CS 726: Homework #1

Elaine Chiu

Please typeset your solutions.

Q 1. All ℓ_p norms are related via the following inequalities:

$$(\forall q > p \geq 1)(\forall \mathbf{x} \in \mathbb{R}^d): \ \|\mathbf{x}\|_q \leq \|\mathbf{x}\|_p \leq d^{\frac{1}{p}-\frac{1}{q}}\|\mathbf{x}\|_q.$$

Provide examples of non-zero vectors (vectors whose elements are not all zeros) for which these inequalities are tight (satisfied with equality).

Note: Obviously, the left and the right inequality cannot be both satisfied at the same time, so you need to come up with two separate vectors for which the left and the right inequalities are tight. [10pts]

Solution:

First prove that we can find an example of $\|\mathbf{x}\|q = \|\mathbf{x}\|p$ for some p, q and $\mathbf{x}$. Let's take $\mathbf{x} = (1,0,0,0,0,0)^T$. Then we have $\|\mathbf{x}\|q = (\sum_{i=1}^{6}|x_i|^q)^{\frac{1}{q}} = 1 = \sum_{i=1}^{6}|x_i|^p)^{\frac{1}{p}}$.

Secondly, we want to prove that we can have $\|\mathbf{x}\|p = d^{\frac{1}{p}-\frac{1}{q}}\|\mathbf{x}\|q$ for some p, q, and $\mathbf{x}$. Consider $p = 1$ and $q = 2$, and $\mathbf{x} = (1,1)^T$. Then we have $\|\mathbf{x}\|_1 = \sum_1^2|x_i| = 2 = 2^{1-\frac{1}{2}}(x_1^2 + x_2^2)^{\frac{1}{2}} = 2^{\frac{1}{1}-\frac{1}{2}}\|\mathbf{x}\|_2$

Q 2. Let p, q be such that $p \geq 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. Suppose that you are given a differentiable function $f : \mathbb{R}^d \rightarrow \bar{\mathbb{R}}$, a number $p \geq 1$, and a constant L_p such that:

$$(\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d): \ \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_q \leq L_p\|\mathbf{x} - \mathbf{y}\|_p,$$

where $\nabla f(\mathbf{x})$ denotes the gradient (the vector of partial derivatives) of f at $\mathbf{x}$. What is the smallest constant L_2 (as a function of p, d, and L_p) for which the following holds:

$$(\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d): \ \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq L_2\|\mathbf{x} - \mathbf{y}\|_2? \qquad [7pts]$$

How large can L_2 be in the worst case (over the possible values of p)? [3pts]

Solution:

I am going to use norm inequalities to solve this problem. We have the following observations.

- $$(\forall q > p \geq 1)(\forall \mathbf{x} \in \mathbb{R}^d): \ \|\mathbf{x}\|_q \leq \|\mathbf{x}\|_p \leq d^{\frac{1}{p}-\frac{1}{q}}\|\mathbf{x}\|_q.$$
- If $q > 2$, then $p < 2$ for $\frac{1}{p} + \frac{1}{q} = 1$.
- For $q < 2$, the upper bound for $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2$ is $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_q$
- For $q > 2$, the upper bound for $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2$ is $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq d^{\frac{1}{2}-\frac{1}{q}}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_q$
- We relate the gradient and the norm with:

  $$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_q \leq L_p\|\mathbf{x} - \mathbf{y}\|_p$$
- For $q < 2$, thus $p > 2$, $\|\mathbf{x} - \mathbf{y}\|_p \leq \|\mathbf{x} - \mathbf{y}\|_2$

- For $q > 2$, thus $p < 2$, $\|\mathbf{x} - \mathbf{y}\|_p \leq d^{\frac{1}{p} - \frac{1}{2}} \|\mathbf{x} - \mathbf{y}\|_2$

Therefore, For $p > 2$:

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_q \leq L_p \|\mathbf{x} - \mathbf{y}\|_p \leq L_p \|\mathbf{x} - \mathbf{y}\|_2$$

