Advanced Linear Algebra Week 4 Day 1

2018/10/01 – Jonathan Hayase, updated by Prof. Weiqing Gu

1 Normed Linear (Vector) Space

1.1 Definition

A linear space X over K ($K = \mathbb{R}$ or $\mathbb{C}$) is a normed linear space if there exists $\|\cdot\| : X \to \mathbb{R}$ satisfying[1]

1. For $x \in X$, $\|x\| \geq 0$ and $x = 0$ if and only if $x = 0$.
2. Triangle inequality: For $x, y \in X$, $\|x + y\| \leq \|x\| + \|y\|$.
3. For $x \in X$ and $k \in K$, $\|kx\| = |k|\|x\|$.

1.2 Examples

1. $X = \mathbb{R}^n$ and $\|x\| = \sqrt{\langle x, x \rangle}$.
2. $X = K^n$ and define $\|x\|_\infty = \left\|(x_1, \ldots, x_n)\right\|_\infty = \max\left\{|x_1|, \ldots, |x_n|\right\}$.

   Note property 1. and 3. use

   For 3, $\|x + y\|_\infty =$[2]
3. $X = K^n$ where $x = (x_1, \ldots, x_n) \in X$. Define $\|x\|_1 = \sum_{i=1}^n |x_i|$.
4. $X = K^n$ with x as above. For $p \geq 1$, define the p-norm as

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

   It's each to check 1 and 3. The proof of 2 requires Holder's Inequality.

2 Holder's Inequality

Statement

For $1/p + 1/q = 1$

$$\left\|\langle x, y \rangle\right\| \leq \|x\|_p \|y\|_y$$

Note that if $p = q = 2$, then this is the Schwarz Inequality.

[1] missed something small here
[2] skipped some things here

$$
\begin{aligned}
\|x+y\|_p &= \left(\sum_{i=1}^{n} \|x_i + y_i\|^p\right) 6{}^1/_p \\
&\le \left(\sum_i \big(\|x_i\| + \|y_i\|\big)^p\right)^{1/p}
\end{aligned}
$$

Note that

$$
\sum \big(\|x_i\| + \|y_i\|\big)^p = \sum_i \big(\|x_i\| + \|y_i\|\big)\big(\|x_i\| + \|y_i\|\big)^{p/q}
$$

since $p - 1 = {}^p/_q$. Thus we have

$$
\begin{aligned}
\left(\sum_i \big(\|x_i\| + \|y_i\|\big)^p\right)^{1/p} &= \sum_i \|x_i\| \big(\|x_i\| + \|y_i\|\big)^{p/q} + \sum_i \|y_i\| \big(\|x_i\| + \|y_i\|\big)^{p/q}. \\
&\le \left(\sum_i \|x_i\|\right)^{1/p} \left(\left(\sum_i \big(\|x_i\| + \|y_i\|\big)^{p/q}\right)^q\right)^{1/q} + \left(\sum_i \|y_i\|\right)^{1/p} \left(\left(\sum_i \big(\|x_i\| + \|y_i\|\big)^{p/q}\right)^q\right)^{1/q} \\
&= \left[\left(\sum_i \|x_i\|^p\right)^{1/p} + \left(\sum_i \|y_i\|^p\right)^{1/p}\right] \left[\sum_i \big(\|x_i\| + \|y_i\|\big)^p\right]^{1/q} \\
\implies \left(\sum_i \big(\|x_i\| + \|y_i\|\big)^p\right)^{\cancel{1-1/q} \to 1/p} &\le \|x\|_p + \|y\|_p.
\end{aligned}
$$

Therefore

$$
\underbrace{\left(\sum_i \|x_i + y_i\|^p\right)^{1/p}}_{\|x+y\|_p} \le \|x\|_p + \|y\|_p
$$

[3]

Q: How are the different norms related.

Definition: Norms $|\cdot|$ and $\|\cdot\|$ on X are called **equivalent** if there exists constants c, C such that, for all $x \in X$,

$$
|x| < c\|x\| \text{ and } \|x\| < C|x|
$$

Theorem: Any two norms on a finite dimensional linear space X are equivalent.

[3] I'm not sure I follow exactly what happened here.

