אופטימיזציה 1 - 098311

גיליון בית מס' 1 - חורף תשפ"א 2021

אסף גנדלר 301727715 עידו צ'רנינסקי 312544596

31 באוקטובר 2020

Problem 1:

Prove that the induced norm of $A \in R^{mxn}$ is given by:

$$\|A\|_\infty = \max_{i=1,2,\dots,m} \left\{ \sum_{j=1}^{n} |A_{i,j}| \right\}$$

proof:

by definition:

$$\|A\|_\infty = \max_{x \in R^n} \{ \|Ax\|_\infty : \|x\|_\infty \leq 1 \}$$

let's show that if $x \in R^n$ and $\|x\|_\infty \leq 1$ then:

$$\|Ax\|_\infty \leq \max_{i=1,2,..,m} \sum_{j=1,2,\dots,n} |A_{i,j}|$$

$$\|Ax\|_\infty = \max_{i=1,2,..,m} \{ |(Ax)_i| \} = \max_{i=1,2,..,m} \{ |A_{i,:}x| \} = \max_{i=1,2,..,m} \left\{ \left| \sum_{j=1,2,\dots,n} A_{i,j}x_j \right| \right\}$$

using the triangle inequality:

$$\begin{aligned} \|Ax\|_\infty &= \max_{i=1,2,..,m} \left\{ \left| \sum_{j=1,2,\dots,n} A_{i,j}x_j \right| \right\} \leq \max_{i=1,2,..,m} \left\{ \sum_{j=1,2,\dots,n} |A_{i,j}x_j| \right\} = \max_{i=1,2,..,m} \left\{ \sum_{j=1,2,\dots,n} |A_{i,j}| \, |x_j| \right\} \\ &\leq \max_{i=1,2,..,m} \left\{ \sum_{j=1,2,\dots,n} |A_{i,j}| \max_{k=1,2,..,n} \{ |x_k| \} \right\} = \max_{i=1,2,..,m} \left\{ \sum_{j=1,2,\dots,n} |A_{i,j}| \, \|x\|_\infty \right\} \\ &= \max_{i=1,2,..,m} \left\{ \|x\|_\infty \sum_{j=1,2,\dots,n} |A_{i,j}| \right\} \leq \max_{i=1,2,..,m} \left\{ \sum_{j=1,2,\dots,n} |A_{i,j}| \right\} \end{aligned}$$

we have proved that if $x \in R^n$ and $\|x\|_\infty \leq 1$ then:

$$\|Ax\|_\infty \leq \max_{i=1,2,..,m} \left\{ \sum_{j=1,2,\dots,n} |A_{i,j}| \right\}$$

now we will prove that the upper bound is attained:

let:

$$k = \arg \max_{i=1,2,..,m} \left\{ \sum_{j=1,2,...,n} |A_{i,j}| \right\}$$

let's choose:

$$x^* = \begin{pmatrix} sign\left(A_{k,1}\right) \\ sign\left(A_{k,2}\right) \\ . \\ . \\ . \\ sign\left(A_{k,n}\right) \end{pmatrix}$$

notice that:

$$\|x^*\|_\infty = 1 \leq 1$$

and:

$$\|Ax^*\|_\infty = \max_{i=1,2,..,m} \left\{ \left| \sum_{j=1,2,...,n} A_{i,j} x_j^* \right| \right\}$$

let's find the this maximum:

$$\left| \sum_{j=1,2,...,n} A_{i,j} x_j^* \right| \leq \sum_{j=1,2,...,n} |A_{i,j}| \left|x_j^*\right| = \sum_{j=1,2,...,n} |A_{i,j}| \leq \sum_{j=1,2,...,n} |A_{k,j}|$$

in addition for $i = k$:

$$\left| \sum_{j=1,2,...,n} A_{i,j} x_j^* \right| = \left| \sum_{j=1,2,...,n} A_{k,j} x_j^* \right| = \left| \sum_{j=1,2,...,n} A_{k,j} sign\left(A_{k,j}\right) \right| = \left| \sum_{j=1,2,...,n} |A_{k,j}| \right| = \sum_{j=1,2,...,n} |A_{k,j}|$$

