098311 Optimization 1 Spring 2018
HW 1

Chen Tessler 305052680
Orr Krupnik 302629027

April 9, 2018

Problem 1. Prove that the induced ℓ_∞ norm of $A \in \mathbb{R}^{m\times n}$ is given by

$$||A||_\infty = \max_{i=1,\ldots,m} \sum_{j=1}^{n} |A_{i,j}|$$

Solution By definition,

$$\begin{aligned} ||A||_\infty = ||A||_{\infty,\infty} &= \max_x \{||Ax||_\infty : ||x||_\infty \leq 1\} \\ &= \max_x \left\{ \max_{i=1,\ldots,m} \left\{ \sum_{j=1}^{n} |A_{i,j} x_j| \right\} : ||x||_\infty \leq 1 \right\} \end{aligned}$$

Notice that the $\max_x$ operation is performed over a weighted sum of rows in A. Since we are limited by $||x||_\infty \leq 1$, for any row the maximal value is received when $x = \mathbf{e}$. Plugging this in, we get:

$$... = \max_{i=1,\ldots,m} \left\{ \sum_{j=1}^{n} |A_{i,j}\mathbf{e}_j| \right\} = \max_{i=1,\ldots,m} \sum_{j=1}^{n} |A_{i,j}|$$

$\blacksquare$

Problem 2. Prove that for any $x \in \mathbb{R}^n$, it holds that:

$$||x||_\infty = \lim_{p\to\infty} ||x||_p$$

Solution Let us define $x^* = \max_{j=1,\ldots,n} |x_j|$
By definition of $||\cdot||_p$, we have:

$$\lim_{p\to\infty} ||x||_p = \lim_{p\to\infty} \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$$

Additionally, we have that:

$$|x^*| = \sqrt[p]{|x^*|^p} \leq \sqrt[p]{\sum_{i=1}^{n} |x_i|^p} \leq \sqrt[p]{\sum_{i=1}^{n} |x^*|^p} = \sqrt[p]{n|x^*|^p} = \sqrt[p]{n}|x^*|$$

Therefore, since $\lim_{p\to\infty} \sqrt[p]{n}|x^*| = |x^*|$, we can use the Sandwich Theorem and attain:

$$\lim_{p\to\infty} ||x||_p = |x^*| = \max_{j=1,\ldots,n} |x_j| \equiv ||x||_\infty$$

$\blacksquare$

Problem 3.

a) Suppose that R^m and R^n are equipped with norms $||\cdot||_a$ and $||\cdot||_b$ respectively. Prove the formula:

$$||A||_{a,b} = \max_x \{||Ax||_b : ||x||_a = 1\}$$

b) Let $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times k}$, and assume $\mathbb{R}^m, \mathbb{R}^n$ and $\mathbb{R}^k$ are equipped with the norms $||\cdot||_c$, $||\cdot||_b$ and $||\cdot||_a$ respectively. Prove the following relation:

$$||AB||_{a,c} \leq ||A||_{b,c}||B||_{a,b}$$

Solution

a) We note first that for the case of $A = 0^{m\times n}$, the formula holds trivially. Therefore, for the remainder of our proof, we assume A has at least one non-zero element.
By definition,

$$||A||_{a,b} = \max_x \{||Ax||_b : ||x||_a \leq 1\} = \max_x \left\{ \sqrt[b]{\sum_{i=1}^{n} |(Ax)_i|^b} : \sqrt[a]{\sum_{i=1}^{n} |x_i|^a} \leq 1 \right\}$$

Let us assume the solution for the above optimization problem is some $\hat{x}$, which satisfies $||\hat{x}|| < 1$. In this case, we will have:

$$||A||_{a,b} = ||A\hat{x}||_b$$

However, let us define $\hat{x}^\epsilon = \hat{x} + \epsilon \cdot \mathbf{e}_k$ for some $1 \leq k \leq n$ (for which $\exists i : |A_{i,k}| > 0$) and $\epsilon > 0$, such that $||\hat{x}^\epsilon|| \leq 1$. In this case, we have:

$$||A\hat{x}||_b = \sqrt[b]{\sum_{i=1}^{n} |(A\hat{x})_i|^b} < \sqrt[b]{\sum_{i=1}^{n} |(A\hat{x}^\epsilon)_i|^b} = ||A\hat{x}^\epsilon||_b$$