3 Banach Space

3.1 Definition

A Banach space is a vector space over a field K (say $K \in \{\mathbb{R}^n, \mathbb{C}^n\}$) which is equipped with a norm $\|\cdot\|$, which is complete with respect to the norm.

Suppose we have a curve (shaped like a hand). Rotating it does not change the shape. (In the complexes, we have $z_i \sim e^{i\theta} z_i$.)

3.2 Example

Consider the set space $C\,[a,b]$, the set of continuous functions on $[a,b] \subset \mathbb{R}$ with norm

$$\|f\| = \sup_{x \in [a,b]} \big|f(x)\big|$$

This norm does not come from an inner product. Therefore, all Hilbert spaces are Banach spaces, but the inverse is not true.

4 Bounded Linear Operator

4.1 Definition

Suppose we have normed linear spaces $\big(V,|\cdot|\big)$ and $\big(W,|\cdot|\big)$. And some $L : V \to W$. A linear operator (or transformation) L between two normed linear spaces if there exists an $M \in \mathbb{R}$ such that

$$\big\|L(v)\big\|_W \leq M|v|_v$$

or

$$\frac{\big\|L(v)\big\|_W}{|v|_v} \leq M.$$

4.2 Example 1

The shift operator on ℓ^2 space of all sequences $(x_0, x_1, \ldots, x_n, \ldots)$. Where $x_0^1 + x_1^2 + \cdots < \infty$ and

$$L(x_0, x_1, x_2, \ldots) = (0, x_0, x_1, \ldots)$$

is bounded. Its operator number[4] is 1.

4.3 Example 2

Integral transformation

$$K : [a,b] \times [c,d] \to \mathbb{R}$$

And

$$L\,f(y) = \int_{x=a}^{b} K(x,y) f(x) \,\mathrm{d}x$$

L is bounded.

[4] Not sure this word is right

5 Schatten p-norm

5.1 Definition

Let H_1, H_2 be separable Hilbert spaces and T be a linear bounded operator from H_1 to H_2. For $p \in [1, \infty)$, define the Schatten p-norm of this operator as

$$\|T\|_p \triangleq \left(\sum_{?} S_n^p(t) \right)^{1/p}$$

where $S_1(T) \geq S_2(T) \geq \cdots S_n(T) \geq \cdots \geq 0$, are the singular values of T.

If we have $T \leftrightarrow A$ then if $A^T A$ has eigenvalues $\lambda_1, \ldots, \lambda_n$, then $S_i = \sqrt{\lambda_i}$.

Note: $A^T A$ is semi-positive definite because

$$x^T \left(A^T A \right) x = (Ax)^T (Ax) = \|Ax\|^2 \geq 0$$

for all x. Thus Ax is semi-positive definite. So, $\lambda_i \geq 0$ for $i = 1, 2, \ldots$.

Note if $p = 2$,

$$\|T\|_2 = \sum_i \left(\sqrt{\lambda_i} \right)^2 = \lambda_1 + \cdots + \lambda_n = \operatorname{tr} T$$

Importantly, we can show $\|T\|_p^p = \operatorname{tr} \left(|T|^p \right)$.

6 Frechét Derivative

6.1 Definition

A Frechét derivative is a derivative on a Banach space.

It enables us to do calculus of variations.

Definition: Let $\left(V, |\cdot|_V\right)$ and $\left(W, |\cdot|_W\right)$ be normed spaces. Let $U \subseteq V$. Then $f : U \subseteq V \to W$ is called Frechét differentiable if there exists a bounded linear operator $A : V \to W$ such that

$$\lim_{h \to 0} \frac{\left\| f(x+h) - f(x) - Ah \right\|_W}{\|h\|_V} = 0$$

Recall: Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, then we have

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + \frac{1}{2} (\boldsymbol{x} - \boldsymbol{x}_0)^T \nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + \cdots .$$

Here,

$$f(x+h) = f(x) + Ah + o(h)$$

if there exists such an operator A then it is unique and we define

$$\mathrm{D} f(x) = A$$

and call this the Frechét derivative of f at x.

Say f is c' if $\mathrm{D} f : U \to B(v, w), x \to \mathrm{D} f(x) : V \to W$ Continuous for each value of x_0. Theorem: f is $c' \implies f$ is differentiable.

