \geq 0$$

Therefore, in a neighborhood of $\mathbf{x}^*$, $q(\mathbf{x}) \geq q(\mathbf{x}^*)$, so q has a local minimum.

(c) $\Rightarrow$ (d)

Assume that $\mathbf{x}^*$ is a local minimum point. At $\mathbf{x}^*$: The gradient is zero: $\nabla q(\mathbf{x}^*) = A\mathbf{x}^* - B = \mathbf{0}$, so $A\mathbf{x}^* = B$, that is, $B \in \text{range}(A)$. The Hessian matrix A is positive semi - definite is a local minimum point, so, $A \succeq 0$.

From $A \succeq 0$ and $A\mathbf{x}^* = B$, consider the function values:

$$q(\mathbf{x}) - q(\mathbf{x}^*) = \left(\frac{1}{2}\mathbf{x}^T A\mathbf{x} - B^T\mathbf{x}\right) - \left(\frac{1}{2}(\mathbf{x}^*)^T A\mathbf{x}^* - B^T\mathbf{x}^*\right)$$

Substitute $B = A\mathbf{x}^*$:

$$B^T\mathbf{x} = (A\mathbf{x}^*)^T\mathbf{x} = \mathbf{x}^T A\mathbf{x}^*, \quad B^T\mathbf{x}^* = (A\mathbf{x}^*)^T\mathbf{x}^* = (\mathbf{x}^*)^T A\mathbf{x}^*$$

Therefore:

$$q(\mathbf{x}) - q(\mathbf{x}^*) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{x}^T A\mathbf{x}^* - \frac{1}{2}(\mathbf{x}^*)^T A\mathbf{x}^* + (\mathbf{x}^*)^T A\mathbf{x}^* = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{x}^T A\mathbf{x}^* + \frac{1}{2}(\mathbf{x}^*)^T A\mathbf{x}^*$$

Thus:

$$q(\mathbf{x}) - q(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T A(\mathbf{x} - \mathbf{x}^*)$$

Because $A \succeq 0$, we have $(\mathbf{x}-\mathbf{x}^*)^T A(\mathbf{x}-\mathbf{x}^*) \geq 0$. So $q(\mathbf{x}) - q(\mathbf{x}^*) \geq 0$, that is, $q(\mathbf{x}) \geq q(\mathbf{x}^*)$ holds for all $\mathbf{x} \in \mathbb{R}^n$. Therefore, $\mathbf{x}^*$ is a global minimum point, that is, q has a global minimum. So $(c) \Rightarrow (d)$.

(d) $\Rightarrow$ **(a)** Easy to prove.

Problem3.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. For $t \in \mathbb{R}$, define

$$\mathcal{L}(t) = \{x \in \mathbb{R}^n : f(x) \leq t\}$$

Solution.

For $\forall t < t_0$, we can get $\mathcal{L}(t) \subset \mathcal{L}(t_0)$. Because of the Boundedness of $\mathcal{L}(t_0)$, $\mathcal{L}(t)$ is bounded. Therefore, we We just need to prove $\forall t > t_0$, $\mathcal{L}(t)$ is bounded.

Assume exists $t_1 > t_0$ such that $\mathcal{L}(t_1)$ is unbounded. Then there exists a sequence $\{x_k\} \subset \mathcal{L}(t_1)$ satisfying $\|x_k\|_2 \to \infty$. Take $x_0 \in \mathcal{L}(t_0)$.

Consider the direction vectors $\mathbf{d}_k = \frac{\mathbf{x}_k - \mathbf{x}_0}{\|\mathbf{x}_k - \mathbf{x}_0\|_2}$. Since $\|\mathbf{d}_k\|_2 = 1$ and the unit sphere is compact, there exists a subsequence that converges to a unit vector $\mathbf{d}$, without loss of generality, assume $\mathbf{d}_k$ converges to $\mathbf{d}$.

Fix $\alpha > 0$, and let $\theta_k = \frac{\alpha}{\|\mathbf{x}_k - \mathbf{x}_0\|_2}$. When k is sufficiently large, $\theta_k \in (0, 1)$ and $\theta_k \to 0$. Let:

$$\mathbf{y}_k = (1-\theta_k)\mathbf{x}_0 + \theta_{\mathbf{x}} x_k = \mathbf{x}_0 + \alpha \mathbf{d}_k$$

By convexity:

$$f(\mathbf{y}_k) \leq (1-\theta_k)t_0 + \theta_k t_1 = t_0 + \theta_k(t_1 - t_0)$$

For any $\epsilon > 0$, when k is large enough, $\theta_k < \frac{\epsilon}{t_1 - t_0}$. Thus:

$$f(\mathbf{y}_k) < t_0 + \epsilon \implies y_k \in \mathcal{L}(t_0 + \epsilon)$$

From $\mathbf{d}_k \to \mathbf{d}$, we know that $\mathbf{y}_k \to \mathbf{y} = \mathbf{x}_0 + \alpha \mathbf{d}$. Using the continuity of f :

$$f(\mathbf{y}) = \lim f(\mathbf{y}_k) \leq t_0 + \epsilon$$

Due to the arbitrariness of $\epsilon > 0$, we get $f(\mathbf{y}) \leq t_0$, that is, $\mathbf{y} \in \mathcal{L}(t_0)$.

Since $\alpha > 0$ is arbitrary, the ray $\{\mathbf{x}_0+\alpha\mathbf{d} : \alpha \geq 0\} \subset \mathcal{L}(t_0)$, but this ray is unbounded, which contradicts the boundedness of $\mathcal{L}(t_0)$.

Therefore, for all $t > t_0$, $\mathcal{L}(t)$ is bounded.

Problem4.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and $K \subset \mathbb{R}^n$ be a compact set. Prove that f is Lipschitz continuous on K.

Solution.

Lemma: Subgradients on a compact set must be bounded.

Proof:

Take $\delta > 0$, and define $K_\delta = \{\mathbf{y} : \mathrm{d}(\mathbf{y}, K) = inf_{\mathbf{z}\in K}\|\mathbf{y}-\mathbf{z}\| \leq \delta\}$. Since K is compact, K_δ is compact.

Since f is convex function, f is continuous on the compact set K_δ, so there $\exists$:

$$M_\delta = \sup_{\mathbf{z}\in K_\delta} f(\mathbf{z}), \quad m_\delta = \inf_{\mathbf{z}\in K_\delta} f(\mathbf{z}), \quad \omega = M_\delta - m_\delta < \infty$$

For any $\mathbf{x} \in K$ and $g \in \partial f(\mathbf{x})$, let $d = g/\|g\|$ (if $g \neq 0$) and $\mathbf{y} = \mathbf{x}+\delta d \in \overline{B}(\mathbf{x}, \delta) \subset K_\delta$. By the definition of subgradients:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + g^\top(\mathbf{y} - \mathbf{x}) = f(\mathbf{x}) + \delta\|g\|$$

From $f(\mathbf{y}) \leq M_\delta$ and $f(\mathbf{x}) \geq m_\delta$, we can get:

$$\delta\|g\| \leq f(\mathbf{y}) - f(\mathbf{x}) \leq \omega \implies \|g\| \leq \frac{\omega}{\delta} \qquad \square$$

For any $\mathbf{x}, \mathbf{y} \in K$, consider $(1-t)\mathbf{x} + t\mathbf{y} \quad (t \in [0, 1])$. By convexity, there $\exists g_t \in \partial f((1-t)\mathbf{x} + t\mathbf{y})$ such that:

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \frac{d}{dt} f((1-t)\mathbf{x} + t\mathbf{y})dt = \int_0^1 g_t^\top(\mathbf{y} - \mathbf{x})dt$$

$$|f(\mathbf{y}) - f(\mathbf{x})| \leq \int_0^1 |g_t^\top(\mathbf{y} - \mathbf{x})|dt \leq \int_0^1 \|g_t\| \cdot \|\mathbf{y} - \mathbf{x}\|dt$$

From the lemma we know the subgradient on K set is bounded. And

$$|f(\mathbf{y}) - f(\mathbf{x})| \leq \frac{\omega}{\delta}\|\mathbf{y} - \mathbf{x}\|$$

Therefore, the convex function f is L- Lipschitz continuous on the compact set K, where $L = \frac{\omega}{\delta}$.

Problem5.

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable convex function, ∇f is L-Lipschitz continuous, and x^* is a minimizer of f. Prove that $\|x - t\nabla f(x) - x^*\|_2 \leq \|x - x^*\|_2$ for all $t \in [0, 2/L]$.

Solution.

We need prove that $\| \mathbf{x} - t\nabla f(\mathbf{x}) - \mathbf{x}^*\|_2 \leq \|\mathbf{x} - \mathbf{x}^*\|_2$.

That is

$$\| \mathbf{x} - t\nabla f(\mathbf{x}) - \mathbf{x}^*\|_2^2 \leq \|\mathbf{x} - \mathbf{x}^*\|_2^2$$

Consider that $\| \mathbf{x} - t\nabla f(\mathbf{x}) - \mathbf{x}^*\|_2^2$

$$\| \mathbf{x} - t\nabla f(\mathbf{x}) - \mathbf{x}^*\|_2^2 = \|\mathbf{x} - \mathbf{x}^*\|_2^2 + t^2\| \nabla f(\mathbf{x})\|_2^2 - 2t\langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{x}^*\rangle$$

That is prove

$$t^2\| \nabla f(\mathbf{x})\|_2^2 \leq 2t\langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{x}^*\rangle$$

Because of $t > 0$, collating the above inequation, that is

$$t\| \nabla f(\mathbf{x})\|_2^2 \leq 2\langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{x}^*\rangle$$

f is convex function, and $\nabla f(\mathbf{x})$ is L-Lipschitz continuous. From the question(1), we know that

$$\frac{1}{L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2 \leq \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y}\rangle$$

Let $\mathbf{y} = \mathbf{x}^*$, that is

$$\frac{1}{L}\|\nabla f(\mathbf{x})\|_2^2 \leq \langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{x}^*\rangle$$

So

$$\frac{2}{L}\|\nabla f(\mathbf{x})\|_2^2 \leq 2\langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{x}^*\rangle$$

We can easily know that, $\| \mathbf{x} - t\nabla f(\mathbf{x}) - \mathbf{x}^*\|_2 \leq \|\mathbf{x} - \mathbf{x}^*\|_2$. where $t \in (0, \frac{2}{L})$. If and only if $\mathbf{x} = \mathbf{x}^*$ the inequation takes an equal sign, otherwise the inequation strictly holds.

Problem6.

Find a convex function that is differentiable on an open convex set but not continuously differentiable on the same set —or prove that such a function does not exist.

Solution.

I don't think such a function exists.