The above is contrary to our assumption that $\hat{x}$ is the value of x which maximizes $\max_x\{||Ax||_b : ||x||_a \leq 1\}$. Therefore, we must select x for which $||x||_a = 1$ to solve the optimization problem, and we get

$$||A||_{a,b} = \max_x \{||Ax||_b : ||x||_a = 1\}$$

$\blacksquare$

b) We define $x_* = \max_x\{||ABx||_c : ||x||_a = 1\}$. Using the formula we have proven in sec. (a), we can write:

$$\begin{aligned}||AB||_{a,c} &= \max_x\{||ABx||_c : ||x||_a = 1\} = ||ABx_*||_c \\ &\overset{(1)}{\leq} ||A||_{b,c}||Bx_*||_b \overset{(2)}{\leq} ||A||_{b,c}||B||_{a,b}||x_*||_a \\ &= ||A||_{b,c}||B||_{a,b}\end{aligned}$$

Where $(1), (2)$ both use the inequality from lec. 1 slide 9: $||Ax||_b \leq ||A||_{a,b}||x||_a$. $\blacksquare$

Problem 4. Let $A \in \mathbb{R}^{m\times n}$. Prove that:

a) $||A||_F^2 = \sum_{i=1}^n \lambda_i(A^TA)$

b) $||A||_2 \leq ||A||_F \leq \sqrt{\min\{m,n\}}||A||_2$

c) $||A||_2^2 \leq ||A||_1||A||_\infty$

Solution

a) We use the definition of the Frobenius norm:

$$\begin{aligned}||A||_F^2 &= \sum_{i=1}^n\sum_{j=1}^m A_{i,j}^2 = \sum_{i=1}^n\sum_{j=1}^m A_{j,i}^TA_{i,j} \overset{(1)}{=} \\ &= \sum_{i=1}^n (A^TA)_{i,i} = tr(A^TA) = \sum_{i=1}^n \lambda_i(A^TA)\end{aligned}$$

Where (1) arises directly from the structure of A^TA by the rules of matrix multiplication: $(A^TA)_{i,k} = \sum_{j=1}^m A_{j,i}^TA_{k,j}$. $\blacksquare$

b) By definition, and using the equality we have proven in the previous section:

$$||A||_2 = \sqrt{\lambda_{max}(A^TA)} \leq \sqrt{\sum_{i=1}^n \lambda_i(A^TA)} = ||A||_F$$

The above holds, since A^TA is a positive semi-definite matrix ($\forall x \in \mathbb{R}^n,\ x^TA^TAx = (Ax)^TAx \geq 0$), and therefore has only non-negative eigenvalues (see below). Additionally, A^TA has a rank of $min\{m,n\}$, and consequentially, at most $min\{m,n\}$ non-zero eigenvalues. This gives us:

$$\begin{aligned}||A||_F &= \sqrt{\sum_{i=1}^n \lambda_i(A^TA)} = \sqrt{\sum_{i=1}^{min\{m,n\}} \lambda_i(A^TA)} \leq \\ &\leq \sqrt{\sum_{i=1}^{min\{m,n\}} \lambda_{max}(A^TA)} = \sqrt{\min\{m,n\}}||A||_2\end{aligned}$$