the upper bound is attained thus:

$$\|Ax^*\|_\infty = \max_{i=1,2,..,m} \left\{ \left| \sum_{j=1,2,...,n} A_{i,j} x_j^* \right| \right\} = \sum_{j=1,2,...,n} |A_{k,j}| = \max_{i=1,2,..,m} \left\{ \sum_{j=1,2,...,n} |A_{i,j}| \right\}$$

so we have proved that for $x \in R^n$ and $\|x\|_\infty \leq 1$:

$$\|Ax\|_\infty \leq \max_{i=1,2,..,m} \left\{ \sum_{j=1,2,...,n} |A_{i,j}| \right\}$$

and:

$$\|Ax^*\|_\infty = \max_{i=1,2,..,m} \left\{ \sum_{j=1,2,...,n} |A_{i,j}| \right\}$$

thus:

$$\boxed{\|A\|_\infty = \max_{x \in R^n} \{\|Ax\|_\infty : \|x\|_\infty \leq 1\} = \max_{i=1,2,..,m} \left\{ \sum_{j=1,2,...,n} |A_{i,j}| \right\}}$$

■

Problem 2:

Prove that for $x \in R^n$:

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

proof:

$$\|x\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p} = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

lets mark:

$$|x_j| = \max_{i=1,2,\ldots,n} \{|x_i|\} = \|x\|_\infty$$

For start let's assume $|x_j|$ is unique meaning:

$$|x_j| > |x_i|, \quad \forall i \neq j$$

then:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} = \left(\sum_{i=1}^{n} |x_j|^p \frac{|x_i|^p}{|x_j|^p} \right)^{\frac{1}{p}} = \left(|x_j|^p \sum_{i=1}^{n} \left| \frac{x_i}{x_j} \right|^p \right)^{\frac{1}{p}} =$$

$$= |x_j| \left(\sum_{i=1}^{n} \left| \frac{x_i}{x_j} \right|^p \right)^{\frac{1}{p}} = |x_j| \left(1 + \sum_{\substack{i=1 \\ i \neq j}}^{n} \left| \frac{x_i}{x_j} \right|^p \right)^{\frac{1}{p}}$$

because $|x_j| > |x_i|$ than:

$$\lim_{p \to \infty} \left| \frac{x_i}{x_j} \right|^p = 0$$

and:

$$\boxed{\lim_{p \to \infty} \|x\|_p = \lim_{p \to \infty} |x_j| \left(1 + \sum_{\substack{i=1 \\ i \neq j}}^{n} \left| \frac{x_i}{x_j} \right|^p \right)^{\frac{1}{p}} = |x_j| \cdot 1^0 = |x_j| = \|x\|_\infty}$$

If $|x_j|$ is not unique, let's say that there are k elements in the vector with the same maximum absolute value, than:

$$\|x\|_p = |x_j| \left(\sum_{i=1}^{n} \left| \frac{x_i}{x_j} \right|^p \right)^{\frac{1}{p}} = |x_j| \left(k + \sum_{\substack{i=1 \\ |x_i| \neq |x_j|}}^{n} \left| \frac{x_i}{x_j} \right|^p \right)^{\frac{1}{p}}$$

and:

$$\boxed{\lim_{p\to\infty} \|x\|_p = |x_j| \left(k + \sum_{\substack{i=1 \\ |x_i| \neq |x_j|}}^{n} \left| \frac{x_i}{x_j} \right|^p \right)^{\frac{1}{p}} = |x_j| \cdot k^0 = |x_j| = \|x\|_\infty}$$

$\blacksquare$

Problem 3:

a)

part 1:

Prove that for $A \in R^{mxn}$:

$$\|A\|_{a,b} = \max_{x \in R^n} \{\|Ax\|_a : \|x\|_b = 1\}$$

by definition:

$$\|A\|_{a,b} = \max_{x \in R^n} \{\|Ax\|_a : \|x\|_b \leq 1\}$$

let's assume by contradiction that:

$$x^* = \arg\max_{x \in R^n} \{\|Ax\|_a : \|x\|_b \leq 1\}$$

has a norm that satisfy:

$$||x^*||_b < 1$$

define:

$$y = \frac{x^*}{||x^*||_b}$$

$$||y||_b = \left|\left|\frac{x^*}{||x^*||_b}\right|\right|_b = \frac{||x^*||_b}{||x^*||_b} = 1 \leq 1$$

$$\|Ay\|_a = \left\|A\frac{x^*}{||x^*||_b}\right\|_a = \frac{1}{||x^*||_b}\|Ax^*\|_a > \|Ax^*\|_a$$

and that's a contradiction to the fact that:

$$x^* = \arg\max_{x \in R^n} \{\|Ax\|_a : \|x\|_b \leq 1\}$$

so x^* has to satisfy:

$$||x^*||_b = 1$$

and:

$$\boxed{\|A\|_{a,b} = \max_{x \in R^n} \{\|Ax\|_a : \|x\|_b \leq 1\} = \max_{x \in R^n} \{\|Ax\|_a : \|x\|_b = 1\}}$$

note:

we divided by $||x^*||_b$ assuming it's larger than zero. there is a singular case in which $A^{mxn} = 0^{mxn}$ and than any vector could be the one that attains the maximum value, including $x^* = 0_n$,in this case the statement still holds.

part 2:

Prove that for $A \in R^{mxn}$:

$$\|A\|_{a,b} = \max_{x \in R^n} \left\{ \frac{\|Ax\|_a}{\|x\|_b} : x \neq 0_n \right\}$$

let's mark:

$$x^* = \arg\max_{x \in R^n} \left\{ \frac{\|Ax\|_a}{\|x\|_b} : x \neq 0_n \right\}$$

and define:

$$y = \frac{x^*}{||x^*||_b}$$

$$||y||_b = 1$$

$$\frac{\|Ay\|_a}{\|y\|_b} = \frac{\left\| A \frac{x^*}{||x^*||_b} \right\|_a}{\left\| \frac{x^*}{||x^*||_b} \right\|_b} = \frac{\cancel{\frac{1}{||x^*||_b}} \|Ax^*\|_a}{\cancel{\frac{1}{||x^*||_b}} \|x^*\|_b} = \frac{\|Ax^*\|_a}{\|x^*\|_b}$$

we see an ambiguity, that for every vector that solves $\max_{x \in R^n} \left\{ \frac{\|Ax\|_a}{\|x\|_b} : x \neq 0_n \right\}$ we can find another vector with norm that equals to 1 that achieves the same value, thus we can limit the search domain and the two problems are equivalent.

$$\boxed{\max_{x \in R^n} \left\{ \frac{\|Ax\|_a}{\|x\|_b} : x \neq 0_n \right\} = \max_{x \in R^n} \left\{ \frac{\|Ax\|_a}{\|x\|_b} : \|x\|_b = 1 \right\} = \max_{x \in R^n} \left\{ \|Ax\|_a : \|x\|_b = 1 \right\}}$$

■

b)

prove that:

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

let $A \in \mathbb{R}^{mxn}$ and $B \in \mathbb{R}^{nxk}$

$$\|AB\|_{c,a} = \max_{x \in R^n} \left\{ \|ABx\|_a : \|x\|_c \leq 1 \right\}$$

using the inequality from the lecture, for $\|x\|_c \leq 1$:

$$\|ABx\|_a \leq \|A\|_{b,a} \|Bx\|_b \leq \|A\|_{b,a} \|B\|_{c,b} \|x\|_c \leq \|A\|_{b,a} \|B\|_{c,b}$$

hence:

$$\boxed{\|AB\|_{c,a} = \max_{x \in R^n} \{\|ABx\|_a : \|x\|_c \leq 1\} \leq \max_{x \in R^n} \left\{\|A\|_{b,a} \|B\|_{c,b} : \|x\|_c \leq 1\right\} = \|A\|_{b,a} \|B\|_{c,b}}$$

