CS 726: Homework #1

Posted: 09/10/2020, due: 09/21/2020 at 5pm on Canvas

Please typeset or write your solutions neatly! If we cannot read it, we cannot grade it.

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:

It is straightforward to show that any standard basis vector $\mathbf{e}_i$ for $i \in \{1, 2, ..., d\}$ will satisfy the left inequality tightly. For the right inequality, we can also show that equality holds when $\mathbf{x} = c\mathbf{1}$ for all $c \in \mathbb{R}$ where $\mathbf{1}$ is the vector with all coordinates equal to 1.

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 $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ and a constant L_p such that:

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

What is the smallest constant L_2 (as a function of p and L_p) for which the following holds:

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

How large can L_2 be in the worst case (depending on the choice of p)? [3pts]

Solution:

If $p = 2$, we have $L_2 = L_p$.

If $p > 2$, then $1 \leq q < 2 < p$ since $\frac{1}{p} + \frac{1}{q} = 1$. By the ℓ_p-norm inequalities stated in Q1, $\forall \mathbf{x}, \mathbf{y} \in \bar{\mathbb{R}}$ we have

$$\begin{aligned} \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 &\leq \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_q \\ &\leq L_p \|\mathbf{x} - \mathbf{y}\|_p \\ &\leq L_p \|\mathbf{x} - \mathbf{y}\|_2 . \end{aligned}$$

if $p < 2$, then $1 \leq p < 2 < q$. We have

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

Note that the inequalities we have used from Q1 are tight in the sense that we cannot find better multiplicative constants in general, hence our constants L_2 in the above cases are the smallest in general. In the worst case when $p = 1$, L_2 is only bounded above by dL_p. This tells us that if our problem was smooth w.r.t. the ℓ_1 norm, but we chose the ℓ_2 norm instead, we could be paying a factor of d in the smoothness constant. We will see later that for many commonly used algorithms this translates into a slow down by a factor of d or $\sqrt{d}$, which is prohibitive for large-scale problems.

Q 3 (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:

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

Solution:

The easiest way to prove the above inequality is by induction on k. The base case when $k = 1$ is trivial since $\alpha_1 = 1$. Let us assume that the above inequality holds when $k = m$ where $m \in \mathbb{Z}^+$, i.e., for any $\mathbf{x}_1, ...\mathbf{x}_m \in \mathbb{R}^d$ and any non-negative scalars $\alpha_1, ..., \alpha_m$ such that $\sum_{i=1}^m \alpha_i = 1$, we have

$$f\Big(\sum_{i=1}^m \alpha_i \mathbf{x}_i\Big) \leq \sum_{i=1}^m \alpha_i f(\mathbf{x}_i).$$

Now for any $\mathbf{x}'_1, ...\mathbf{x}'_{m+1} \in \mathbb{R}^d$ and any non-negative scalars $\alpha'_1, ..., \alpha'_{m+1}$ such that $\sum_{i=1}^{m+1} \alpha'_i = 1$, we have

$$\begin{aligned}
f\Big(\sum_{i=1}^{m+1} \alpha'_i \mathbf{x}'_i\Big) &= f\Big((1-\alpha'_{m+1})\sum_{i=1}^m \frac{\alpha'_i}{1-\alpha'_{m+1}}\mathbf{x}'_i + \alpha'_{m+1}\mathbf{x}'_{m+1}\Big) \\
&\leq (1-\alpha'_{m+1}) f\Big(\sum_{i=1}^m \frac{\alpha'_i}{1-\alpha'_{m+1}}\mathbf{x}'_i\Big) + \alpha'_{m+1} f(\mathbf{x}'_{m+1}) \quad \text{by the definition of convexity for } f \\
&\leq (1-\alpha'_{m+1}) \sum_{i=1}^m \frac{\alpha'_i}{1-\alpha'_{m+1}} f(\mathbf{x}'_i) + \alpha'_{m+1} f(\mathbf{x}'_{m+1}) \quad \text{by the assumption of induction} \\
&= \sum_{i=1}^{m+1} \alpha'_i f(\mathbf{x}'_i)
\end{aligned}$$

Hence, by the principle of induction, we conclude that the inequality holds for all $k \in \mathbb{Z}^+$ with any $\mathbf{x}_1, ...\mathbf{x}_k \in \mathbb{R}^d$ and any non-negative scalars $\alpha_1, ..., \alpha_k$ such that $\sum_{i=1}^k \alpha_i = 1$.

Q 4. 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, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, $f(\mathbf{y}) \leq \underline{\lim}_{\alpha \downarrow 0} f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y})$. [10pts]

(ii) If, $\forall \mathbf{x} \in \mathbb{R}^d$, $|f(\mathbf{x})| \leq M$, 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) Notice that for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we have $\liminf_{\alpha \downarrow 0} f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) \geq \liminf_{\bar{\mathbf{x}} \to \mathbf{y}} f(\bar{\mathbf{x}})$ since the right-hand side of the inequality is relaxed to allow any sequence $\bar{\mathbf{x}}$ tending to $\mathbf{y}$. Therefore, we can conclude that the inequality in hint holds when f is lower semicontinuous by definition.