Assume that there exists a convex function f that is differentiable but not continuously differentiable on an open convex set $U \subseteq \mathbb{R}^n$. Then there exists a point $\mathbf{x} \in U$ and a sequence $\{\mathbf{x}_k\} \subseteq U$ converging to $\mathbf{x}$ (i.e., $\mathbf{x}_k \to \mathbf{x}$), but the sequence of gradients $\{\nabla f(\mathbf{x}_k)\}$ does not converge to $\nabla f(\mathbf{x})$. That is:

$$\nabla f(\mathbf{x}_k) \nrightarrow \nabla f(\mathbf{x}) \quad \text{as} \quad k \to \infty$$

Since U is an open set and $\mathbf{x} \in U$, there exists a neighborhood $K \subseteq U$ containing $\mathbf{x}$. Because f is convex on U, it is Lipschitz continuous on K . Let the Lipschitz constant be L. If f is differentiable, then the gradient is bounded on K: for all $\mathbf{y} \in K$, $\|\nabla f(\mathbf{y})\| \leq L$.

The sequence $\{\nabla f(\mathbf{x}_k)\}$ is bounded, so it has a convergent subsequence. Assume the entire sequence converges (otherwise take a subsequence), that is:

$$\nabla f(\mathbf{x}_k) \to \mathbf{g} \quad \text{as} \quad k \to \infty$$

where $\mathbf{g} \neq \nabla f(\mathbf{x})$.

Since f is convex and differentiable on U, for any $\mathbf{y} \in U$, the subgradient inequality holds:

$$f(\mathbf{y}) \geq f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{y} - \mathbf{x}_k \rangle$$

Take the limit as $k \to \infty$: $f(\mathbf{x}_k) \to f(\mathbf{x})$ (because f is continuous; a convex function is continuous on an open set). $\nabla f(\mathbf{x}_k) \to \mathbf{g}$. $\mathbf{x}_k \to \mathbf{x}$, so $\mathbf{y} - \mathbf{x}_k \to \mathbf{y} - \mathbf{x}$.

Thus:

$$f(\mathbf{y}) \geq \lim_{k \to \infty} \left[f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{y} - \mathbf{x}_k \rangle \right] = f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle$$

This shows that $\mathbf{g}$ is a subgradient of f at $\mathbf{x}$, i.e., $\mathbf{g} \in \partial f(\mathbf{x})$. But $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$. Therefore, we must have $\mathbf{g} = \nabla f(\mathbf{x})$, which contradicts the assumption $\mathbf{g} \neq \nabla f(\mathbf{x})$.

Problem7.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function. Given any $d \in \mathbb{R}^n$ with $\|d\|_2 = 1$, the function $t \mapsto f(td)$ has a local minimum at $t^* = 0$. Is it guaranteed that f has a local minimum at $x^* = 0$?

Solution.

Let $g_d(t) = f(t\mathbf{d})$, $g_d'(t) = \langle \nabla f(t\mathbf{d}), \mathbf{d}\rangle$, $g_d''(t) = \langle \nabla^2 f(t\mathbf{d})\mathbf{d}, \mathbf{d}\rangle = \|\nabla^2 f(t\mathbf{d})\mathbf{d}\|_2^2$. Since $t \mapsto f(t\mathbf{d})$ has a local minimum at $t^* = 0$, we can know that

$$g_d'(0) = \langle \nabla f(0), \mathbf{d}\rangle = 0$$

$$g_d''(0) = \nabla^2 f(0) \succeq 0$$

Suppose f does not have a local minimum at $\mathbf{0}$. Then there exists a sequence $\{\mathbf{x}_k\} \subset \mathbb{R}^n$ such that:

$$\mathbf{x}_k \to \mathbf{0}, \quad f(\mathbf{x}_k) < f(\mathbf{0}), \quad \forall k \in \mathbb{N}$$

and define the unit vector:

$$\mathbf{d}_k = \frac{\mathbf{x}_k}{\|\mathbf{x}_k\|_2}$$

The unit sphere $S^{n-1} = \{\mathbf{d} \in \mathbb{R}^n : \|\mathbf{d}\|_2 = 1\}$ is compact, so $\{\mathbf{d}_k\}$ has a convergent subsequence. Without loss of generality, let $\mathbf{d}_k \to \mathbf{d}_*$, and $\|\mathbf{x}_k\|_2 \to 0$.

$$g_{\mathbf{d}_*}(t) = f(t\mathbf{d}_*)$$

By the condition, $g_{\mathbf{d}_*}$ has a local minimum at $t = 0$, so there exists $\delta > 0$ such that:

$$g_{\mathbf{d}_*}(t) \geq g_{\mathbf{d}_*}(0) = f(\mathbf{0}) \quad \forall t \in (-\delta, \delta)$$

$\forall \epsilon > 0$, since $\|\mathbf{x}_k\|_2 \to 0$ and $\mathbf{d}_k \to \mathbf{d}_*$, for sufficiently large k, $\|\mathbf{x}_k\|_2 < \delta$ and $\|\mathbf{d}_k - \mathbf{d}_*\|_2 < \epsilon$.

Because f is continuous and continuous on compact sets. Consider the points $\|\mathbf{x}_k\|_2\mathbf{d}_k$ and $\|\mathbf{x}_k\|_2\mathbf{d}_*$:

$$\|\|\mathbf{x}_k\|_2\mathbf{d}_k - \|\mathbf{x}_k\|_2\mathbf{d}_*\|_2 = \|\mathbf{x}_k\|_2\|\mathbf{d}_k - \mathbf{d}_*\|_2 \to 0 \quad (k \to \infty)$$

By continuity:

$$\lim_{k\to\infty} |f(\|\mathbf{x}_k\|_2\mathbf{d}_k) - f(\|\mathbf{x}_k\|_2\mathbf{d}_*)| = 0$$

But by definition:

$$f(\|\mathbf{x}_k\|_2\mathbf{d}_k) = f(\mathbf{x}_k) < f(\mathbf{0})$$

and since $\|\mathbf{x}_k\|_2 < \delta$, we have:

$$f(\|\mathbf{x}_k\|_2\mathbf{d}_*) = g_{\mathbf{d}_*}(\|\mathbf{x}_k\|_2) \geq f(\mathbf{0})$$

Thus:

$$f(\|\mathbf{x}_k\|_2\mathbf{d}_*) - f(\|\mathbf{x}_k\|_2\mathbf{d}_k) \geq f(\mathbf{0}) - f(\mathbf{x}_k) > 0$$

Take the limit as $k \to \infty$:

$$\lim_{k\to\infty} [f(\|\mathbf{x}_k\|_2\mathbf{d}_*) - f(r_k\mathbf{d}_k)] \geq \lim_{k\to\infty} [f(\mathbf{0}) - f(\mathbf{x}_k)] > 0$$

we can get:

$$0 \geq \lim_{k\to\infty} [f(\mathbf{0}) - f(\mathbf{x}_k)] > 0$$

The contradiction shows that the assumption is wrong, so f has a local minimum at $\mathbf{x}^*$

Problem8.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function. Suppose that there exists a unique point $x^* \in \mathbb{R}^n$ such that $\nabla f(x^*) = 0$. In addition, x^* is a local minimizer of f. Is it guaranteed that x^* is a global minimizer of f?

Solution.

x^* is not necessarily a global minimizer. The following is a counterexample.

Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x, y) = x^2 + y^2(1 - x)^3$$

This function is twice continuously differentiable.

Calculate the gradient:

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

where

$$\frac{\partial f}{\partial x} = 2x - 3y^2(1 - x)^2, \quad \frac{\partial f}{\partial y} = 2y(1 - x)^3$$

$\frac{\partial f}{\partial y} = 0$ gives $2y(1-x)^3 = 0$, so $y = 0$ or $x = 1$.

If $x = 1$, then $\frac{\partial f}{\partial x} = 2(1) - 3y^2(1-1)^2 = 2 \neq 0$. Thus, $x = 1$ does not satisfy the condition that the gradient is zero.

If $y = 0$, then $\frac{\partial f}{\partial x} = 2x - 0 = 2x$. Setting this equal to zero gives $x = 0$.

Therefore, the unique critical point is $(x, y) = (0, 0)$.

At $(0, 0)$, $f(0, 0) = 0$.

The Hessian matrix is:

$$H_f(x, y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

where

$$\frac{\partial^2 f}{\partial x^2} = 2 + 6y^2(1-x), \quad \frac{\partial^2 f}{\partial y^2} = 2(1-x)^3, \quad \frac{\partial^2 f}{\partial x \partial y} = -6y(1-x)^2$$

At $(0, 0)$:

$$H_f(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

The eigenvalues are $2 > 0$, so the matrix is positive definite. Therefore, $(0, 0)$ is a local minimizer.

Take the point $(2, 3)$:

$$f(2, 3) = 2^2 + 3^2(1-2)^3 = 4 + 9 \cdot (-1) = 4 - 9 = -5 < 0 = f(0, 0)$$

Thus, $f(2, 3) < f(0, 0)$, so $(0, 0)$ is not a global minimizer.

x^* is not necessarily a global minimizer.

Problem9.

Let $\{X_k\}$ be a sequence of independent random variables such that (a) for each $k \geq 1$, X_k is either 0 or 1; (b) there exists a constant $p \in (0, 1)$ such that $\mathbb{P}(X_k = 1) \geq p$ for each $k \geq 1$.

For all $t \in [0, p]$, prove that

$$\mathbb{P}\left(\sum_{k=1}^{n} X_k \leq tn\right) \leq \exp\left[-\frac{(p-t)^2}{2p} n\right]$$

Provide an interpretation for this bound.

Solution.

Since $\mathbb{P}(X_k = 1) \geq p$ and the goal is to find an upper bound for $\mathbb{P}(S_n \leq tn)$ (where $t \leq p$), consider the case when $\mathbb{P}(X_k = 1) = p$ for all k. In this case, the probability $\mathbb{P}(S_n \leq tn)$ reaches the maximum. Therefore, to find the upper bound, we can assume that each X_k is an independent Bernoulli random variable with parameter p, that is, $S_n \sim \text{Binomial}(n, p)$.