$\blacksquare$

Problem 4:

let $A \in \mathbb{R}^{mxn}$

a)

prove that:

$$\|A\|_F^2 = \sum_{i=1}^{n} \lambda_{i\{A^TA\}}$$

$$\|A\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} A_{ij}^2 = \sum_{i=1}^{n} \left(A^TA\right)_{i,i} = Tr\left(A^TA\right) = \sum_{i=1}^{n} \lambda_{i\{A^TA\}}$$

$$\boxed{\|A\|_F^2 = \sum_{i=1}^{n} \lambda_{i\{A^TA\}}}$$

$\blacksquare$

b)

prove that:

$$\frac{1}{\sqrt{n}}\|A\|_\infty \le \|A\|_2 \le \sqrt{m}\,\|A\|_\infty$$

let's first prove a similar inequality for vectors:

let $x \in \mathbb{R}^n$ then:

$$n\|x\|_\infty^2 = \sum_{i=1}^{n}\|x\|_\infty^2 \ge \sum_{i=1}^{n} x_i^2 = \|x\|_2^2$$

$$\sqrt{n}\,\|x\|_\infty \ge \|x\|_2$$

and:

$$\|x\|_\infty = \sqrt{\|x\|_\infty^2} \le \sqrt{\sum_{i=1}^{n} x_i^2} = \|x\|_2$$

hence:

$$\|x\|_\infty \le \|x\|_2 \le \sqrt{n}\,\|x\|_\infty$$

now moving to matrices:

let's start from the right side

$$\|A\|_2 = \max_{x\in R^n}\left\{\|Ax\|_2 : \|x\|_2 \le 1\right\}$$

for $\|x\|_2 \leq 1$:

$$\|Ax\|_2 \leq \sqrt{m}\,\|Ax\|_\infty \leq \sqrt{m}\,\|A\|_\infty\,\|x\|_\infty \leq \sqrt{m}\,\|A\|_\infty\,\|x\|_2 \leq \sqrt{m}\,\|A\|_\infty$$

hence:

$$\|A\|_2 = \max_{x\in R^n}\{\|Ax\|_2 : \|x\|_2 \leq 1\} \leq \max_{x\in R^n}\left\{\sqrt{m}\,\|A\|_\infty : \|x\|_2 \leq 1\right\} = \sqrt{m}\,\|A\|_\infty$$

$$\|A\|_2 \leq \sqrt{m}\,\|A\|_\infty$$

for the left, side:

$$\|A\|_\infty = \max_{x\in R^n}\{\|Ax\|_\infty : \|x\|_\infty \leq 1\}$$

for $\|x\|_\infty \leq 1$

$$\|Ax\|_\infty \leq \|Ax\|_2 \leq \|A\|_2\,\|x\|_2 \leq \|A\|_2\,\sqrt{n}\,\|x\|_\infty \leq \sqrt{n}\,\|A\|_2$$

hence:

$$\|A\|_\infty = \max_{x\in R^n}\{\|Ax\|_\infty : \|x\|_\infty \leq 1\} \leq \max_{x\in R^n}\left\{\sqrt{n}\,\|A\|_2 : \|x\|_\infty \leq 1\right\} = \sqrt{n}\,\|A\|_2$$

$$\frac{1}{\sqrt{n}}\,\|A\|_\infty \leq \|A\|_2$$

to summarize:

$$\boxed{\frac{1}{\sqrt{n}}\,\|A\|_\infty \leq \|A\|_2 \leq \sqrt{m}\,\|A\|_\infty}$$

■

Problem 5:

a)

let $x \in \bigcup_{i \in I} A_i$

$$\exists j \in I : x \in A_j$$

A_j is an open set, thus x is an interior point of A_j, and by definition

$$\exists r > 0 : B(x, r) \subseteq A_j \subseteq \bigcup_{i \in I} A_i$$

so:

$$\exists r > 0 : B(x, r) \subseteq \bigcup_{i \in I} A_i$$

therefore x is an interior point of $\bigcup_{i \in I} A_i$.

