Problem 1  (15 pts)

Prove that the inverse image of a convex set in $\mathbb{R}^n$ under a perspective function, $f : \mathbb{R}^n \rightarrow \mathbb{R}^{n-1}$, is convex.

Answer

The definition of inverse image of a set $C$ under a mapping $f$ is

$$f^{-1}(C) = \{ x \mid x \in \text{dom}(f) \}.$$  

The perspective mapping is defined by $f(x, t) = \frac{1}{t}x$ of which domain is $\mathbb{R}^n \times \mathbb{R}_{++}$. Therefore, the corresponding inverse image of a set $C$ under perspective mapping is

$$f^{-1}(C) = \left\{ (x, t) \mid x \in \mathbb{R}^n, t \in \mathbb{R}_{++}, \frac{1}{t}x \in C \right\}.$$  

To check the convexity of the inverse image, we arbitrarily select two elements, $(x_1, t_1)$, $(x_2, t_2)$, in $f^{-1}(C)$, and then prove that their convex combination is still in the inverse image. Convex combination of these two element is

$$(x, t) = \theta(x_1, t_1) + (1 - \theta)(x_2, t_2)$$

$$= (\theta x_1 + (1 - \theta)x_2, \theta t_1 + (1 - \theta)t_2),$$
where \( \theta \in [0, 1] \). First, \( x \) and \( t \) are still in \( \mathbb{R}^n \) and \( \mathbb{R}_{++} \) respectively, because both \( \mathbb{R}^n \) and \( \mathbb{R}_{++} \) are closed under scaling and addition. Second, we have

\[
\frac{x}{t} = \frac{\theta x_1 + (1 - \theta) x_2}{\theta t_1 + (1 - \theta) t_2} = \frac{\theta x_1}{\theta t_1 + (1 - \theta) t_2} + \frac{(1 - \theta) x_2}{\theta t_1 + (1 - \theta) t_2}
\]

\[
= \frac{x_1}{\theta t_1 + (1 - \theta) t_2} t_1 + \frac{x_2}{\theta t_1 + (1 - \theta) t_2} t_2
\]

\[
= \frac{\theta}{\theta t_1 + (1 - \theta) t_2} x_1 t_1 + \frac{1 - \theta}{\theta t_1 + (1 - \theta) t_2} x_2 t_2,
\]

where \( \hat{\theta} \in [0, 1] \). Equation (1) shows that \( x/t \) can be represented as a convex combination of two elements in \( C \), and therefore \( x/t \) belongs to \( C \) due to the convexity of \( C \). We have proved that \( (x, t) \) satisfies all constraints of the inverse image of \( C \), so the inverse image is a convex set.

**Problem 2 (20 pts)**

Let \( x \in \mathbb{R}^n \).

(a) Is \( f(x) = \|x\|_2^4 \) a strictly convex function?

(b) Is \( \nabla^2 f(x) \succ 0 \quad \forall x? \)

**Answer**

(a) We have known that \( x^T x \) is a strictly convex function because its Hessian matrix is \( 2I \), where \( I \) is an identity matrix. Therefore, we have

\[
(\theta x_1 + (1 - \theta) x_2)^T (\theta x_1 + (1 - \theta) x_2) < \theta x_1^T x_1 + (1 - \theta) x_2^T x_2
\]

when \( x_1 \neq x_2 \). We can show that \( \|x\|_2^4 \) is strictly convex because

\[
\left[ (\theta x_1 + (1 - \theta) x_2)^T (\theta x_1 + (1 - \theta) x_2) \right]^2 < \left[ \theta x_1^T x_1 + (1 - \theta) x_2^T x_2 \right]^2
\]

\[
\leq \theta (x_1^T x_1)^2 + (1 - \theta) (x_2^T x_2)^2.
\]

The last inequality has only “\( \leq \)” because \( x_1^T x_1 = x_2^T x_2 \) may occur.

(b) Let \( n = 1 \).

\[ \nabla^2 f(x) = 12x^2 = 0 \text{ if } x = 0. \]

Thus, \( \nabla^2 f(x) \) is not always positive definite.
Common mistake:

- If $f$ and $g$ are strictly convex, $f \circ g$ may not be.

Problem 3  (15 pts)

A differentiable function $f$ is defined as a strongly convex function if there exists a constant $m > 0$ such that for all points $x, y$ in its domain

$$(\nabla f(x) - \nabla f(y))^T(x - y) \geq m\|x - y\|^2.$$ 

Consider the following one-variable functions and answer questions.

(a) Is $e^x$ strongly convex?

(b) Is $x^2$ strongly convex?

(c) Is $x^4$ strongly convex?

You cannot just answer Yes or No. You need to prove your results.