For the lower tail of a binomial distribution, the standard Chernoff bound states: Let $\mu = \mathbb{E}[S_n] = np$. For $\delta \in [0, 1]$, we have

$$\mathbb{P}(S_n \leq (1-\delta)\mu) \leq \exp\left(-\frac{\delta^2 \mu}{2}\right)$$

Let $(1-\delta)\mu = tn$. Substitute $\mu = np$:

$$(1-\delta)np = tn \implies 1 - \delta = \frac{t}{p} \implies \delta = 1 - \frac{t}{p} = \frac{p-t}{p}$$

Substitute into the Chernoff bound:

$$\mathbb{P}(S_n \leq tn) \leq \exp\left(-\frac{\left(\frac{p-t}{p}\right)^2 \cdot (np)}{2}\right) = \exp\left(-\frac{(p-t)^2 \cdot np}{2p^2}\right) = \exp\left(-\frac{(p-t)^2 n}{2p}\right)$$

In the general case, $\mathbb{P}(X_k = 1) = p_k \geq p$. To prove the upper bound, we use the general form of the Chernoff bound: For any $\lambda \leq 0$, we have

$$\mathbb{P}(S_n \leq tn) \leq e^{-\lambda tn} \prod_{k=1}^{n} \mathbb{E}[e^{\lambda X_k}]$$

For each k, the moment - generating function $\mathbb{E}[e^{\lambda X_k}] = 1 - p_k + p_k e^{\lambda}$. Consider the function $h(p) = 1 - p + pe^{\lambda}$. Its derivative is

$$\frac{\partial h}{\partial p} = -1 + e^{\lambda}$$

Since $\lambda \leq 0$, $e^{\lambda} \leq 1$, so $\frac{\partial h}{\partial p} \leq 0$, that is, $h(p)$ is non - increasing in p. Therefore, when $p_k \geq p$,

$$\mathbb{E}[e^{\lambda X_k}] = h(p_k) \leq h(p) = 1 - p + pe^{\lambda}$$

Thus,

$$\prod_{k=1}^{n} \mathbb{E}[e^{\lambda X_k}] \leq (1 - p + pe^{\lambda})^n$$

So,

$$\mathbb{P}(S_n \leq tn) \leq e^{-\lambda tn}(1 - p + pe^{\lambda})^n$$

This is the same as in the case of independent and identically distributed Bernoulli(p) random variables. By choosing λ, we can obtain the same upper bound.

Problem10.

Recall that a consistent matrix norm on $\mathbb{R}^{n\times n}$ is a function $\psi : \mathbb{R}^{n\times n} \to \mathbb{R}$ that satisfies the following conditions. (a) Absolute homogeneity: $\psi(\alpha A) = |\alpha|\psi(A)$ for all $A \in \mathbb{R}^{n\times n}$ and $\alpha \in \mathbb{R}$. (b) Triangle inequality: $\psi(A+B) \leq \psi(A) + \psi(B)$ for all $A, B \in \mathbb{R}^{n\times n}$. (c) Positive definiteness: $\psi(A) \geq 0$ for all $A \in \mathbb{R}^{n\times n}$, and $\psi(A) = 0$ if and only if $A = 0$. (d) Consistency: $\psi(AB) \leq \psi(A)\psi(B)$ for all $A, B \in \mathbb{R}^{n\times n}$.

Solution.

ρ is not a consistent matrix norm on $\mathbb{R}^{n\times n}$, ρ satisfies (a), and violates (b), (c), (d).

(a)

If λ is an eigenvalue of A, then $\alpha\lambda$ is an eigenvalue of αA. Thus, $\rho(\alpha A) = \max|\alpha\lambda| = |\alpha|\max|\lambda| = |\alpha|\rho(A)$.

(b)

Let

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

Then $\rho(A) = 0$, $\rho(B) = 0$, but

$$A + B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

The eigenvalues are $1, -1$, so $\rho(A+B) = 1$.

So $1 = \rho(A+B) > \rho(A) + \rho(B) = 0$.

(c)

Let

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

The eigenvalues are $0, 0$, so $\rho(A) = 0$, but $A \neq 0$.

(d)

Let

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

Then

$$AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

The eigenvalues are $1, 0$, so $\rho(AB) = 1$.

We have $\rho(A) = 0$, $\rho(B) = 0$, so $\rho(A)\rho(B) = 0$, but $1 = \rho(AB) > \rho(A)\rho(B) = 0$.

Problem11.

For any $x \in \mathbb{R}^n$, define

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}, \quad p \in (0, \infty)$$

(a) Given $p \in (0, 1]$, prove that $\|x + y\|_p^p \leq \|x\|_p^p + \|y\|_p^p$ for all $x, y \in \mathbb{R}^n$. (b) Given $p \in (0, 1]$, prove that $\|x+y\|_p \leq 2^{\frac{1}{p}-1}(\|x\|_p + \|y\|_p)$ for all $x, y \in \mathbb{R}^n$. (c) Given $p \in (0, 1]$, prove that $\|x + y\|_p \geq \|x\|_p + \|y\|_p$ for all $x, y \in \mathbb{R}^n$ whose entries are all nonnegative.

(d) Given $x \in \mathbb{R}^n$, prove that $\|x\|_p$ is a decreasing function of $p \in (0, \infty)$.

(e) Given $x \in \mathbb{R}^n \setminus \{0\}$, prove that $\log \|x\|_p$ is a convex function of $p \in (0, \infty)$.

(f) Recall that, for a matrix $A \in \mathbb{R}^{n \times n}$, $\|A\|_p$ is defined by

$$\|A\|_p = \max_{\|x\|_p = 1} \|Ax\|_p$$

As a function of $p \in (0, +\infty)$, is $\|A\|_p$ increasing, decreasing, or neither?

Solution.

(a)

$$\|\mathbf{x} + \mathbf{y}\|_p^p \leq \|\mathbf{x}\|_p^p + \|\mathbf{y}\|_p^p \Leftrightarrow \sum_{i=1}^{n} (x_i + y_i)^p \leq \sum_{i=1}^{n} x_i^p + \sum_{i=1}^{n} y_i^p$$

we only need to prove that, for every $x_i, y_i > 0$, $(x_i + y_i)^p \leq x_i^p + y_i^p$

Consider $f(t) = t^p (t \geq 0, p \in (0, 1])$, $f''(t) = p(p-1)t^(p-2)$, easy to know $f''(t) \geq 0$, so $f(t)$ is a concave function. Therefore, we can know $x_i, y_i > 0$, $(x_i + y_i)^p \leq x_i^p + y_i^p$

(b)

From the (a), we have $\|\mathbf{x}+\mathbf{y}\|_p^p \le \|\mathbf{x}\|_p^p + \|\mathbf{y}\|_p^p$, so

$$\begin{aligned}
\|\mathbf{x}+\mathbf{y}\|_p &\le (\|\mathbf{x}\|_p^p + \|\mathbf{y}\|_p^p)^{\frac{1}{p}} \\
&= (\|\mathbf{x}\|_p + \|\mathbf{y}\|_p)[(\frac{\mathbf{x}_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p})^p + (\frac{\mathbf{y}_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p})^p]^{\frac{1}{p}} \\
&= (\|\mathbf{x}\|_p + \|\mathbf{y}\|_p)[(\frac{\mathbf{x}_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p})^p + (1-\frac{\mathbf{x}_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p})^p]^{\frac{1}{p}}
\end{aligned}$$

Let $g(t) = t^p + (1-t)^p$, where $0 < p < 1, t \in (0,1)$

$g'(t) = pt^{p-1} + (1-t)^{p-1}$. When t=0.5, $g'(t) = 0$, $g_{max} = g(0.5) = 2^{1-p}$

$$\frac{\mathbf{x}_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p} \in (0,1)$$

So

$$[(\frac{\mathbf{x}_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p})^p + (1-\frac{\mathbf{x}_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p})^p]_{max} = 2^{1-p}$$

From the above inequality, we can have:

$$\|\mathbf{x}+\mathbf{y}\|_p \le (2^{p-1})^{\frac{1}{p}}(\|\mathbf{x}\|_p + \|\mathbf{y}\|_p) = 2^{\frac{1}{p}-1}(\|\mathbf{x}\|_p + \|\mathbf{y}\|_p)$$

(c)

$$\|\mathbf{x}+\mathbf{y}\|_p \ge \|\mathbf{x}\|_p + \|\mathbf{y}\|_p \Leftrightarrow \frac{\|\mathbf{x}+\mathbf{y}\|_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p} \ge 1$$

$$\frac{\|\mathbf{x}+\mathbf{y}\|_p}{\|\mathbf{x}\|_p + \|\mathbf{y}\|_p} = [\frac{\sum(x_i+y_i)^p}{(\|\mathbf{x}\|_p + \|\mathbf{y}\|_p)^p}]^{\frac{1}{p}}$$

For function $g(u) = u^p$, $p \in (0,1)$, g is the concave function, so we can have

$$[\frac{\sum(x_i+y_i)^p}{(\|\mathbf{x}\|_p + \|\mathbf{y}\|_p)^p}]^{\frac{1}{p}} \ge [\frac{\sum(x_i^p+y_i^p)}{(\|\mathbf{x}\|_p + \|\mathbf{y}\|_p)^p}]^{\frac{1}{p}} = \frac{(\|\mathbf{x}\|_p^p + \|\mathbf{y}\|_p^p)^{\frac{1}{p}}}{(\|\mathbf{x}\|_p + \|\mathbf{y}\|_p)} \ge 1$$

(d) We need to prove that for $q > p > 0$, $\|\mathbf{x}\|_q \le \|\mathbf{x}\|_p$. If $\mathbf{x} = \mathbf{0}$, then all norms are 0, and the statement holds. Let $\mathbf{x} \ne \mathbf{0}$. Set $\|\mathbf{x}\|_p = 1$, then $\sum |x_i|^p = 1$. We need to prove $\|\mathbf{x}\|_q \le 1$, that is: $(\sum |x_i|^q)^{1/q} \le 1$

Let $y_i = |x_i|^p \ge 0$, then $\sum y_i = 1$, and:

$$\|\mathbf{x}\|_q = \left(\sum |x_i|^q\right)^{1/q} = \left(\sum (|x_i|^p)^{q/p}\right)^{1/q} = \left(\sum y_i^{q/p}\right)^{1/q}$$

Let $r = q/p > 1$, then $\|\mathbf{x}\|_q = (\sum y_i^r)^{1/q}$. Since $\sum y_i = 1$ and $y_i \geq 0$, we have $y_i \leq 1$. Since $r > 1$ and $y_i \in [0, 1]$, we have $y_i^r \leq y_i$. Thus:

$$\sum y_i^r \leq \sum y_i = 1$$

Therefore:

$$\left(\sum y_i^r\right)^{1/q} \leq (1)^{1/q} = 1$$

That is, $\|\mathbf{x}\|_q \leq 1 = \|\mathbf{x}\|_p$. Equality holds when $\mathbf{x}$ has only one non - zero component.

(e)

Failed to prove it

$$log\||\mathbf{x}\|_p = \frac{log(\sum x_i^p)}{p}$$

Let $h(p) = log(\sum x_i^p)$, $g(p) = \frac{h(p)}{p}$

$$h'(p) = \frac{\sum x_i^p log x_i}{\sum x_i^p}, \quad h''(p) = \frac{(\sum x_i^p (log x_i)^2)(\sum x_i^p) - (\sum x_i^p log x_i)^2}{(\sum x_i^p)^2} \geq 0$$

h is a convex function.

$$g'(t) = \frac{ph'(p) - h(p)}{p^2}, \quad g''(t) = \frac{p^2 h''(p) - 2ph'(p) + 2h(p)}{p^3}$$

Since $p > 0$, that is $p^3 > 0$, we only need to prove $p^2 h''(p) - 2ph'(p) + 2h(p) \geq 0$

(f) $\|A\|_p$ is neither monotonically increasing nor monotonically decreasing for $p > 0$.