For $p < 2$:

$$\begin{aligned}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 &\leq d^{\frac{1}{2} - \frac{1}{q}} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_q \\ &\leq d^{\frac{1}{2} - \frac{1}{q}} L_p \|\mathbf{x} - \mathbf{y}\|_p \\ &\leq d^{\frac{1}{2} - \frac{1}{q}} L_p d^{\frac{1}{p} - \frac{1}{2}} \|\mathbf{x} - \mathbf{y}\|_2 \\ &= d^{\frac{1}{p} - \frac{1}{q}} L_p \|\mathbf{x} - \mathbf{y}\|_2 \\ &= d^{\frac{2}{p} - 1} L_p \|\mathbf{x} - \mathbf{y}\|_2\end{aligned}$$

Thus for $p > 2$, the smallest L_2 is L_p; for $p < 2$, the smallest L_2 is $d^{\frac{2}{p}-1} L_p$, and across all of the possible p, the worst case is $p = 1$, then $L_2 = d \cdot L_p$.

Q 3. Let $\mathcal{X} \subseteq \mathbb{R}^d$ and $\mathcal{Y} \subseteq \mathbb{R}^d$ be convex sets.

(i) Prove that $\mathcal{X} \cap \mathcal{Y}$ is convex. [5pts]

(ii) Provide an example that shows that $\mathcal{X} \cup \mathcal{Y}$ is not necessarily convex. [5pts]

Solution:

(i) Let Z_1 and $Z_2 \in \mathcal{X} \cap \mathcal{Y}$, we want to show that $\alpha Z_1 + (1 - \alpha) Z_2 \in \mathcal{X} \cap \mathcal{Y}$, for $\alpha \in (0, 1)$. With the given condition, since Z_1 and $Z_2 \in \mathcal{X}$, so we have: $Z_1 \in \mathcal{X}$ and $Z_2 \in \mathcal{X}$, thus by $\mathcal{X}$ being convex, we have $\alpha Z_1 + (1 - \alpha) Z_2 \in \mathcal{X}$. Similarly, since Z_1 and $Z_2 \in \mathcal{Y}$, so we have $\alpha Z_1 + (1 - \alpha) Z_2 \in \mathcal{Y}$. Finally, since $\alpha Z_1 + (1 - \alpha 1) Z_2$ are in both $\mathcal{X}$ and $\mathcal{Y}$, so we have $\alpha Z_1 + (1 - \alpha) Z_2 \in \mathcal{X} \cap \mathcal{Y}$. As a result, $\mathcal{X} \cap \mathcal{Y}$ is also convex.

(ii) Consider the $x - y$ 2D coordinate system in $\mathbb{R}^2$. Define $\mathcal{X}$ to be the points in the form of $(x, 0)$, $x \in \mathbb{R}$; and define $\mathcal{Y}$ to be the points in the form of $(0, y)$, $y \in \mathbb{R}$. Then $\mathcal{X}$ and $\mathcal{Y}$ are both convex sets. However, if we take $(x_0, 0)$ and $(0, y_0)$ for $x_0, y_0 \in \mathbb{R}$, they are in $\mathcal{X} \cup \mathcal{Y}$, but $\alpha(x_0, 0) + (1 - \alpha)(0, y_0) = (\alpha x_0, (1 - \alpha) y_0 \notin \mathcal{X} \cup \mathcal{Y}$, thus the union is not a convex set.

Q 4 (Jensen's Inequality)**.** Let $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ be a convex function. Prove that for any sequence of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathbb{R}^d$ and any sequence of non-negative scalars $\alpha_1, \alpha_2, \ldots, \alpha_k \geq 0$ such that $\sum_{i=1}^k \alpha_i = 1$ we have:

$$f\Big(\sum_{i=1}^k \alpha_i \mathbf{x}_i\Big) \leq \sum_{i=1}^k \alpha_i f(\mathbf{x}_i). \qquad [10\text{pts}]$$

Solution:

Notice that by the definition of convexity, we have, when $k = 2$,

$$f(\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2) \leq \alpha_1 f(\mathbf{x}_1) + \alpha_2 f(\mathbf{x}_2)$$