Answer

(a) Assume $e^x$ is strongly convex. Then there exists an $m > 0$ such that

$$(e^x - e^y)(x - y) \geq m(x - y)^2, \quad \forall x > y.$$ 

Consider $y = x - 1$. Then

$$e^x - e^{x-1} \geq m, \quad \forall x.$$ 

That is,

$$e^{x-1}(e - 1) \geq m, \quad \forall x.$$ 

Taking limit on both sides

$$0 = \lim_{x \to -\infty} e^{x-1}(e - 1) \geq \lim_{x \to -\infty} m = m.$$ 

We obtain a contradiction. Therefore, $e^x$ is not strongly convex.

(b) Consider $f(x) = x^2$.

$$(f'(x) - f'(y))(x - y) = (2x - 2y)(x - y) = 2(x - y)^2.$$ 

Set $m = 2$. Then we have

$$(f'(x) - f'(y))(x - y) \geq m(x - y)^2, \quad \forall x, y.$$ 

Thus, $x^2$ is strongly convex.
(c) Consider \( f(x) = x^4 \).

\[
(f'(x) - f'(y))(x - y) = (4x^3 - 4y^3)(x - y) = 4(x - y)^2(x^2 + xy + y^2).
\]

Assume \( f(x) \) is strongly convex. There exists an \( m > 0 \) such that

\[
4(x^2 + xy + y^2)(x - y)^2 \geq m(x - y)^2, \quad \forall x > y.
\]

By \( x > y \), we have

\[
4(x^2 + xy + y^2) \geq m, \quad \forall x > y.
\]

If we set \( x \to 0, y \to 0 \),

\[
0 = 4(x^2 + xy + y^2) \geq m.
\]

\( x^4 \) is not strongly convex because there is no \( m > 0 \) satisfying the inequality condition.

**Problem 4  (20 pts)**

Let

\[
B_1 = \{(x, y)|x^2 + y^2 \leq 1\}
\]
\[
B_2 = \{(x, y)|(x - 2)^2 + y^2 \leq 1\}
\]

(a) Show that \( B_1 \) and \( B_2 \) are convex subsets of \( \mathbb{R}^2 \).

(b) Find a hyperplane properly separating \( B_1 \) and \( B_2 \).

(c) Can you separate \( B_1 \) and \( B_2 \) strictly?

(d) Put \( B'_1 = B_1 \setminus \{(1,0)\} \) and \( B'_2 = B_2 \setminus \{(1,0)\} \). Show that \( B'_1 \) and \( B'_2 \) are convex subsets. Can you separate \( B'_1 \) and \( B'_2 \) strictly? What about \( B'_1 \) and \( B'_2' \)?

You need to rigorously prove your answers.

**Answer**

(a) Let

\[
B_a = \{(x, y)|(x - a)^2 + y^2 \leq 1\},
\]
where \( a \) is a constant. Assume \((x_1, y_1)\) and \((x_2, y_2)\) belong to \( B_a \). We know

\[
f(x) = x^2
\]

is convex because \( f''(x) = 2 > 0 \). Then

\[
f(x) = (x - a)^2
\]

is also convex by the composition of affine function. With \( 0 \leq \theta \leq 1 \), we have

\[
(\theta x_1 + (1-\theta)x_2 - a)^2 + (\theta y_1 + (1-\theta)y_2)^2
\]

\[
= (\theta(x_1 - a) + (1-\theta)(x_2 - a))^2 + (\theta y_1 + (1-\theta)y_2)^2
\]

\[
\leq \theta(x_1 - a)^2 + (1-\theta)(x_2 - a)^2 + \theta y_1^2 + (1-\theta)y_2^2
\]

\[
\leq \theta + (1-\theta) = 1.
\]

Therefore, \( B_a \) is a convex set, and \( B_1 \) and \( B_2 \) are also convex because \( B_1 \) and \( B_2 \) are two special cases of \( B_a \) by setting \( a = 0 \) and \( 2 \), respectively.

(b) By the following figure, it is clear that \( x = 1 \) is a hyperplane separating \( B_1 \) and \( B_2 \).

(c) \((1, 0)\) belongs to both \( B_1 \) and \( B_2 \), so we cannot separate \( B_1 \) and \( B_2 \).

(d) We have proved that \( B_1 \) and \( B_2 \) are convex, so we only need to prove that \((1, 0)\) is not in the line segment of any two points in \( B'_1 \) and \( B'_2 \). Assume \((x_1, y_1)\) and \((x_2, y_2)\) belong to \( B'_1 \) and \( x_1 \leq x_2 < 1 \). If there exists \( 0 \leq \theta \leq 1 \) such that \( \theta x_1 + (1-\theta)x_2 = 1 \), we have

\[
x_1 \leq \theta x_1 + (1-\theta)x_2 = 1 \leq x_2,
\]

which is a contradiction because \( x_1 \leq x_2 < 1 \). The proof for \( B'_2 \) is similar.