Proving the non-negativity of the eigenvalues of A^TA is as follows:
For every eigenvector x_i of the matrix A^TA, we have that $A^TAx_i = \lambda_i x_i$. Multiplying by x_i^T we have: $\lambda_i x_i^T x_i = x_i^T A^T A x_i \geq 0$ (since A^TA is positive semi-definite). Finally, $x_i^T x_i$ is a non-negative scalar for any vector x_i, so in order to maintain the inequality, we require $\forall i\text{,}\ \ \lambda_i \geq 0$. ■

c) We begin by showing that $||A||_2^2 \leq ||A^TA||_1$:
By definition:

$$\begin{aligned}||A^TA||_1 &= \max_x\{||A^TAx||_1 : ||x||_1 = 1\} \overset{(a)}{=} ||A^TAx_*||_1 \overset{(b)}{\geq} ||A^TAv||_1 \\ &\overset{(c)}{=} ||\lambda_{max}(A^TA)v||_1 = |\lambda_{max}(A^TA)| \cdot ||v||_1 \geq \lambda_{max}(A^TA)||v||_1 \\ &= \lambda_{max}(A^TA) \overset{(d)}{=} ||A||_2^2\end{aligned}$$

Where:
(a) is by defining x_* as the optimal $x \in \{x : ||x||_1 = 1\}$ which maximizes $||A^TAx||_1$.
(b) since v is some $v \in \{x : ||x||_1 = 1\}$ and (c) by selecting v such that $A^TAv = \lambda_{max}(A^TA)v$ which exists since A^TA is a symmetric matrix.
(d) follows by definition of $||\cdot||_2$ as seen in class.

Finally we now show that $||A^T||_1 = ||A||_\infty$ which will conclude our proof.

$$||A^T||_1 = \max_{j'} \sum_{i'=1}^{n} |A^T_{i',j'}| = \max_{j'} \sum_{i'=1}^{n} |A_{j',i'}| \overset{(e)}{=} \max_{i} \sum_{j=1}^{n} |A_{i,j}| = ||A||_\infty$$

where (e) is by switching indices such that $i = j'$ and $j = i'$.
Combining all parts of the proof conducted above, in addition to the proof from Problem 2(b), we get:

$$||A||_2^2 \leq ||A^TA||_1 \leq ||A^T||_1||A||_1 = ||A||_\infty||A||_1$$

■

Problem 5. Let $\{A_i\}_{i\in I} \subseteq \mathbb{R}^n$ be a collection of closed sets, where I is a given index set.

a) Show that $\bigcap_{i\in I} A_i$ is a closed set.

b) Show that if I is finite, then $\bigcup_{i\in I} A_i$ is closed.

c) Is section (b) true for infinite index set I?

Solution

a) We denote $B = \bigcap_{i\in I} A_i$. By definition of an intersection over sets, any point in B must also exist in all intersecting sets $\{A_i\}_{i\in I}$. Given a converging sequence $\{x_i\}_{i=0}^\infty \subseteq B$ which converges to x_*, by definition of the intersection $\{x_i\}_{i=0}^\infty \subseteq A_i\text{,}\ \ \forall i \in I$.

However, each set A_i is a closed set which by definition entails that $x_* \in A_i$, $\forall i \in I$ which as shown above leads to the conclusion that $x_* \in B$.

We have shown that any convergence point of a sequence $\{x_i\}_{i=0}^{\infty} \subseteq B$ such that $x_i \xrightarrow{i\to\infty} x_*$ is contained in B ($x_* \in B$) $\Rightarrow B$ is a closed set.

b) We denote $C = \bigcup_{i\in I} A_i$. By definition, a set is closed if it contains all the limits of convergent sequences of vectors in the set. Lemma: For any sequence $\{x_i\}_{i=1}^{\infty} \subseteq C$ which converges to some point x_*, there exists some sub-sequence $\{x_{i_k}\}_{k=0}^{\infty} \subseteq A_j$ for some $j \in I$, which converges to x_*.

Proof of the Lemma: assume $\forall j \in I$ there does not exist any sub-sequence $\{x_{i_k}\}_{k=0}^{\infty} \subseteq A_j$, which converges to x_*. Then, for any $j \in I$, there exists some N_j for which $\forall n > N_j, x_n \notin A_j$. Notice that for $\bar{N} = \max_j N_j$, $\{x_n\}_{n=\bar{N}}^{\infty} \nsubseteq A_j \;\; \forall j \in I$ hence for $\bar{N} = \max_j N_j$, $\{x_n\}_{n=\bar{N}}^{\infty} \nsubseteq \bigcup_{j\in I} A_j = C$. This is a contradiction, since $\{x_i\}_{i=0}^{\infty} \subseteq C$ by definition.