For $\alpha_1 + \alpha_2 = 1$. Now, we extend this result by induction. Assume that we have

$$\sum_{i=1}^k \alpha_i = 1$$

Such that

$$f(\sum_{i=1}^{k}\alpha_i\mathbf{x}_i) \le \sum_{i=1}^{k}\alpha_i f(\mathbf{x}_i)$$

Then we want to prove that for $\sum_i^{k+1} b_i = 1$, we have

$$f(\sum_{i}^{k+1} b_i y_i) \le \sum_{i=1}^{k+1} b_i f(y_i)$$

$$\begin{aligned}
f(\sum_{i=1}^{k+1} b_i y_i) &= f(\sum_{i=1}^{k} b_i y_i + b_{k+1}y_{k+1}) \\
&= f((1-b_{k+1})(\frac{\sum_{i=1}^{k} b_i y_i}{1-b_{k+1}}) + b_{k+1}y_{k+1}) \\
&\le (1-b_{k+1})f(\frac{\sum_{i}^{k} b_i y_i}{(1-b_{k+1})}) + b_{k+1}f(y_{k+1}) \\
&\le (1-b_{k+1})\frac{\sum_{i=1}^{k} b_i f(y_i)}{(1-b_{k+1})}) + b_{k+1}f(y_{k+1}) \\
&= \sum_{i=1}^{k+1} b_i f(y_i)
\end{aligned}$$

The first inequality is from the convexity, and the second inequality is from the given induction condition. For $k = 2$ as the starting step, we can extend Jensen's inequality to $k = 2, 3, ...\mathbb{N}$

Q 5. Let $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ be an extended real valued *convex* function.

(i) Assuming that f is lower semicontinuous, prove the following: if there exists a point $\mathbf{x}$ such that $f(\mathbf{x}) = -\infty$, then f is not real-valued anywhere – it equals either $-\infty$ or $+\infty$ everywhere.
Hint: Argue first that, under these assumptions, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, $f(\mathbf{y}) \le \underline{\lim}_{\alpha\downarrow 0} f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y})$. [10pts]

(ii) If, $\forall \mathbf{x} \in \mathbb{R}^d$, $|f(\mathbf{x})| \le M$, for some constant $M < \infty$, then f must be a constant function (i.e., taking the same value for all $\mathbf{x} \in \mathbb{R}^d$). [10pts]

Solution:

(i) We first argue that given f is lower semicontinuous, we would have $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, $f(\mathbf{y}) \le \underline{\lim}_{\alpha\downarrow 0} f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y})$. Recall that a function is lower semicontinuous if:

$$\underline{\lim}_{x\to x_0} f(x) \ge f(x_0)$$

$$\underline{\lim}_{\alpha\to 0} f(\alpha x + (1-\alpha)y) \ge f(\alpha x + (1-\alpha)y|_{\alpha=0}) = f(y)$$

Secondly, by convexity, we have that

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)y$$

Therefore, combining two inequalities gives: $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$,

$$f(y) \le \underline{\lim}_{\alpha\to 0} \alpha f(x) + (1-\alpha)y$$

Now, assume that we do not have f is not real-valued everywhere, that is, we can find a point y such that

$$-\infty < f(y) = C < \infty$$

However, this would give us a contradiction as:

$$C = f(y) \nleq \lim_{\alpha \to 0} \alpha f(x) + (1-\alpha)y = -\infty$$

Thus, we prove the claim.

(ii) We want to prove this claim by contradiction. Assume that f is not constant, we we can find $x, y \in \mathbb{R}^d$ such that $f(x) > f(y)$. Recall that convexity gives that

$$af(x_0) + (1-a)f(y_0) \geq f(ax_o + (1-a)y_0)$$

for $a \in (0,1)$. Let $x = ax_0 + (1-a)y_0$, and $y = y_0$, then $\frac{x-(1-a)y}{a} = x_0$, then we have

$$af(\frac{x-(1-a)y}{a}) + (1-a)f(y) \geq f(x)$$

by plugging in.

$$\frac{f(x) - (1-a)f(y)}{a} \leq f(\frac{x-(1-a)y}{a})$$

However, we supposed that

$$f(x) > f(y)$$

so we can rewrite the left-hand side as:

$$\frac{f(x) - f(y) + af(y)}{a}$$

Which gives:

$$f(\frac{x-(1-a)y}{a}) \geq \frac{f(x) - f(y)}{a} + f(y)$$

However, $a \in (0,1)$, thus the right-hand side can get to ∞ as $a \to 0$, which leads to $f(\frac{x-(1-a)y}{a}) = \infty$, contradiction.