Consider the matrix $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$.

When $p = 1$:

$$\|\mathbf{x}\|_1 = \left(|x_1|^1 + |x_2|^1\right)^2$$

Easy to know that $\|A\|_1 = 1$

When $p = 2$:

$$\|\mathbf{x}\|_2 = \left(|x_1|^2 + |x_2|^2\right)^{1/2} = 1 = x_1^2 + x_2^2$$

$\|A\mathbf{x}\|_2 = |x_1 + x_2|$. Let $L(x_1, x_2, \lambda) = |x_1 + x_2| - \lambda(x_1^2 + x_2^2)$.

$$\frac{\partial L}{\partial x_1} = | - 2\lambda x_1| = 0$$

$$\frac{\partial L}{\partial x_2} = |-2\lambda x_2| = 0$$
$$\frac{\partial L}{\partial \lambda} = |-2\lambda x_2| = 0$$

We can know that $|x_1 + x_2|$ achieve the maximum value, when $x_1 = x_2$, that is, $|x_1 + x_2| \leq \sqrt{2}\|\mathbf{x}\|_2 = \sqrt{2}$. Therefore, $\|A\|_2 = \sqrt{2}$.

$\|A\|_1 = 1 < \sqrt{2} = \|A\|_2$, that is, when p increases from 1 to 2, $\|A\|_p$ increases.

Consider the matrix $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

When $p = 0.5$:

$$\|A\mathbf{x}\|_{0.5} = \left(|x_1 + x_2|^{0.5} + |x_1 + x_2|^{0.5}\right)^2 = (2|x_1 + x_2|^{0.5})^2 = 4|x_1 + x_2|$$

$\|\mathbf{x}\|_{0.5} = \left(|x_1|^{0.5} + |x_2|^{0.5}\right)^2 = 1$, that is, $|x_1|^{0.5} + |x_2|^{0.5} = 1$. Let $a = |x_1|^{0.5}, b = |x_2|^{0.5}$, then $a + b = 1$, $a, b \geq 0$. Then $\|A\mathbf{x}\|_{0.5} = 4|x_1 + x_2| \leq 4(|x_1| + |x_2|) = 4(a^2 + b^2)$. Since $a + b = 1$, $a^2 + b^2 = (a+b)^2 - 2ab = 1 - 2ab$. The maximum value is achieved when $ab = 0$. Therefore, $\|A\|_{0.5} = 4$.

When $p = 1$:

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| = 1$$

$\|A\mathbf{x}\|_1 = |x_1 + x_2| + |x_1 + x_2| = 2|x_1 + x_2| \leq 2(|x_1| + |x_2|) = 2$. Therefore, $\|A\|_1 = 2$.

$\|A\|_{0.5} = 4 > 2 = \|A\|_1$. That is, when p increases from 0.5 to 1, $\|A\|_p$ decreases.

$\|A\|_p$ is neither monotonically increasing nor monotonically decreasing.

Problem12.

For any matrix $A \in \mathbb{R}^{n\times n}$ and any vector $x \in \mathbb{R}^n$, prove that $\max_{\|d\|\leq 1} \|A(x + d)\| \geq \|A\|$.Here, $\|\cdot\|$ denotes a vector norm on $\mathbb{R}^n$ and the operator norm on $\mathbb{R}^{n\times n}$ induced by this vector norm.

Solution.

Failed to prove it

Problem13.

Consider matrices $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{n\times m}$.

(a) Show that AB and BA share the same set of nonzero eigenvalues.

Optional Requirements:

- Give a proof without using determinants or matrix decomposition.
- Give a proof from a geometric point of view.
- Give a proof from an algebraic point of view.

(b) If λ is a nonzero eigenvalue of AB and BA, show that the geometric multiplicity of λ is the same with respect to AB and BA.

(c) Prove the same conclusion as above for the algebraic multiplicity.

Solution. **(a1)**

The eigenvalues of AB are all eigenvalues of BA:

$\lambda \neq 0$ is an eigenvalue of AB and the corresponding eigenvector is $\mathbf{x} \in \mathbb{C}^m$, that is $AB\mathbf{x} = \lambda\mathbf{x}$, so

$$BAB\mathbf{x} = B(\lambda\mathbf{x})$$
$$BA(B\mathbf{x}) = \lambda(B\mathbf{x})$$

It shows that $B\mathbf{x}$ is an eigenvector of BA, λ is an eigenvalue of AB. If $B\mathbf{x} = \mathbf{0}$, then the original equation becomes $AB\mathbf{x} = \lambda\mathbf{x} = \mathbf{0}$. Since $\lambda \neq 0$, we must have $\mathbf{x} = \mathbf{0}$, which contradicts the fact that an eigenvector is non - zero. Therefore, $B\mathbf{x} \neq \mathbf{0}$, that is, λ is a non-zero eigenvalue of BA.

The eigenvalues of BA are all eigenvalues of AB:

If $\lambda \neq 0$ is an eigenvalue of BA, and the corresponding eigenvector is $\mathbf{y} \in \mathbb{C}^n$, that is, $BA\mathbf{y} = \lambda\mathbf{y}$.

$$AB\mathbf{y} = A(\lambda\mathbf{y}) \implies AB(A\mathbf{y}) = \lambda(A\mathbf{y})$$

Similarly, if $A\mathbf{y} = \mathbf{0}$, then $BA\mathbf{y} = \lambda\mathbf{y} = \mathbf{0}$, which leads to $\mathbf{y} = \mathbf{0}$, a contradiction. Therefore, $A\mathbf{y} \neq \mathbf{0}$, that is, λ is a non-zero eigenvalue of AB.

(a2)

Consider linear mappings: Let $A : \mathbb{C}^n \to \mathbb{C}^m$ and $B : \mathbb{C}^m \to \mathbb{C}^n$ be linear mappings. Then $AB : \mathbb{C}^m \to \mathbb{C}^m$ and $BA : \mathbb{C}^n \to \mathbb{C}^n$.

The eigenvalues of AB are all eigenvalues of BA:

If $\lambda \neq 0$ is an eigenvalue of AB, then there $\exists$ a non-zero vector $\mathbf{x} \in \mathbb{C}^m$ such that $AB\mathbf{x} = \lambda\mathbf{x}$. $B : \mathbb{C}^m \to \mathbb{C}^n$, then $\mathbf{y} = B\mathbf{x} \in \mathbb{C}^n$. We have $\mathbf{y} \neq 0$ (as mentioned before). Then:

$$BA\mathbf{y} = BA(B\mathbf{x}) = B(AB\mathbf{x}) = B(\lambda\mathbf{x}) = \lambda B\mathbf{x} = \lambda\mathbf{y}$$

This shows that $\mathbf{y}$ is an eigenvector of BA corresponding to the eigenvalue λ.

The eigenvalues of BA are all eigenvalues of AB:

If $\lambda \neq 0$ is an eigenvalue of BA, then there exists a non-zero vector $\mathbf{z} \in \mathbb{C}^n$ such that $BA\mathbf{z} = \lambda\mathbf{z}$. $A : \mathbb{C}^n \to \mathbb{C}^m$, then $\mathbf{w} = A\mathbf{z} \in \mathbb{C}^m$. We have $\mathbf{w} \neq 0$ (as mentioned before). Then:

$$AB\mathbf{w} = AB(A\mathbf{z}) = A(BA\mathbf{z}) = A(\lambda\mathbf{z}) = \lambda A\mathbf{z} = \lambda\mathbf{w}$$

This shows that $\mathbf{w}$ is an eigenvector of AB corresponding to the eigenvalue λ.

(a3)

See (a1) for details.

(b)

Let $\lambda \neq 0$ be a common eigenvalue of AB and BA. Define the eigenspaces:

$E_\lambda(AB) = \{\mathbf{x} \in \mathbb{C}^m \mid AB\mathbf{x} = \lambda\mathbf{x}\}$, $E_\lambda(BA) = \{\mathbf{y} \in \mathbb{C}^n \mid BA\mathbf{y} = \lambda\mathbf{y}\}$.

Since $\lambda \neq 0$, we can construct linear mappings:

Define $T : E_\lambda(AB) \to E_\lambda(BA)$ as $T(\mathbf{x}) = B\mathbf{x}$, where $\mathbf{x} \in E_\lambda(AB)$.

Define $S : E_\lambda(BA) \to E_\lambda(AB)$ as $S(\mathbf{y}) = A\mathbf{y}$, where $\mathbf{y} \in E_\lambda(BA)$.

Let $\mathbf{x}_1, \mathbf{x}_2 \in E_\lambda(AB)$, and $T(\mathbf{x}_1) = T(\mathbf{x}_2)$, that is, $B\mathbf{x}_1 = B\mathbf{x}_2$.

Then $B(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$.

Since $\mathbf{x}_1 - \mathbf{x}_2 \in E_\lambda(AB)$, we have $AB(\mathbf{x}_1 - \mathbf{x}_2) = \lambda(\mathbf{x}_1 - \mathbf{x}_2)$.

But $B(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$, so:

$$AB(\mathbf{x}_1 - \mathbf{x}_2) = A(B(\mathbf{x}_1 - \mathbf{x}_2)) = A(\mathbf{0}) = \mathbf{0}$$

Thus, $\lambda(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$. Since $\lambda \neq 0$, we get $\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0}$, that is, $\mathbf{x}_1 = \mathbf{x}_2$.

Therefore, T is injective.

Similarly, we can prove that S is injective.

Since $T : E_\lambda(AB) \to E_\lambda(BA)$ is injective. Therefore

$$\dim E_\lambda(AB) \leq \dim E_\lambda(BA)$$

Similarly,

$$\dim E_\lambda(BA) \leq \dim E_\lambda(AB)$$

So

$$\dim E_\lambda(AB) = \dim E_\lambda(BA)$$

Therefore, the geometric multiplicities of λ in AB and BA are the same.

(c)

Denote the characteristic polynomial of AB as $f_{AB}(\lambda) = |(\lambda I_m - AB)|$, and the characteristic polynomial of BA as $f_{BA}(\lambda) = |(\lambda I_n - BA)|$.