every point in $\bigcup_{i \in I} A_i$ is an interior point, hence $\bigcup_{i \in I} A_i$ is an open set.

b)

let $x \in \bigcap_{i \in I} A_i$

$$\forall j \in I : x \in A_j$$

$\forall j \in I, A_j$ is an open set, thus x is an interior point of A_j ,by definition:

$$\exists r_j > 0 : B(x, r_j) \subseteq A_j$$

define:

$$\tilde{r} = \min_{j \in I} \{r_j\}$$

$\tilde{r}$ exists because I is finite (completeness theorem).

in addition:

$$\tilde{r} > 0$$

since $\forall j \in I : \tilde{r} \leq r_j$:

$$\forall j \in I : B(x, \tilde{r}) \subseteq B(x, r_j) \subseteq A_j$$

thus:

$$B(x, \tilde{r}) \subseteq \bigcap_{i \in I} A_i$$

therefore x is an interior point of $\bigcap_{i \in I} A_i$.

every point in $\bigcap_{i \in I} A_i$ is an interior point, hence $\bigcap_{i \in I} A_i$ is an open set.

c)

let $n = 1$ and define:

$$I = \mathbb{N}$$
$$A_i = \left(-1, \frac{1}{i}\right)$$

$\forall i \in I$, A_i is an open set.

let's examine the intersection.

first it is trivial that every point $x \in (-1, 0)$ belongs to the intersection.

secondly, 0 also belongs to the intersection because:

$$\forall i \in I : -1 \leq 0 \leq \frac{1}{i} \longrightarrow 0 \in \bigcap_{i \in I} A_i$$

in addition, any point larger than 0 doesn't belong to the intersection because:

$$\forall \epsilon > 0, \exists i = \left\lceil \frac{1}{\epsilon} \right\rceil \in I : \epsilon \notin A_i$$

therefore 0 is a boundary point.

hence:

$$\bigcap_{i \in I} A_i = (-1, 0]$$

and of course this is not an open set, because it contains a boundary point.

note: we could generalized this example to any n by padding the vector with zeros.

Problem 6:

a)

$$f(x) = x^T A x$$

$f(x)$ is a quadratic combination of the coordinates of vector x thus, it consists of sums and multiplications of the vector coordinates, which results in an elementary function that is continuously differentiable infinite time. Therefore, $f(x)$ is continuously differentiable infinite times.

let's calculate the partial derivative by definition:

$$\begin{aligned}
\frac{\partial f(x)}{\partial x_k} &= f'(x; e_k) = \lim_{t\to 0} \frac{f(x+te_k) - f(x)}{t} = \\
&= \lim_{t\to 0} \frac{(x+te_k)^T A (x+te_k) - x^T A x}{t} = \\
&= \lim_{t\to 0} \frac{x^T A x + t x^T A e_k + t e_k^T A x + t^2 e_k^T e_k - x^T A x}{t} = \\
&= \lim_{t\to 0} \frac{t e_k^T A^T x + t e_k^T A x + t^2 \|e_k\|^2}{t} = \\
&= \lim_{t\to 0} \frac{t\left(e_k^T A^T x + e_k^T A x + t\right)}{t} = \lim_{t\to 0} e_k^T A^T x + e_k^T A x + t = \\
&= e_k^T A^T x + e_k^T A x = \left((Ae_k)^T + \left(A^T e_k\right)^T\right) x = \left(A_{k,:} + (A_{:,k})^T\right) x
\end{aligned}$$

thus:

$$\nabla f(x) = \left(A + A^T\right) x$$

if we derive a second time:

$$\begin{aligned}
\frac{\partial^2 f(x)}{\partial x_k x_l} &= \lim_{t\to 0} \frac{f'(x+te_l; e_k) - f'(x; e_k)}{t} = \\
&= \lim_{t\to 0} \frac{\left((Ae_k)^T + \left(A^T e_k\right)^T\right)(x+te_l) - \left((Ae_k)^T + \left(A^T e_k\right)^T\right) x}{t} = \\
&= \lim_{t\to 0} \frac{\left((Ae_k)^T + \left(A^T e_k\right)^T\right) te_l}{t} = \lim_{t\to 0} \left((Ae_k)^T + \left(A^T e_k\right)^T\right) e_l = \\
&= \left((Ae_k)^T + \left(A^T e_k\right)^T\right) e_l = e_k^T A^T e_l + e_k^T A e_l = A_{k,l}^T + A_{k,l} = A_{l,k} + A_{k,l}
\end{aligned}$$

therefore:

$$\nabla^2 f(x) = A + A^T$$

b)

$$f(x) = \left| \|x-a\|_2 - \delta \right|$$

$$f'(x;d) = \lim_{t\to 0^+} \frac{f(x+td) - f(x)}{t} = \lim_{t\to 0^+} \frac{\left|\|x+td-a\|_2 - \delta\right| - \left|\|x-a\|_2 - \delta\right|}{t}$$

if $\|x-a\|_2 > \delta > 0$ there exists small enough t such that:

$$\left|\|x+td-a\|_2 - \delta\right| = \|x+td-a\|_2 - \delta$$

and then:

$$f'(x;d) = \lim_{t\to 0^+} \frac{\|x+td-a\|_2 - \delta - \|x-a\|_2 + \delta}{t} = \lim_{t\to 0^+} \frac{\|x-a+td\|_2 - \|x-a\|_2}{t}$$

this is exactly the definition of the directional derivative of the l_2 norm of $(x-a)$ which we saw in the tutorial equals to (when $\|x-a\|_2 > 0$):

$$f'(x;d) = \frac{d^T(x-a)}{\|x-a\|_2}$$

if $\|x-a\|_2 < \delta$ there exists small enough t such that:

$$\left|\|x+td-a\|_2 - \delta\right| = \delta - \|x+td-a\|_2$$

and then:

$$f'(x;d) = \lim_{t\to 0^+} \frac{\delta - \|x+td-a\|_2 + \|x-a\|_2 - \delta}{t} = \lim_{t\to 0^+} -\frac{\|x-a+td\|_2 - \|x-a\|_2}{t} =$$
$$= -\lim_{t\to 0^+} \frac{\|x-a+td\|_2 - \|x-a\|_2}{t}$$

which again like we saw in the tutorial equals to:

$$f'(x;d) = \begin{cases} -\|d\|_2 & x-a=0 \\ -\frac{d^T(x-a)}{\|x-a\|_2} & 0 < \|x-a\|_2 < \delta \end{cases}$$

if $\|x-a\|_2 = \delta > 0$ then:

$$f'(x;d) = \lim_{t\to 0^+} \frac{\left|\|x+td-a\|_2 - \|x-a\|_2\right| - \left|\|x-a\|_2 - \|x-a\|_2\right|}{t} =$$
$$= \lim_{t\to 0^+} \frac{\left|\|x-a+td\|_2 - \|x-a\|_2\right|}{t} = \lim_{t\to 0^+} \left|\frac{\|x-a+td\|_2 - \|x-a\|_2}{t}\right|$$

which is again the definition of the directional derivative:

$$f'(x;d) = \left|\frac{d^T(x-a)}{\|x-a\|_2}\right| = \frac{\left|d^T(x-a)\right|}{\|x-a\|_2}$$

to summarize:

$$
\boxed{f^{'}(x;d) = \begin{cases} \frac{d^T(x-a)}{\|x-a\|_2} & \|x-a\|_2 > \delta \\ \frac{\left|d^T(x-a)\right|}{\|x-a\|_2} & \|x-a\|_2 = \delta \\ -\frac{d^T(x-a)}{\|x-a\|_2} & 0 < \|x-a\|_2 < \delta \\ -\|d\|_2 & \|x-a\|_2 = 0 \end{cases}}
$$