Q 6. Let $f : \mathbb{R}^d \to \bar{\mathbb{R}}$. Prove that f is convex if and only if its epigraph, defined as

$$\text{epi}(f) = \{(\mathbf{x}, a) : \mathbf{x} \in \mathbb{R}^d, a \in \mathbb{R}, f(\mathbf{x}) \leq a\},$$

is convex. [15pts]

Solution:

- First, we prove that if f is convex, its epigraph is also convex. To prove that its epigraph is convex, we need to show that if:

  $Z_1, Z_2 \in \text{epi}(f)$, then $tZ_1 + (1-t)Z_2 \in \text{epi}(f)$, for $t \in (0,1)$.

  Let Z_1 be the tuple (X_1, a_1), then we have

  $$f(X_1) \leq a_1$$

  ; similarly we have

  $$f(X_2) \leq a_2$$

  since it is in $\text{epi}(f)$. Notice that:

  $$tZ_1 + (1-t)Z_2 = (tX_1 + (1-t)X_2, ta_1 + (1-t)a_2)$$

  , and by f being convex:

  $$f(tX_1 + (1-t)X_2) \leq tf(x_1) + (1-t)f(x_2) \leq ta_1 + (1-t)a_2.$$

  As a result, given $Z_1, Z_2 \in \text{epi}(f)$, we also have $tZ_1 + (1-t)Z_2 \in \text{epi}(f)$, so $\text{epi}(f)$ is a convex set.

- Secondly, we prove if $\mathrm{epi}(f)$ is a convex set, then f is convex. If $\mathrm{epi}(f)$ is convex, then if $(x, f(x) \in \mathrm{epi}(f)$, and $(y, f(y)) \in \mathrm{epi}(f)$, then we have $t(x, f(x) + (1-t)(y, f(y) = (tx + (1-t)y, tf(x) + (1-t)f(y)) \in \mathrm{epi}(f)$. This implies $f(tx + (1-t)y) \le tf(x) + (1-t)f(y))$, thus f is a convex function.

Q 7. Let $\mathbf{A}$ be a real symmetric $d \times d$ matrix with eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_d$. Prove that, $\forall \mathbf{x} \in \mathbb{R}^d$:

(i) $\mathbf{x}^T\mathbf{A}\mathbf{x} \ge \lambda_1 \|\mathbf{x}\|_2^2$; [5pts]

(ii) $\mathbf{x}^T\mathbf{A}\mathbf{x} \le \lambda_d \|\mathbf{x}\|_2^2$. [5pts]

Solution:

We can prove the above two inequalities together. Recall that if we have linearly independent d $d \times 1$ column vectors, then it forms the basis of $\mathbb{R}^d$. Also, recall that for any $d \times d$ real symmetric matrix, we would have d distinct eigenvalues $(\lambda_1, ...\lambda_d)$, and has d orthogonal eigenvectors. Therefore, the eigenvectors form a basis of $\mathbb{R}^d$. Thus, for $\mathbf{x} \in \mathbb{R}^d$, we can write it into a linear combination of the eigenvectors as $\mathbf{x} = c_1v_1 + c_2v_2 +, ...c_dv_d$, for $v_1, v_2, ...v_d$ are the eigenvectors of $\mathbf{A}$. Thus,

$$\begin{aligned}\mathbf{x}^T\mathbf{A}\mathbf{x} &= \mathbf{x}^T\sum_{i=1}^{d}\lambda_i v_i v_i^T\mathbf{x}\\ &= \sum_{i=1}^{d}\lambda_i c_i^2 v_i^T v_i v_i^T v_i\end{aligned}$$

The first equality comes from the spectral decomposition of $\mathbf{A} = \sum_{i=1}^{d}\lambda_i v_i v_i^T$, for λ_i are the eigenvalues and v_i's are the eigenvectors corresponding to each eigenvalues. The second equality comes from that the eigenvectors are orthogonal to each other, thus $v_i^T v_j = 0$, for $i \ne j$, and from we know that $\mathbf{x}$ can be written as the linear combination of the eigenvectors from $\mathbf{A}$.