Since A_j is closed, this means $x_* \in A_j$. By definition of C, this also means $x_* \in C$ for any such convergent series. By definition, C is closed.

c) Note that in section (b) we assume that any convergent sequence in C has some natural N for which $\{x_i\}_{i=N}^{\infty} \subseteq A_j$. This may not hold true for an infinite I. For instance, consider the collection of sets $A_i = \left[\frac{1}{i}, 1\right]$. The convergent set of points $x_n = \frac{1}{n}$, which is in the union, converges to 0 which is not in any of the sets A_i. Therefore, $\bigcup_{i\in I} A_i$ is not closed in this case.

Problem 6.

a) Let $f(x) = \max\{f_1(x), f_2(x), ..., f_m(x)\}$ where $f_i(x)$ is a differentiable function for all $i = 1, ..., m$. Show that for a given point $x \in \mathbb{R}^n$ and a nonzero vector $d \in \mathbb{R}^n$

$$f'(x; d) = \max_{i\in I_x} f_i'(x; d)$$

where $I_x = \{i \in \{1, ..., m\} : f_i(x) = f(x)\}$.

b) For any $x \in \mathbb{R}^n$ and any nonzero vector $d \in \mathbb{R}^n$, compute the directional derivative $f'(x; d)$ of

$$f(x) = \ln(e^{x_1} + e^{x_2} + ... + e^{x_n}) + max\{||x - a||, ||x - b||\}$$

Where $a, b \in \mathbb{R}^n$

Solution

a) We begin by showing that $I_x \cap I_{x+td} \neq \emptyset$.

If indeed $I_x \cap I_{x+td} = \emptyset$, this entails that $\exists f_*(x) \in \{f_1(x), f_2(x), ..., f_m(x)\}$ such that $f_*(x) \notin I_x$ and $f_*(x + td) \in I_{x+td}$. We denote $f(x) - f_*(x) = \epsilon$, note that $f(x + td) = f_*(x + td)$ and $f_*(x) < f(x)$.

$$
\begin{aligned}
0 &= \Big|f(x+td) - f_*(x+td)\Big| \\
&\overset{(a)}{=} \Big|f(x) + \nabla f(x_*)^T td + -(f_*(x) + \nabla f_*(x_{**})^T td)\Big| \\
&= \Big|f(x) - f_*(x) + (\nabla f(x_*) - \nabla f_*(x_{**}))^T td)\Big| \\
&= \Big|\epsilon + (\nabla f(x_*) - \nabla f_*(x_{**}))^T td)\Big| \overset{t\to 0}{>} 0, \ \forall \epsilon > 0
\end{aligned}
$$

where the final inequality is true, as for any bounded derivatives there exists a $T > 0$ such that for any $0 < t < T$ the above holds. Hence $\epsilon \equiv 0 \to f(x) = f_*(x)$ which in turn means that $I_x \cap I_{x+td} \neq \emptyset$. (a) is from the Linear Approximation Theorem where $x_*, x_{**} \in [x, x+td]$.

We now continue to prove $f'(x; d) = \max_{i \in I_x} f_i'(x; d)$. By definition,

$$
\begin{aligned}
f'(x;d) &= \lim_{t\to 0^+} \frac{f(x+td) - f(x)}{t} = \lim_{t\to 0^+} \frac{\max_{i\in 1,\ldots,m} f_i(x+td) - \max_{i \in 1,\ldots,m} f_i(x)}{t} \\
&\overset{(1)}{=} \lim_{t\to 0^+} \frac{\max_{i\in I_{x+td}} f_i(x+td) - \max_{i\in I_x} f_i(x)}{t} \\
&\overset{(2)}{=} \lim_{t\to 0^+} \frac{\max_{i\in I_x} f_i(x+td) - \max_{i\in I_x} f_i(x)}{t} \\
&\overset{(3)}{=} \max_{i\in I_x} f_i'(x;d)
\end{aligned}
$$

Where (1) follows from the definition of I_x (and I_{x+td} respectively), (2) follows the fact that $I_x \cap I_{x+td} \neq \emptyset$ (see proof above) and (3) follows directly from the definition of $f'(x; d)$.