Consider the polynomials:

$$\lambda^n f_{AB}(\lambda) = \lambda^n |(\lambda I_m - AB)|$$

and

$$\lambda^m f_{BA}(\lambda) = \lambda^m |(\lambda I_n - BA)|$$

There exists an invertible matrix P such that $AB = P^{-1}BAP$.

$$\begin{aligned}\lambda^n |\lambda I_m - AB| &= \lambda^n \lambda^m \left| \left(I_m - \frac{1}{\lambda} AB \right) \right| = \lambda^n \lambda^m \left| P^{-1}(I_n - \frac{1}{\lambda} BA)P \right| \\ &= \lambda^n \lambda^m \left| I_n - \frac{1}{\lambda} BA \right| = \lambda^m |\lambda I_n - BA|.\end{aligned}$$

there is an identity:

$$\lambda^n |(\lambda I_m - AB)| = \lambda^m |(\lambda I_n - BA)|$$

In the polynomial $\lambda^n f_{AB}(\lambda)$, the multiplicity of λ_0 as a root is equal to its algebraic multiplicity in $f_{AB}(\lambda)$.

Similarly, in the polynomial $\lambda^m f_{BA}(\lambda)$, the multiplicity of λ_0 as a root is equal to its algebraic multiplicity in $f_{BA}(\lambda)$.

Since the above identity shows that $\lambda^n f_{AB}(\lambda)$ and $\lambda^m f_{BA}(\lambda)$ are the same polynomial, their roots and their multiplicities are completely the same. Therefore, for any non-zero eigenvalue $\lambda_0 \neq 0$, its algebraic multiplicities in AB and BA are the same.

Problem14.

Consider a polynomial $p \in \mathbb{C}[x]$ and a matrix $A \in \mathbb{C}^{n\times n}$.

(a) For any $\lambda \in \mathbb{C}$, show that λ is an eigenvalue of A if and only if $p(\lambda)$ is an eigenvalue of $p(A)$.

Optional Requirements:

- Give a proof without using determinants or matrix decomposition.
- Give a proof from a geometric point of view.
- Give a proof from an algebraic point of view.

(b) Suppose that the eigenvalues of A are $\lambda_1, \lambda_2, \ldots, \lambda_n$, multiple eigenvalues counted with multiplicity. Show that the eigenvalues of $p(A)$ are $p(\lambda_1), p(\lambda_2), \ldots, p(\lambda_n)$, multiple eigenvalues counted with multiplicity.

Solution.

(a1)

$\Rightarrow$

If λ is an eigenvalue of A, then there $\exists$ a non-zero vector $\mathbf{v} \in \mathbb{C}^n$ such that:

$$A\mathbf{v} = \lambda\mathbf{v}$$

For a polynomial $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$, we have:

$$p(A) = a_0I + a_1A + a_2A^2 + \cdots + a_mA^m$$

Since $A\mathbf{v} = \lambda\mathbf{v}$, we can obtain that:

$$A^k\mathbf{v} = \lambda^k\mathbf{v} \quad \text{for all } k \geq 0$$

Therefore:

$$p(A)\mathbf{v} = (a_0I + \cdots + a_mA^m)\mathbf{v} = a_0\mathbf{v} + \cdots + a_m\lambda^m\mathbf{v} = p(\lambda\mathbf{v})$$

This shows that $p(\lambda)$ is an eigenvalue of $p(A)$.

$\Leftarrow$

If $p(\lambda)$ is an eigenvalue of $p(A)$, then there $\exists$ a non-zero vector $\mathbf{v} \in \mathbb{C}^n$ such that:

$$p(A)\mathbf{v} = p(\lambda)\mathbf{v}$$

Since $p(A) = a_0 I + a_1 A + a_2 A^2 + \cdots + a_m A^m$, we have:

$$(a_0 I + a_1 A + a_2 A^2 + \cdots + a_m A^m)\mathbf{v} = p(\lambda)\mathbf{v}$$

Let $q(x) = p(x) - p(\lambda)$. Then $q(\lambda) = 0$, and $q(A)\mathbf{v} = 0$. Since $q(x)$ is a polynomial and $q(\lambda) = 0$, we can write $q(x) = (x - \lambda)r(x)$, where $r(x)$ is a polynomial.

Therefore:

$$q(A) = (A - \lambda I)r(A)$$

Since $q(A)\mathbf{v} = 0$, we have:

$$(A - \lambda I)r(A)\mathbf{v} = 0$$

If $r(A)\mathbf{v} \neq 0$, then $A - \lambda I$ must have a non-zero vector $r(A)\mathbf{v}$ that makes it zero, which means λ is an eigenvalue of A. If $r(A)\mathbf{v} = 0$, we can continue to apply this process recursively, eventually, we will get that λ is an eigenvalue of A.

(a2)

Failed to prove it

(a3)

See (1) for details.

(b)

In a(1) we have already proven that if λ is an eigenvalue of A, $p(\lambda)$ is an eigenvalue of $p(A)$. Therefore, we will only consider multiplicities in this question.

Let λ_i be an eigenvalue of A, and its algebraic multiplicity is k. The eigenspace $V_{\lambda_i} = \{v \in \mathbb{C}^n \mid (A - \lambda_i I)^m v = 0 \text{ for some } m\}$ satisfies $\dim V_{\lambda_i} = k$. On V_{λ_i}, A can be expressed as $\lambda_i I + N_i$, where N_i is a nilpotent operator. Then:

$$p(A)|_{V_{\lambda_i}} = p(\lambda_i I + N_i)$$

Expand p as a Taylor series at λ_i:

$$p(\lambda_i I + N_i) = p(\lambda_i)I + p'(\lambda_i)N_i + \frac{p''(\lambda_i)}{2!}N_i^2 + \cdots$$

Since N_i is nilpotent, this series is finite. Choose a basis of V_{λ_i} such that the matrix of N_i is upper triangular matrix , then the matrix of $\lambda_i I + N_i$ is upper triangular, with all diagonal elements being λ_i. Therefore, the matrix of $p(\lambda_i I + N_i)$ is also upper triangular, with all diagonal elements being $p(\lambda_i)$. This shows that all eigenvalues of $p(A)|_{V_{\lambda_i}}$ are $p(\lambda_i)$, and the algebraic multiplicity is $\dim V_{\lambda_i} = k$.

Over the complex number field $\mathbb{C}$, the space $\mathbb{C}^n$ can be decomposed into the direct sum of generalized eigenspaces:

$$\mathbb{C}^n = \bigoplus_{\lambda} V_\lambda$$

where λ runs over the distinct eigenvalues of A, and $\dim V_\lambda$ is equal to the algebraic multiplicity of λ. Therefore, the matrix of $p(A)$ is a block - diagonal matrix with respect to this decomposition, and each block corresponds to $p(A)|_{V_\lambda}$. The eigenvalues (including algebraic multiplicities) of $p(A)$ are the union of the eigenvalues of all blocks, that is:

For each distinct eigenvalue λ of A, the algebraic multiplicity of the value $p(\lambda)$ is $\dim V_\lambda$.

Therefore, the eigenvalues of $p(A)$ are $p(\lambda_1), p(\lambda_2), \ldots, p(\lambda_n)$ (multiple eigenvalues are counted according to their algebraic multiplicities).

Problem15.

Let $n > 1$. Define $A \in \mathbb{R}^{n\times n}$ to be the matrix with entries

$$A_{ij} = \begin{cases} 1 & \text{if } i = j, \quad i, j = 1, 2, \ldots, n. \\ x & \text{if } i \neq j, \end{cases}$$

(a) Find the eigenvalues of A. Specify their multiplicities.

(b) Prove that A is positive definite if and only if $-1/(n-1) < x < 1$.

Solution.

(a)

Let $A = (1-x)I + xJ$, where I is the identity matrix and J is the all - one matrix.

The rank of the all - one matrix J is 1. Its non - zero eigenvalue is n (with algebraic multiplicity 1), and the remaining eigenvalues are 0 (with algebraic multiplicity $n-1$).

After multiplying J by x, the non - zero eigenvalue becomes $x \cdot n$, and the remaining eigenvalues are still 0. After adding $(1-x)I$, each eigenvalue increases $(1-x)$.

Therefore, the eigenvalues of matrix A are:

$(n-1)x + 1$, with corresponding algebraic multiplicity 1;

$1-x$, with corresponding algebraic multiplicity $n-1$.

(b)

The matrix is positive definite $\Leftrightarrow$ all its eigenvalues greater than 0. From (a), we can easy to know A is positive definite $\Leftrightarrow (n-1)x+1>0$ and $1-x>0 \Leftrightarrow \frac{1}{1-n} < x < 1$.

Problem16.

Suppose that $m \geq n$. Define $S = \{X \in \mathbb{C}^{m\times n} : X^H X = I_n\}$. Given $X \in \mathbb{C}^{m\times n}$, let $\operatorname{dist}(X,S)$ be the distance from X to S in Frobenius norm.

(a) Prove that $\operatorname{dist}(X,S) \leq \|I_n - X^H X\|_F$

(b) Prove that there does not exist a constant C such that $\|I_n - X^H X\|_F \leq C \operatorname{dist}(X,S)$ for all $X \in \mathbb{C}^{m\times n}$.

Solution.

(a)

Consider the singular value decomposition of matrix X: $X = U\Sigma V^H$, where U and V are unitary matrices, and Σ is a diagonal matrix whose diagonal entries are the singular values $\sigma_1, \sigma_2, \dots, \sigma_n$. Construct the matrix Y in set S as $Y = U\begin{bmatrix} I_n \\ 0 \end{bmatrix} V^H$, where $\begin{bmatrix} I_n \\ 0 \end{bmatrix}$ is an $m \times n$ matrix.

$\forall Y \in S$, $Y^H Y = I_n$.

We can get $\|X-Y\|_F = \sqrt{\sum_{i=1}^n (\sigma_i - 1)^2}$, $\|I_n - X^H X\|_F = \sqrt{\sum_{i=1}^n (1-\sigma_i^2)^2}$.

$$\sum_{i=1}^n (\sigma_i - 1)^2 \leq \sum_{i=1}^n (1-\sigma_i^2)^2$$

Therefore, $\sqrt{\sum_{i=1}^n (\sigma_i - 1)^2} \leq \sqrt{\sum_{i=1}^n (1-\sigma_i^2)^2}$, that is, $\operatorname{dist}(X,S) \leq \|I_n - X^H X\|_F$.

(b)

Construct a sequence of matrices X_k, where the first singular value is k and others are 1. At this time, $\|I_n - X_k^H X_k\|_F = |1-k^2|$, and $\operatorname{dist}(X_k, S) = |k-1|$. When $k \to \infty$, the ratio $\frac{|1-k^2|}{|k-1|} = k+1 \to \infty$, which shows that there is no such constant C.

Problem17.

Let $A \in \mathbb{C}^{m\times n}$ be a nonsingular matrix, and

$$J = \begin{pmatrix} 0 & A \\ A^H & 0 \end{pmatrix}.$$

(a) If the eigenvalues of A^HA are $\sigma_1, \ldots, \sigma_n$, multiplicity included, prove that the eigenvalues of J are $\sqrt{\sigma_1}, -\sqrt{\sigma_1}, \ldots, \sqrt{\sigma_n}, -\sqrt{\sigma_n}$, multiplicity included.