Finally, notice that $\|\mathbf{x}\|_2^2 = \sum_{i=1}^{d} c_i^2 v_i^T v_i v_i^T v_i$, thus:

$$\lambda_1\|\mathbf{x}\|_2^2 = \lambda_1\sum_{i=1}^{d}c_i^2 v_i^T v_i v_i^T v_i \le \sum_{i=1}^{d}\lambda_i c_i^2 v_i^T v_i v_i^T v_i = \mathbf{x}^T\mathbf{A}\mathbf{x} \le \lambda_d\sum_{i=1}^{d}c_i^2 v_i^T v_i v_i^T v_i = \lambda_d\|\mathbf{x}\|_2^2$$

Q 8. Let $\mathbf{A}$ be a $d \times d$ matrix defined by: $A_{ii} = 2$ for $1 \le i \le d$, $A_{i,i+1} = A_{i+1,i} = -1$, for $1 \le i \le d-1$ and $A_{d,1} = A_{1,d} = -1$. That is, $\mathbf{A}$ is defined as:

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 & 0 & \dots & 0 & -1\\ -1 & 2 & -1 & 0 & \dots & 0 & 0\\ 0 & -1 & 2 & -1 & \dots & 0 & 0\\ 0 & 0 & -1 & 2 & \dots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ -1 & 0 & 0 & 0 & \dots & -1 & 2 \end{bmatrix}.$$

Is $\mathbf{A}$ positive semidefinite (PSD)? What is its smallest eigenvalue? Justify your answers. [15pts]

Solution:

First, answer if $\mathbf{A}$ is positive semidefinite. For $\mathbf{A}$ to be positive semidefinite, we require that for $\mathbf{x} \in \mathbb{R}^d$, that we have $\mathbf{x}^T\mathbf{A}\mathbf{x} \ge 0$. Define $\mathbf{x} = (x_1, x_2, ...x_d)^T$ is a $d \times 1$ column vector, and $\mathbf{A} = [\mathbf{A}_1|\mathbf{A}_2|...\mathbf{A}_d]$, that $\mathbf{A}_i$ is the ith column of $\mathbf{A}$. Then, $\mathbf{A}\mathbf{x} = x_1 \cdot \mathbf{A}_1 + x_2 \cdot \mathbf{A}_2$, and $\mathbf{x}^T = [x_1, ...x_d]$.

$$\begin{aligned}
\mathbf{x}^T\mathbf{A}\mathbf{x} &= \sum_{j=1}^{d}\sum_{i=1}^{d} x_j x_i A_{ij} \\
&= \sum_{i=1}^{d} x_i^2 A_{ii} + \sum_{i \neq j}^{d-1} x_i x_j A_{ij} \\
&= 2 \cdot (\sum_{i=1}^{d} x_i^2 - \sum_{i=1}^{d-1} x_i x_{i+1} - x_d x_1)
\end{aligned}$$

Look at $\sum_{i=1}^{d} x_i^2 - x_1 x_d - x_1 x_2 - x_2 x_3 - ...x_{d-1} x_d$, recall that Young's inequality gives that:

$$ab \leq \frac{a^2}{2} + \frac{b^2}{2}$$

Thus the above summation is nonnegative, the matrix is positive semidefinite.

As for the smallest eigenvalue, first, we know that every positive semidefinite matrix has nonnegative eigenvalues. Furthermore, we know, if a matrix has linearly dependent columns or rows, then it is singular, and it would have zero determinant. The determinant is the product of the eigenvalues. We can see that this matrix has linearly dependent rows, thus determinant is zero, thus we have at least one zero-eigenvalue. Together with the fact that all eigenvalues are nonnegative, the smallest eigenvalue is zero.