6.2 Example 1

Let $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$ and consider some $f : \mathbb{R}^n \to \mathbb{R}^m$.

- If $n = m = 1$, then f is a function.
- If $n > 1$ and $m = 1$ then f is a "graph".
- If $n = 1$ and $m > 1$ then f is a curve.
- $n > 1$ and $m > 1$

In this case, the Frechét Derivative is our usual derivative.

$$\mathrm{D}f = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_n}{\partial x_1} & \cdots & \dfrac{\partial f_n}{\partial x_n} \end{bmatrix} \begin{array}{l} \leftarrow \nabla f_1 \\ \\ \\ \leftarrow \nabla f_n \end{array}$$

6.3 Example 2

Consider $f : \left(M_{n,m}, \|\cdot\|\right) \to \left(M_{k,l}, \|\cdot\|\right)$

$$\underbrace{X}_{\begin{matrix} \| \\ \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nn} \end{bmatrix} \end{matrix}} = \underbrace{A}_{\text{is fixed}} X$$

f is called a matrix function.

6.4 Example 3

Consider $X \in \mathbb{R}^n \xrightarrow{f} x^T A x$

$$\mathrm{D}f = ?$$

Suppose $n = 2$ then

$$f \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

7 Matrix Calculus

Given dimension compatible matrix-valued forms of matrix variable $f(x)$ and $g(x)$.

$$\nabla_x \left[f(x)^T g(x) \right] = \nabla_x(f) g + \nabla_X(g) f$$

is the product rule.

E.g. $\nabla_x(x^T Ax) = \nabla_x(x)Ax + \nabla_x(Ax)x$.

$$x = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \to \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

In general

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \to \begin{bmatrix} g_{11}\left(x_{11}\right) & g_{12}\left(x_{12}\right) \\ g_{21}\left(x_{21}\right) & g_{22}\left(x_{22}\right) \end{bmatrix}$$

Then

$$\nabla g_{ij} = \begin{bmatrix} \dfrac{\partial g_{ij}}{\partial x_{11}} & \dfrac{\partial g_{ij}}{\partial x_{12}} \\ \dfrac{\partial g_{ij}}{\partial x_{21}} & \dfrac{\partial g_{ij}}{\partial x_{22}} \end{bmatrix}$$

and

$$\nabla^2 g = \begin{bmatrix} \nabla\left(\dfrac{\partial g}{\partial x_{11}}\right) & \cdots & \nabla\left(\dfrac{\partial g}{\partial x_{12}}\right) \\ \vdots & \ddots & \vdots \\ \nabla\left(\dfrac{\partial g}{\partial x_{21}}\right) & \cdots & \nabla\left(\dfrac{\partial g}{\partial x_{22}}\right) \end{bmatrix}$$

Hessian of f where $f : \mathbb{R}^n \to \mathbb{R}$,

$$\nabla f = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} & \cdots & \dfrac{\partial f}{\partial x_n} \end{bmatrix}$$

Then

$$\nabla^2 f = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

Q: How to find the Frechét derivative

$$A \xrightarrow{f} A^{-1}$$
$$M_n \xrightarrow{f} M_n$$

Claim

$$\mathrm{D}_A F(H) = -A^{-1} H A^{-1}$$

for all matrix H.

Assume $\left\|A^{-1}H\right\| \leq 1$, we have

$$\begin{aligned} (A+H)^{-1} &= \left(A\left(I + A^{-1}H\right)\right)^{-1} \\ &= (I + A^{-1}H)^{-1}A^{-1} = \sum_{k=0}^{\infty}(-1)^k (A^{-1}H)^k A^{-1} \\ (A+H)^{-1} &= A^{-1} - A^{-1}HA^{-1} + o(H) \\ \mathrm{D}F(A)(H) &= -A^{-1}HA^{-1} \end{aligned}$$

8 Hellinger Distance

The set of all probability distributions functions ????? space (locally looks like a vector space).

Let P, Q be two probability distributions with density functions p and q. Then

$$H^2(P,Q) = \frac{1}{2}\int (\sqrt{p} - \sqrt{q})^2 \, \mathrm{d}x = 1 - \int \sqrt{p(x)q(x)} \, \mathrm{d}x$$

HW Hint: You can use the Schwarz inequality to show $0 \leq H(P,Q) \leq 1$.