(b) Consider $n \times n$ complex matrices U_1, U_2, V_1, V_2, and Σ. Suppose that Σ is a diagonal matrix whose diagonal entries are all positive. If

$$J = \begin{pmatrix} U_1 & U_2 \\ V_1 & V_2 \end{pmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{pmatrix} \begin{pmatrix} U_1 & U_2 \\ V_1 & V_2 \end{pmatrix}^H$$

is an eigenvalue decomposition of J, prove that

$$A = 2U_1\Sigma V_1^H = -2U_2\Sigma V_2^H.$$

Solution.

(a)

Consider the eigenvalue equation of J: $J\mathbf{v} = \lambda\mathbf{v}$, where $\mathbf{v} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$, $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$. Then

$$J\mathbf{v} = \begin{pmatrix} 0 & A \\ A^H & 0 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} A\mathbf{y} \\ A^H\mathbf{x} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$$

That is,

$$A\mathbf{y} = \lambda\mathbf{x}, \quad (1)$$

$$A^H\mathbf{x} = \lambda\mathbf{y}. \quad (2)$$

If $\lambda = 0$, then $A\mathbf{y} = \mathbf{0}$ and $A^H\mathbf{x} = \mathbf{0}$. Since A is non - singular, we get $\mathbf{y} = \mathbf{0}$ and $\mathbf{x} = \mathbf{0}$, that is, $\mathbf{v} = \mathbf{0}$, a contradiction, so $\lambda \neq 0$. Solve $\mathbf{x} = \lambda^{-1}A\mathbf{y}$ from (1) and substitute into (2):

$$A^H(\lambda^{-1}A\mathbf{y}) = \lambda\mathbf{y} \implies \lambda^{-1}A^HA\mathbf{y} = \lambda\mathbf{y} \implies A^HA\mathbf{y} = \lambda^2\mathbf{y}$$

Therefore, λ^2 is an eigenvalue of A^HA, that is, $\lambda^2 = \sigma_j$. Since $\sigma_j > 0$, we have

$$\lambda = \pm\sqrt{\sigma_j}$$

(b)

Failed to prove it

According to the eigenvalue decomposition, we have:

$$\begin{aligned}
J &= \begin{pmatrix} U_1 & U_2 \\ V_1 & V_2 \end{pmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{pmatrix} \begin{pmatrix} U_1^H & V_1^H \\ U_2^H & V_2^H \end{pmatrix} \\
&= \begin{pmatrix} U_1 & U_2 \\ V_1 & V_2 \end{pmatrix} \begin{pmatrix} \Sigma U_1^H & \Sigma U_2^H \\ -\Sigma V_1^H & -\Sigma V_2^H \end{pmatrix} \\
&= \begin{pmatrix} U_1 \Sigma U_1^H + U_2(-\Sigma V_1^H) & U_1 \Sigma U_2^H + U_2(-\Sigma V_2^H) \\ V_1 \Sigma U_1^H + V_2(-\Sigma V_1^H) & V_1 \Sigma U_2^H + V_2(-\Sigma V_2^H) \end{pmatrix} \\
&= \begin{pmatrix} 0 & A \\ -A^H & 0 \end{pmatrix}
\end{aligned}$$

So

$$U_1 \Sigma U_1^H + U_2(-\Sigma V_1^H) = 0$$

$$U_1 \Sigma U_2^H + U_2(-\Sigma V_2^H) = A$$

$$V_1 \Sigma U_1^H + V_2(-\Sigma V_1^H) = -A^H$$

$$V_1 \Sigma U_2^H + V_2(-\Sigma V_2^H) = 0$$

Problem18.

(a) If $2 \le m \le n+1$, show that there exists $\{v_1, v_2, \ldots, v_m\} \subset \mathbb{R}^n$ such that $v_i^T v_j < 0$ for all distinct indices $i, j \in \{1, 2, \ldots, m\}$.

(b) If $m > n+1$, show that there does not exist $\{v_1, v_2, \ldots, v_m\} \subset \mathbb{R}^n$ such that $v_i^T v_j < 0$ for all distinct indices $i, j \in \{1, 2, \ldots, m\}$.

Solution.

(a)

$\mathbf{e}_i$ is the standard basis vector in $\mathbb{R}^n$, $\mathbf{s} = \mathbf{e}_1 + \mathbf{e}_2 + \cdots + \mathbf{e}_n$ is the all - one vector, and a is a positive number. Let $\mathbf{v}_i = \mathbf{e}_i - a\mathbf{s}$

For different i and j

$$\mathbf{v}_i^T \mathbf{v}_j = (\mathbf{e}_i - a\mathbf{s})^T (\mathbf{e}_j - a\mathbf{s}) = -2a + a^2 n$$

When $a \in (0, 2/n)$, the inner product $-2a + a^2 n < 0$.

Add the vector $\mathbf{v}_{n+1} = -b\mathbf{s}$, where $b > 0$.

$$\mathbf{v}_i^T \mathbf{v}_{n+1} = (\mathbf{e}_i - a\mathbf{s})^T (-b\mathbf{s}) = b(-1 + an)$$

When $a \in (0, 1/n)$, the inner product $b(-1 + an) < 0$.

In conclusion, When $a \in (0, 1/n)$, $\mathbf{v}_i^T \mathbf{v}_j < 0 \quad i \neq j$

(b)

Failed to prove it

Problem19.

Given $A \in \mathbb{R}^{m \times m}$ and $B \in \mathbb{R}^{n \times n}$, prove that the equation

$$AX - XB = C, \quad X \in \mathbb{R}^{m \times n}$$

has a unique solution for all $C \in \mathbb{R}^{m \times n}$ if and only if A and B do not share any eigenvalue. [When $n = 1$, B is a scalar while X and C are m-dimensional vectors; in this case, the conclusion says nothing but $(A - BI)X = C$ has a unique solution for all $C \in \mathbb{R}^m$ if and only if B is not an eigenvalue of A.]

Solution.

Define a linear operator $T : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ as $T(X) = AX - XB$. Then the original equation is equivalent to $T(X) = C$. Since $\mathbb{R}^{m \times n}$ is a finite - dimensional vector space, the following are equivalent for the linear operator T to have a unique solution for all C: T is invertible $\Leftrightarrow$ T is bijective $\Leftrightarrow$ T is injective, that is, the equation $T(X) = 0$ has only the zero solution:$AX - XB = 0 \Leftrightarrow AX = XB$ has a unique solution $X = 0$.

Therefore, we only need to prove that: $AX = XB$ has only the zero solution $\Leftrightarrow$ A and B have no common eigenvalues.

$\Rightarrow$

Suppose A and B have a common eigenvalue λ. Let $\mathbf{u}$ be an eigenvector of A belonging to λ (i.e., $A\mathbf{u} = \lambda\mathbf{u}$, $\mathbf{u} \neq \mathbf{0}$), and let $\mathbf{v}$ be a left eigenvector of B belonging to λ

(i.e., $\mathbf{v}^T B = \lambda \mathbf{v}^T$, $\mathbf{v} \neq \mathbf{0}$). Construct the matrix $X = \mathbf{u}\mathbf{v}^T \in \mathbb{R}^{m\times n}$. $X \neq \mathbf{0}$ because $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$.

$$AX = A(\mathbf{u}\mathbf{v}^T) = (A\mathbf{u})\mathbf{v}^T = (\lambda\mathbf{u})\mathbf{v}^T = \lambda\mathbf{u}\mathbf{v}^T$$

$$XB = (\mathbf{u}\mathbf{v}^T)B = \mathbf{u}(\mathbf{v}^T B) = \mathbf{u}(\lambda\mathbf{v}^T) = \lambda\mathbf{u}\mathbf{v}^T$$

Thus, $AX = \lambda\mathbf{u}\mathbf{v}^T = XB$, that is, $AX = XB$ has a non - zero solution $X = \mathbf{u}\mathbf{v}^T$.

$\Leftarrow$

Lemma:Suppose X satisfies $AX = XB$. $\forall k \geq 0$, we have $A^k X = XB^k$.

Proof: When $k = 0$, $A^0 X = IX = X$, and $XB^0 = XB^0 = XI = X$. Thus, $A^0 X = XB^0$ holds.

Suppose for some integer $k \geq 0$, $A^k X = XB^k$ holds.

$$A^{k+1}X = A \cdot A^k X = A \cdot (XB^k) = (AX)B^k = (XB)B^k = XB^{k+1}$$

Thus, by mathematical induction, we prove $A^k X = XB^k$ holds for all integers $k \geq 0$

Let $f(\lambda)$ be the characteristic polynomial of matrix A. According to the Cayley - Hamilton theorem， $f(A) = 0$.

Since A and B have no common eigenvalues, all eigenvalues of B are not roots of $f(\lambda)$. So $f(B)$ is an invertible matrix.

From the lemma, for any polynomial $p(\lambda)$, we have $p(A)X = Xp(B)$. Let $p(\lambda) = f(\lambda)$. Then by the Cayley - Hamilton theorem, $f(A) = 0$. Thus:

$$0 = f(A)X = Xf(B)$$

Since $f(B)$ is invertible, from $Xf(B) = 0$, we can obtain:

$$Xf(B) = 0 \implies X = Xf(B)f(B)^{-1} = 0 \cdot f(B)^{-1} = 0$$

Thus, when A and B have no common eigenvalues, the equation $AX - XB = 0$ has only the zero solution: $X = 0$

Thus, $AX - XB = C$ has a unique solution for all $C \in \mathbb{R}^{m\times n} \Leftrightarrow A$ and B have no common eigenvalues.

Problem20.

Let X be a random variable and f be a convex function on $\mathbb{R}$. Suppose that both X and $f(X)$ have finite expectations. Prove Jensen's inequality:

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)].$$

Solution.

Suppose the tangent line equation of the function $f(x)$ at $x = x_0$ is:$l(x) = ax + b$, where $a = f'(x_0)$ and $b = f(x_0) - ax_0$.

Since f is a convex functions, it satisfies:

$$f(x_1) \geq f(x_2) + f'(x_2)(x_1 - x_2)$$

So

$$f(x) \geq f(x_0) + f'(x_0)(x - x_0) = ax + b$$

Taking expectations on both sides simultaneously, we have:

$$\mathbb{E}[f(x)] \geq \mathbb{E}[ax + b] = a\mathbb{E}[x] + b$$

We take $x_0 = \mathbb{E}[x]$, and correspondingly $a = f'(x_0)$, $b = f(x_0) - ax_0$. Substituting into the above formula at this time, we have:

$$\mathbb{E}[f(x)] \geq a\mathbb{E}[x] + b = ax_0 + b = f(x_0) = f(\mathbb{E}[x])$$

Problem21.