■

b) We define

$$
\begin{aligned}
\bar{f}(x) &= \max\{\ln(e^{x_1} + e^{x_2} + \ldots + e^{x_n}) + ||x-a||, \ln(e^{x_1} + e^{x_2} + \ldots + e^{x_n}) + ||x-b||\} \\
&= \max\{\bar{f}_1(x), \bar{f}_2(x)\} = f(x)
\end{aligned}
$$

Using the proof from section (a) we have that $\bar{f}'(x; d) = \max_{i\in I_x} \bar{f}_i'(x; d)$ where $I_x = \{i \in \{1, 2\} : \bar{f}_i = \bar{f}(x)\}$.

$$
\begin{aligned}
\bar{f}(x) = f(x) &= \begin{cases} \bar{f}_1(x) & ||x-a|| > ||x-b|| \\ \bar{f}_2(x) & else \end{cases} \\
&= \ln(e^{x_1} + e^{x_2} + \ldots + e^{x_n}) + \begin{cases} ||x-a|| & , \ ||x-a|| > ||x-b|| \\ ||x-b|| & , \ else \end{cases}
\end{aligned}
$$

Denote $g(x) = \ln(e^{x_1} + e^{x_2} + ... + e^{x_n})$. Since $\forall x_i,\ e^{x_i} > 0$ and $\ln(\cdot)$ is defined and continuous for any positive input we can define the derivative through the gradient

$$\begin{aligned} g'(x;d) &= \nabla g(x)^T d = \Big(\frac{\partial g}{\partial x_1}(x), \frac{\partial g}{\partial x_2}(x), ..., \frac{\partial g}{\partial x_n}(x)\Big)d \\ &= \Big(\frac{e^{x_1}}{e^{x_1} + e^{x_2} + ... + e^{x_n}}, \frac{e^{x_2}}{e^{x_1} + e^{x_2} + ... + e^{x_n}}, ..., \frac{e^{x_n}}{e^{x_1} + e^{x_2} + ... + e^{x_n}}\Big)d \end{aligned}$$

Also we denote $r(x) = ||x - c|| = \sqrt{(x-c)^T(x-c)} = \sqrt{x^T x - x^T c - c^T x + c^T c} = \sqrt{\sum_{i=1}^n (x_i^2 - x_i c_i - x_i c_i + c_i^2)}$ for some $c \in \mathbb{R}^n$

$$\begin{aligned} r'(x;d) &= \nabla r(x)^T d = \Big(\frac{\partial r}{\partial x_1}(x), \frac{\partial r}{\partial x_2}(x), ..., \frac{\partial r}{\partial x_n}(x)\Big)d \\ &= \Big(\frac{2x_1 - c_1}{2r(x)}, \frac{2x_2 - c_2}{2r(x)}, ..., \frac{2x_n - c_n}{2r(x)}\Big)d \end{aligned}$$

due to linearity of the derivative operator, i.e. $f' = (g + r)' = g' + r'$ we conclude

$$\begin{aligned} f'(x;d) = \Bigg(&\Big(\frac{e^{x_1}}{e^{x_1} + e^{x_2} + ... + e^{x_n}}, \frac{e^{x_2}}{e^{x_1} + e^{x_2} + ... + e^{x_n}}, ..., \frac{e^{x_n}}{e^{x_1} + e^{x_2} + ... + e^{x_n}}\Big) \\ &+ \Big(\frac{2x_1 - c_1}{2r(x)}, \frac{2x_2 - c_2}{2r(x)}, ..., \frac{2x_n - c_n}{2r(x)}\Big)\Bigg)d \end{aligned}$$

where $c = arg\max_{a,b}\{||x - a||, ||x - b||\}$.