We cannot separate \( B'_1 \) and \( B_2 \) strictly. Assume there exists a hyperplane separating \( B'_1 \) and \( B_2 \) strictly. That is,

\[
ax + by > C \quad \text{if} \ (x, y) \in B'_1 \quad \text{and} \quad ax + by < C \quad \text{if} \ (x, y) \in B_2.
\]
We know that \((1, 0) \in B_2\), so \(a < C\). As \(\{(x, y) | ax + by < C\}\) is open, there exists an \(\epsilon\) such that

\[
ax + by < C, \quad \forall (x, y) \in \{(x, y) | (x - 1)^2 + y^2 < \epsilon\}.
\]

Let \(0 < \epsilon' < \epsilon\). Then we have

\[
a(1 - \epsilon') + b(0) < C,
\]

which is a contradiction to \(a(1 - \epsilon') + b(0) > C\) due to \((1 - \epsilon', 0) \in B'_1\).

\(B'_1\) and \(B'_2\) can be strictly separable by \(x = 1\) because

\[
x < 1 \text{ if } (x, y) \in B'_1, \quad \text{and} \quad x > 1 \text{ if } (x, y) \in B'_2.
\]

**Problem 5 (15 pts)**

In our lecture, we proved that log-sum-exp \(f(x) = \log \sum_{j=1}^{n} e^{x_j}\) is convex. We are interested in the following maximum entropy problem in machine learning: given training instances \((x_i, y_i), i = 1, \ldots, l\), with \(y_i \in \{1, \ldots, k\}\) to indicate \(x_i\)'s class. Maxent solves

\[
\min_{w_1, \ldots, w_k} \sum_{i=1}^{l} - \log P(y_i | x_i), \tag{3}
\]

where

\[
P(y | x) = \frac{e^{w_{y_i}^T x}}{e^{w_1^T x} + \ldots + e^{w_k^T x}}.
\]

Use “operations that preserve convexity” and the property that log-sum-exp is convex to rigorously prove that (3) is convex in \(w_1, \ldots, w_k\).

**Answer**

\[
- \log P(y_i | x_i) = - \log \frac{e^{w_{y_i}^T x_i}}{e^{w_1^T x_i} + \ldots + e^{w_k^T x_i}} = -w_{y_i}^T x_i + \log \sum_{j=1}^{k} e^{w_j^T x_i}.
\]

Because the sum of convex functions is convex, it is sufficient to prove that

\[
\log \sum_{j=1}^{k} e^{w_j^T x}
\]

is convex in \(w_1, \ldots, w_k\). Consider

\[
f(z) = \log \sum_{j=1}^{k} e^{z_j} \quad \text{from } \mathbb{R}^n \to \mathbb{R}^1
\]
and
\[
g(w_1, \ldots, w_k) = \begin{bmatrix} x^T & 0 & \cdots & 0 \\ 0 & x^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x^T \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix}
\]
from \( \mathbb{R}^{nk} \to \mathbb{R}^k \),

where \( n \) is the dimension of \( x \). Then
\[
\log \sum_{j=1}^k e^{w_j^T x} = f(g(w_1, \ldots, w_k)).
\]

Because \( f \) is convex and \( g \) is an affine function, we obtain the desired result.

**Problem 6 (15 pts)**

A function's gradient is Lipschitz continuous if
\[
\| \nabla f(x) - \nabla f(y) \| \leq L \| x - y \|, \forall x, y,
\]
where \( L \) is a constant.

(a) Give a function that is strongly convex and has Lipschitz continuous gradient.

(b) Does the property in (a) hold for any strongly convex function? If yes, prove the property. Otherwise, give a counter example.

**Answer**

(a) \( f(x) = x^2, x \in \mathbb{R} \) is strongly convex because
\[
(f'(x) - f'(y))(x - y) = 2(x - y)^2 \geq m(x - y)^2 \quad \forall m \leq 2.
\]

This function's gradient is Lipschitz continuous because
\[
f'(x) = 2x \text{ and } \| 2x - 2y \| \leq L \| x - y \|,
\]
where the Lipschitz constant \( L \) can be any positive number larger or equal to 2.

(b) Let \( x, y \in \mathbb{R} \). Consider \( f(x) = x^4 + x^2 \). We have
\[
(f'(x) - f'(y))(x - y) = 2(x - y)^2(2(x + y)^2 + 2x^2 + 2y^2 + 2) \geq m(x - y)^2, \quad \forall m \leq 4,
\]

\[7\]
and therefore $f$ is a strongly convex problem. Assume that $f$ has Lipschitz continuous gradient; that is, for all $x$ and $y$

$$
\|f'(x) - f'(y)\| = \|x - y\|\|4x^2 + 4xy + 4y^2 + 2\| \leq L\|x - y\|
$$

$$
\Rightarrow \|4x^2 + 4xy + 4y^2 + 2\| \leq L
$$

always holds. The inequality causes a contradiction because for any $L$, we can find $x$ and $y$ large enough such that

$$
\|4x^2 + 4xy + 4y^2 + 2\| > L.
$$

To sum up, $f(x) = x^4 + x^2$ is strongly convex but does not has Lipschitz continuous gradient. Another example is $f(x) = e^x + e^{-x}$. 