For any convex function f on $[0, 1]$, prove that

$$f\left(\frac{1}{2}\right) \leq \int_0^1 f(x)\,dx \leq \frac{1}{2}\big[f(0) + f(1)\big].$$

Solution.

f is a convex function on $[0, 1]$, $\forall x, y \in [0, 1]$ and $t \in [0, 1]$, we have

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

Let $x=0, y=1, t=\frac{1}{2}$, we can get

$$f(\frac{1}{2}) \leq \frac{1}{2}f(0) + \frac{1}{2}f(1)$$

$$f(x) \geq f(0) + \frac{f(1)-f(0)}{1-0}x = f(0) + (f(1)-f(0))x.$$

So

$$\int_0^1 f(x)dx \geq \int_0^1 [f(0) + (f(1)-f(0))x]\, dx = \frac{1}{2}f(0) + \frac{1}{2}f(1)$$

So

$$f(\frac{1}{2}) \leq \int_0^1 f(x)dx$$

By the property of convex functions, for any $x \in [0,1]$, we have:

$$f(x) \leq (1-x)f(0) + xf(1)$$

So

$$\int_0^1 f(x)dx \leq \int_0^1 (1-x)f(0) + xf(1)dx$$

Calculating the right-hand side integral:

$$\int_0^1 (1-x)f(0) + xf(1)dx = \frac{1}{2}f(0) + \frac{1}{2}f(1)$$

So

$$\int_0^1 f(x)\, dx \leq \frac{1}{2}[f(0) + f(1)]$$

Therefore

$$f(\frac{1}{2}) \leq \int_0^1 f(x)dx \leq \frac{1}{2}[f(0) + f(1)]$$

Problem22.

Let X be a random variable. Suppose that f and g are two increasing functions such that $f(X)$ and $g(X)$ are both bounded. Prove

$$\mathbb{E}\big[f(X)g(X)\big] \geq \mathbb{E}\big[f(X)\big]\mathbb{E}\big[g(X)\big].$$

Solution.

Consider two independent and identically distributed random variables X and Y, both of which have the same distribution as X. Since f and g are increasing functions, for any real numbers x and y, we have:

When $x \geq y$, $f(x) \geq f(y)$ and $g(x) \geq g(y)$, so $(f(x)-f(y))(g(x)-g(y)) \geq 0$. When $x < y$, $f(x) \leq f(y)$ and $g(x) \leq g(y)$, so $(f(x)-f(y))(g(x)-g(y)) \geq 0$.

Therefore, for all x, y, we have:

$$(f(x)-f(y))(g(x)-g(y)) \geq 0$$

Take the expectation, we can obtain:

$$\mathbb{E}[(f(X)-f(Y))(g(X)-g(Y))] \geq 0$$

Expand the left - hand side, that is,

$$\mathbb{E}[f(X)g(X) - f(X)g(Y) - f(Y)g(X) + f(Y)g(Y)]$$

Since X and Y are independent and and identical distribution:

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(X)]$$

$$\mathbb{E}[g(Y)] = \mathbb{E}[g(X)]$$

$$\mathbb{E}[f(Y)g(X)] = \mathbb{E}[f(Y)]\mathbb{E}[g(X)] = \mathbb{E}[f(X)]\mathbb{E}[g(X)]$$

$$\mathbb{E}[f(Y)g(Y)] = \mathbb{E}[f(X)g(X)]$$

Substitute these in:

$$\mathbb{E}[f(X)g(X)] - \mathbb{E}[f(X)]\mathbb{E}[g(X)] - \mathbb{E}[f(X)]\mathbb{E}[g(X)] + \mathbb{E}[f(X)g(X)] \geq 0$$

That is:

$$2\mathbb{E}[f(X)g(X)] - 2\mathbb{E}[f(X)]\mathbb{E}[g(X)] \geq 0$$

Therefore:

$$\mathbb{E}[f(X)g(X)] \geq \mathbb{E}[f(X)]\mathbb{E}[g(X)]$$

Problem23.

Suppose that $\{a_k\}$ and $\{b_k\}$ are monotone real sequences with the same monotonicity. Let n be a nonnegative integer. Prove that

$$\sum_{k=0}^{n} a_k b_{n-k} \leq \frac{1}{n+1} \left(\sum_{k=0}^{n} a_k \right) \left(\sum_{k=0}^{n} b_k \right) \leq \sum_{k=0}^{n} a_k b_k.$$

Give as many proofs as possible.

Solution.

From the Chebyshev's inequality we can obtain that for two sequences $\{a_k\}$ and $\{b_k\}$ that are monotonic in the same direction, we have:

$$\frac{1}{n+1}\sum_{k=0}^{n} a_k b_k \geq \left(\frac{1}{n+1}\sum_{k=0}^{n} a_k\right)\left(\frac{1}{n+1}\sum_{k=0}^{n} b_k\right)$$

So

$$\sum_{k=0}^{n} a_k b_k \geq \frac{1}{n+1}\left(\sum_{k=0}^{n} a_k\right)\left(\sum_{k=0}^{n} b_k\right)$$

Consider $\{a_k\}$ and $\{b_{n-k}\}$. Since $\{b_k\}$ and $\{a_k\}$ are monotonic in the same direction, if $\{b_k\}$ is increasing, then $\{b_{n-k}\}$ is decreasing. At this time, $\{a_k\}$ and $\{b_{n-k}\}$ are monotonic in opposite directions. According to Chebyshev's inequality, sequences that are monotonic in opposite directions satisfy:

$$\frac{1}{n+1}\sum_{k=0}^{n} a_k b_{n-k} \leq \left(\frac{1}{n+1}\sum_{k=0}^{n} a_k\right)\left(\frac{1}{n+1}\sum_{k=0}^{n} b_{n-k}\right)$$

Since $\sum_{k=0}^{n} b_{n-k} = \sum_{k=0}^{n} b_k$, the right - hand side becomes:

$$\left(\frac{1}{n+1}\sum_{k=0}^{n} a_k\right)\left(\frac{1}{n+1}\sum_{k=0}^{n} b_k\right)$$

So

$$\sum_{k=0}^{n} a_k b_{n-k} \leq \frac{1}{n+1}\left(\sum_{k=0}^{n} a_k\right)\left(\sum_{k=0}^{n} b_k\right)$$

Problem24.

Prove that a sequence $\{x_k\} \subset \mathbb{R}$ converges if $\sum_{|x_k|>\epsilon} |x_k - x_{k+1}| < \infty$ for all $\epsilon > 0$. Is the converse proposition true?

Solution.

$\sum_{k=1}^{\infty} |x_k - x_{k+1}| < \infty$. We need to prove that $\{x_n\}$ converges. Since $\mathbb{R}$ is complete, a sequence converges if and only if it is a Cauchy sequence.

Suppose $m > n$, then:

$$|x_m - x_n| = \left|\sum_{k=n}^{m-1}(x_k - x_{k+1})\right| \leq \sum_{k=n}^{m-1} |x_k - x_{k+1}|$$

Since $\sum_{k=1}^{\infty} |x_k - x_{k+1}| < \infty$, its partial sum sequence is a Cauchy sequence. Therefore, for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that when $m > n \geq N$,

$$\sum_{k=n}^{m-1} |x_k - x_{k+1}| < \epsilon$$

Thus,

$$|x_m - x_n| \leq \sum_{k=n}^{m-1} |x_k - x_{k+1}| < \epsilon$$

Consider the Converse Proposition

The converse proposition is: If the sequence $\{x_k\}$ converges, then $\sum_{|x_k|>\epsilon} |x_k - x_{k+1}| < \infty$ holds for all $\epsilon > 0$.

Define the sequence $\{x_n\}$: $x_n = (-1)^m \frac{1}{m}$, for $2^m \leq n < 2^{m+1}$.

As $n \to \infty$, $m \to \infty$, and $|(-1)^m \frac{1}{m}| = \frac{1}{m} \to 0$. For any $\epsilon > 0$, take M such that $\frac{1}{M} < \epsilon$, and let $N = 2^M$. Then when $n \geq N$, there exists $m \geq M$ such that $2^m \leq n < 2^{m+1}$, so $|x_n| = \frac{1}{m} \leq \frac{1}{M} < \epsilon$. Thus, $x_n \to 0$.

The sequence is constant on the interval $[2^m, 2^{m+1})$. Therefore, when $k \neq 2^{m+1} - 1$, $|x_k - x_{k+1}| = 0$.

When $k = 2^{m+1} - 1$, we have $x_k = (-1)^m \frac{1}{m}$, $x_{k+1} = (-1)^{m+1} \frac{1}{m+1}$ Then,

$$|x_k - x_{k+1}| = \left|(-1)^m \frac{1}{m} - (-1)^{m+1} \frac{1}{m+1}\right| = \left|(-1)^m \left(\frac{1}{m} + \frac{1}{m+1}\right)\right| = \frac{1}{m} + \frac{1}{m+1}$$

Therefore

$$\sum_{k=1}^{\infty} |x_k - x_{k+1}| = \sum_{m=1}^{\infty} |x_{2^{m+1}-1} - x_{2^{m+1}}| = \sum_{m=1}^{\infty} \left(\frac{1}{m} + \frac{1}{m+1}\right)$$

Since $\sum_{m=1}^{\infty} \frac{1}{m} = \infty$ and $\sum_{m=1}^{\infty} \frac{1}{m+1} = \sum_{m=2}^{\infty} \frac{1}{m} = \infty$, we have:

$$\sum_{m=1}^{\infty} \left(\frac{1}{m} + \frac{1}{m+1}\right) \geq \sum_{m=1}^{\infty} \frac{1}{m} = \infty$$

Thus, $\sum_{k=1}^{\infty} |x_k - x_{k+1}| = \infty$.

In conclusion, the converse proposition does not hold.

Problem25.

Let $\{a_k\}$ and $\{b_k\}$ be nonnegative real sequences. For each index $k \geq 0$, one of the following two conditions holds:

(a) $a_k \leq b_k$ and $a_{k+1} = 2a_k$;

(b) $a_{k+1} = a_k/2$.

Prove that

$$\sum_{k=0}^{\infty} a_k \leq 2a_0 + 4\sum_{k=0}^{\infty} b_k.$$

Solution.

Failed to prove it

Problem26.

Suppose that $X \subset \mathbb{R}^n$ is a compact set, and $T : X \to X$ is a continuous operator satisfying

$$\|T(x) - T(y)\| < \|x - y\| \quad \text{for all distinct } x, y \in X.$$

(a) Show that T has a unique fixed point.

(b) For any $x_0 \in X$, show that the fixed point iteration

$$x_{k+1} = T(x_k)$$

converges to the fixed point.

Solution.

(a)

Consider the function $f(\mathbf{x}) = \|T(\mathbf{x}) - \mathbf{x}\|$. Since T is continuous, $f(\mathbf{x})$ is continuous on the compact set X. f attains its minimum value on X. Let the minimum value be attained at $\mathbf{x}^* \in X$, i.e., $f(\mathbf{x}^*) = d \geq 0$.