Given $\mathbf{x} \in \mathbb{R}^d$ such that $f(\mathbf{x}) = -\infty$, assume f.p.o.c. that there exists $\mathbf{y} \in \mathbb{R}^d$ such that $f(\mathbf{y})$ is real-valued. As f is convex, for any $\alpha \in (0, 1)$, $f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}) = -\infty$. Taking the limit $\alpha \downarrow 0$, we get that $f(\mathbf{y}) \leq \underline{\lim}_{\alpha \downarrow 0} f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) = -\infty$, which contradicts the assumption that $f(\mathbf{y})$ was real-valued.

(ii) Suppose f.p.o.c. that f is not constant. Then there exist $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ such that $f(\mathbf{x}) - f(\mathbf{y}) > 0$. Let $\beta = f(\mathbf{x}) - f(\mathbf{y})$. For any $\alpha \in (0, 1)$, there exists $\mathbf{z}$ ($\mathbf{z} = \frac{\mathbf{x} - (1-\alpha)\mathbf{y}}{\alpha} \in \mathbb{R}^d$) such that $\mathbf{x} = \alpha \mathbf{z} + (1-\alpha)\mathbf{y}$. By convexity:

$$f(\mathbf{x}) \leq \alpha f(\mathbf{z}) + (1-\alpha) f(\mathbf{y}).$$

Rearranging the last inequality:

$$f(\mathbf{z}) \geq f(\mathbf{y}) + \frac{f(\mathbf{x}) - f(\mathbf{y})}{\alpha} \geq -M + \frac{\beta}{\alpha}.$$

But for any $\alpha < \frac{\beta}{2M}$ we would then have $f(\mathbf{z}) > M$, which is a contradiction.

Q 5. 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:
($\Rightarrow$) Let us denote $\mathcal{X} = \text{epi}(f)$. For any $(\mathbf{x}, a)$ and $(\mathbf{y}, b)$ in $\mathcal{X}$ and for any $\alpha \in (0, 1)$, we want to show that $(\mathbf{z}, c) = \alpha(\mathbf{x}, a) + (1-\alpha)(\mathbf{y}, b)$ is in $\mathcal{X}$. Notice that we have $(\mathbf{z}, c) = (\alpha\mathbf{x} + (1-\alpha)\mathbf{y}, \alpha a + (1-\alpha)b))$, so we have

$$f(\mathbf{z}) = f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y}) \leq \alpha a + (1-\alpha)b = c$$

by the convexity of f. By the definition of an epigraph, $(\mathbf{z}, c) \in \mathcal{X}$ and hence $\mathcal{X}$ is convex.
($\Leftarrow$) For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\alpha \in (0, 1)$, let $\mathbf{z} = \alpha\mathbf{x} + (1-\alpha)\mathbf{y}$. Notice that

$$\alpha(\mathbf{x}, f(\mathbf{x})) + (1-\alpha)(\mathbf{y}, f(\mathbf{y})) = (\mathbf{z}, \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y})) \in \mathcal{X}$$

since $(\mathbf{x}, f(\mathbf{x}))$ and $(\mathbf{y}, f(\mathbf{y}))$ are in $\mathcal{X}$ and $\mathcal{X}$ is convex. We have $f(\mathbf{z}) \leq \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y})$ by the definition of an epigraph and hence f is convex.

Q 6. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex continuously differentiable function. Using the definition of convexity from the class and properties of directional derivatives, prove that it must be $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle .$$ [10pts]

Solution:
If f is convex, then for all $\alpha \in (0, 1)$:

$$f((1-\alpha)\mathbf{x} + \alpha\mathbf{y}) = f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) \leq (1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{y}).$$

Rearranging:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \frac{f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\alpha}.$$

It remains to take $\alpha \downarrow 0$.

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

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

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

Solution:
If $\mathbf{x}$ is a vector of all zeros, both statements hold trivially, so assume that it is not. By a simple rescaling of both sides in (i) and (ii) by $\|\mathbf{x}\|_2^2$, it suffices to prove the statements for $\|\mathbf{x}\|_2 = 1$.
Using the spectral theorem, there exist orthonormal vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_d\}$ such that $\mathbf{A} = \sum_{i=1}^d \lambda_i \mathbf{u}_i \mathbf{u}_i^T$. Hence:

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i=1}^d \lambda_i (\mathbf{x}^T\mathbf{u}_i)^2.$$

It remains to use that:

$$\lambda_1 = \Big(\min_{1\leq i\leq d} \lambda_i\Big)\Big(\sum_{i=1}^d (\mathbf{x}^T\mathbf{u}_i)^2\Big) \leq \sum_{i=1}^d \lambda_i(\mathbf{x}^T\mathbf{u}_i)^2 \leq \Big(\max_{1\leq i\leq d} \lambda_i\Big)\Big(\sum_{i=1}^d (\mathbf{x}^T\mathbf{u}_i)^2\Big) = \lambda_d.$$

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

Take $\mathbf{x}$ to be any vector from $\mathbb{R}^d$. Then:

$$\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle = \sum_{i=1}^{d-1} (x_i - x_{i+1})^2 + (x_1 - x_d)^2 \geq 0,$$

which proves that $\mathbf{A}$ is PSD. To see its smallest eigenvalue, take $\mathbf{x} = \mathbf{1}$. Then we have $\langle \mathbf{A}\mathbf{1}, \mathbf{1} \rangle = 0$ hence the smallest eigenvalue is 0.