If $d = 0$, then $\mathbf{x}^*$ is a fixed point.

If $d > 0$, then $\mathbf{x}^* \neq T(\mathbf{x}^*)$. Consider $T(\mathbf{x}^*) \in X$. According to the problem's condition, $\|T(T(\mathbf{x}^*)) - T(\mathbf{x}^*)\| < \|T(\mathbf{x}^*) - \mathbf{x}^*\| = d$, i.e., $f(T(\mathbf{x}^*)) < d$, which contradicts the fact that d is the minimum value. Therefore, d must be 0, meaning a fixed point exists.

Suppose there exist two distinct fixed points $\mathbf{x}^*$ and $\mathbf{y}^*$, i.e., $T(\mathbf{x}^*) = \mathbf{x}^*$ and $T(\mathbf{y}^*) = \mathbf{y}^*$. According to the problem's condition, when $\mathbf{x} \neq \mathbf{y}$, $\|T(\mathbf{x}) - T(\mathbf{y})\| < \|\mathbf{x} - \mathbf{y}\|$. But $\|T(\mathbf{x}^*) - T(\mathbf{y}^*)\| = \|\mathbf{x}^* - \mathbf{y}^*\|$, which contradicts the condition. Therefore, the fixed point must be unique.

In conclusion, T has exactly one fixed point on X.

(b) Since T satisfies $\|T(x) - T(y)\| < \|x - y\|$, we have:

$$\|x_{k+1} - x_k\| = \|T(x_k) - T(x_{k-1})\| < \|x_k - x_{k-1}\|$$

This shows that the sequence $\{\|x_{k+1} - x_k\|\}$ is a decreasing sequence of positive numbers and converges to some limit $a \geq 0$.

Suppose $a > 0$. Then for any $\epsilon > 0$, there exists N_1 such that when $k > N_1$, $\|x_{k+1} - x_k\| < a + \epsilon$. Since $\{\|x_{k+1} - x_k\|\}$ is decreasing, a must be 0, that is:

$$\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0$$

Since $\|x_{k+1} - x_k\|$ is decreasing and converges to 0, according to the Monotone Convergence Theorem, the series $\sum_{k=0}^{\infty} \|x_{k+1} - x_k\|$ converges.

For any $\epsilon > 0$, there exists N_2 such that when $n > N_2$

$$\sum_{k=n}^{\infty} \|x_{k+1} - x_k\| < \epsilon$$

For any $m > n > N_2$

$$\|x_m - x_n\| \leq \sum_{k=n}^{m-1} \|x_{k+1} - x_k\| < \sum_{k=n}^{\infty} \|x_{k+1} - x_k\| < \epsilon$$

This shows that the sequence $\{x_k\}$ is a Cauchy sequence.

Since X is compact, the Cauchy sequence $\{x_k\}$ must converge to some point x^* in X, that is, $\lim_{k\to\infty} x_k = x^*$.

Since T is continuous, we have:

$$T(x^*) = T\left(\lim_{k\to\infty} x_k\right) = \lim_{k\to\infty} T(x_k) = \lim_{k\to\infty} x_{k+1} = x^*$$

From the result of problem (a), the fixed point of T on X is unique. Therefore, no matter how the initial point x_0 is chosen, the iterative sequence will converge to this unique fixed point x^*.

Problem27.

Let $f : [0, 1] \to [0, 1]$ be a continuous function. Consider the fixed point iteration $x_{k+1} = f(x_k)$ with a certain $x_0 \in [0, 1]$. If $x_k - x_{k+1} \to 0$, is it guaranteed that $\{x_k\}$ converges?

Solution.

$f : [0, 1] \to [0, 1]$ is a continuous function. The sequence x_k is bounded, and by the Bolzano-Weierstrass theorem, we can know that there is a convergent subsequence x_{k_j}.

Let $x_{k_j} \to a \in [0, 1]$. Since f is a continuous function, $x_{k_j+1} = f(x_{k_j}) \to a$. Apply $x_k - x_{k+1} \to 0$ to the subsequence, we can get $f(a) = a$, a is a fixed point of f

Assume that both a and $b \quad (a \neq b)$ are limit points of the sequence $\{x_n\}$. a and b are fixed points of f. Let $|a - b| = d$, and $\exists N > 0$ such that when $k > N$, there is $x_k - x_{k+1} < \frac{d}{3}$

If for some $k > N$, $|x_k - a| < \frac{d}{3}$

$$|x_{k+1} - a| \leq |x_{k+1} - x_k| + |x_k - a| < \frac{2d}{3}$$

$$|x_{k+1} - b| \geq |a - b| - |x_{k+1} - a| > \frac{d}{3}$$

For $k > N$, we have$|x_k - a| < d = |a - b|, \quad |x_k - b| > d/3$.

If the sequence x_k converges to b, then there exists a sufficiently large k such that $|x_k - b| < d/3$, which contradicts $|x_k - b| > d/3$. Therefore, x_k have only one limit point. Sequence x_k convergence

Problem28.

Suppose that f is a twice differentiable function on $[0, 1]$ satisfying

$$f'(0) = 0 = f'(1).$$

Show that there exists a number $\xi \in (0, 1)$ such that

$$|f''(\xi)| = 4|f(0) - f(1)|.$$

Solution.

Since f is a twice differentiable function, ee can perform a Taylor expansion to the second derivative term for f at $x = 0$ and $x = 1$, respectively.

Taylor unfolds at $x = 0$:

$$f(x) = f(0) + f'(0) + \frac{f''(\eta_1)}{2}x^2 = f(0) + \frac{f''(\eta_1)}{2}x^2, \quad \eta_1 \in (0, x)$$

Taylor unfolds at $x = 1$:

$$f(x) = f(1) + f'(1) + \frac{f''(\eta_2)}{2}x^2 = f(1) + \frac{f''(\eta_2)}{2}x^2, \quad \eta_2 \in (x, 1)$$

Substitute $x = \frac{1}{2}$

$$f(\frac{1}{2}) = f(0) + \frac{f''(\eta_1)}{8}$$
$$f(\frac{1}{2}) = f(1) + \frac{f''(\eta_2)}{8}$$

We can have

$$|f''(\eta_2) - f''(\eta_1)| = 8|f(0) - f(1)|$$

Substitute $x = 1$

$$f(1) = f(0) + \frac{f''(\eta_3)}{2}$$

We can have

$$|f(1) - f(0)| = |\frac{f''(\eta_3)}{2}|$$

Since Darboux's theorem, f'' on $[\eta_1, \eta_2]$ can take all the values between $f''(\eta_1)$ and $f''(\eta_2)$.

If $4|f(0) - f(1)|$ is between $f''(\eta_1)$ and $f''(\eta_2)$, $\exists \xi \in (\eta_1, \eta_2)$, such that $f''(\xi) = 4|f(0) - f(1)|$.

If $-4|f(0) - f(1)|$ is between $f''(\eta_1)$ and $f''(\eta_2)$, we can come to the same conclusion.

If $f''(\eta_1) \geq 4|f(0) - f(1)|$ and $f''(\eta_2) \geq 4|f(0) - f(1)|$, $f''(\eta_3) = 2|f(0) - f(1)| < 4|f(0) - f(1)|$, so $\exists \xi \in (\eta_3, \eta_1)$, such that $f''(\xi) = 4|f(0) - f(1)|$.

If $f''(\eta_1) \leq -4|f(0) - f(1)|$ and $f''(\eta_2) \leq -4|f(0) - f(1)|$, we can come to the same conclusion.

Problem29.

Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function.

(a) Suppose that $\lim_{k\to\infty} f(k+x) \to 0$ for all $x \in \mathbb{R}$. Is it guaranteed that $f(x) \to 0$ when $x \to +\infty$?

(b) Suppose that $\lim_{k\to\infty} f(kx) \to 0$ for all $x > 0$. Is it guaranteed that $f(x) \to 0$ when $x \to +\infty$?

Solution.

Failed to prove it

Problem30.

Suppose that f is a continuous function over $[0,1]$ and

$$\int_0^x [f(t)]^2 dt \leq f(x) \quad \text{for all } x \in [0,1].$$

(a) Show that

$$\min_{x\in[0,1]} f(x) \leq 2.$$

(b) Is the bound in (a) tight or not?

Solution.

Failed to prove it

Problem31.

Show that

$$\min_{\|x\|_2=1} \|Ax\|_\infty \leq \frac{1}{n}\|A\|_F$$

for all matrix $A \in \mathbb{R}^{n\times n}$, or find a counterexample.

Solution.

Note: Here we can only prove $\min_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_\infty \leq \frac{\|A\|_F}{\sqrt{n}}$.

Consider the Singular Value Decomposition of A:

$$A = U\Sigma V^T$$

where U and V are orthogonal matrices, and Σ is a diagonal matrix whose diagonal entries are the singular values of A, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

Let $\mathbf{v}_n$ be the right singular vector of A corresponding to the minimum singular value σ_n. According to the properties of singular value decomposition, we have:

$$A\mathbf{v}_n = \sigma_n \mathbf{u}_n$$

where $\mathbf{u}_n$ is the left singular vector and $\|\mathbf{u}_n\|_2 = 1$.

Take $\mathbf{x} = \mathbf{v}_n$. Obviously, $\|\mathbf{x}\|_2 = 1$.

For the vector $A\mathbf{x}$, its infinity norm satisfies:

$$\|A\mathbf{x}\|_\infty \leq \|A\mathbf{x}\|_2$$

From the singular value decomposition, we know that:

$$\|A\mathbf{x}\|_2 = \sigma_n$$

The Frobenius norm of A is:

$$\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2}$$

Since $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$

$$\sigma_n \leq \sqrt{\frac{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2}{n}} = \frac{\|A\|_F}{\sqrt{n}}$$

We can know that

$$\|A\mathbf{x}\|_\infty \leq \|A\mathbf{x}\|_2 = \sigma_n$$

$$\sigma_n \leq \frac{\|A\|_F}{\sqrt{n}}$$

Therefore,

$$\min_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_\infty \leq \frac{\|A\|_F}{\sqrt{n}}$$

Problem32.

Show that there exists a set $S \subset \mathbb{R}^n$ satisfying the following conditions.

(a) $\|x\|_2 = 1$ for all $x \in S$.

(b) $|x^T y| \leq \epsilon$ for all distinct $x, y \in S$.

(c) The cardinality of S is at least $\exp(cn\epsilon^2)$ with a certain absolute constant $c > 0$ that you must specify. [An absolute constant is a number that maintains the same value wherever it appears, e.g., $1, \pi$, and $\log 2$.]

In theory, if a set of unit vectors in $\mathbb{R}^n$ are pairwise orthogonal, then the cardinality of the set cannot exceed n. Use the existence of S to explain why we cannot rely on such a theory in numerical computations.

Solution.

Failed to prove it