For discrete distributions

$$H(P,Q) = \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{k} \left(\sqrt{p} - \sqrt{q}\right)^2}$$

where

$$P = (p_1, \ldots, p_k) \text{ and } Q = (q_1, \ldots, q_n)$$

are probability distributions

9 KL Divergence

Discrete case

$$D_{KL}(P \mid\mid Q) = -\sum_i p(i) \log \frac{q(i)}{p(i)} = \sum_i p(i) \log \frac{p(i)}{q(i)}$$

Continuous case

$$D_{KL}(P \mid\mid Q) = -\int_{\infty}^{\infty} p(x) \log \frac{q(x)}{p(x)} \, \mathrm{d}x$$

Suppose $N_1 = N(\mu_1, \sigma_1^2)$ and $N_2 = N(\mu_2, \sigma_2^2)$. Then

$$D_{KL}(N_1 \| N_2) = \frac{1}{2}\left(\operatorname{tr}\left(\Sigma_1^{-1}\Sigma_o\right)\right) \operatorname{tr}\left(\mu_1 - \mu_0\right)^T - k + \ln\left(\frac{\det \Sigma_1}{\det \Sigma_0}\right)$$

10 Bhattacharyya Distance

$$D_b(P,Q) = -\ln\left(BC(P,Q)\right)$$

where

$$BC(P,Q) = \sum_{x \in X} \sqrt{p(x)q(x)}$$

Where $BC(P,Q)$ is called the Bhattacharyya coefficients.

$$D_B(p,q) = \frac{1}{4}\ln\left(\frac{1}{4}\left(\frac{\sigma_p^2}{\sigma_q^2} + \frac{\sigma_q^2}{\sigma_p^2} + 2\right)\right) + \frac{1}{4}\frac{\left(\mu_p - \mu_q\right)^2}{\sigma_p^2 + \sigma_q^2}$$

11 Rayleigh quotient

If M is positive definite, we can use M to define an inner product

$$\langle x, Mx\rangle = x^T Mx$$

Want to compare $\langle\cdot,\cdot\rangle_M$ with $\langle\cdot,\cdot\rangle_i = x^T x = \|x\|$.

$$xMx = q(x)$$

where q is a quadratic form, and

$$x^T Mx = \sum_{i,j=1}^{n} m_{ij} x_i x_j$$

In many applications, we want to maximize or minimize $q(x)$ subject to $\|x\| = 1$.

Note: If we solve this $q(x_0) \geq q(x)$, $\|x_0\| = 1$ and for all x with $\|x\| = 1$, for all x,

$$q\left(\frac{x}{\|x\|}\right) \leq q(x_0) \implies \frac{1}{\|x\|^2 q(x)} \leq q(x_0) = \max\left\{\frac{q(x)}{\|x\|^2}\right\}$$

Definition the Rayleigh quotion R of M determined by

$$Ra : X \setminus \{0\} \to R$$

by

$$R(x) = R_M(x) = \frac{q(x)}{p(x) =}$$

$\langle x, x\rangle$for $x \neq 0$.

Key: We can use $R(x)$ to calculate eigenvalues if we estimate eigenvectors.

Claim: R_M is continuous on $S = \big\{x \in X \,|\, \|x\| = 1\big\}$. Hint: $x_k \to x$, $\|x_k\| = 1$ show that $Mx_k \to Mx$.

$$\big|\langle x_k, Mx_k\rangle - \langle x, Mx\rangle\big|$$

And you can prove it...

Claim 2: Let $R(x_0) = \max R(x) \,|\, \|x\| = 1$. Then $R(x_0)$ $(= q(x_0))$ is an eigenvalue and x_0 is an eigenvector.

Claim 3: We can decompose $X = \operatorname{span}\{x_0\} \oplus \hat{X}$, where $\operatorname{span}\{x_0\} \perp \hat{X}$. Then for $\hat{x} \in \hat{X}$, $\langle M\hat{x}, x_0\rangle = \langle \hat{x}, Mx\rangle = \langle \hat{x}, ax_0\rangle = a\langle \hat{x}, x_0\rangle = 0$.

Now, M can be viewed as $\hat{X} \to \hat{X}$. We can iterate to find the eigenvalues and vectors.

Max/min principles any eigenvalue $a_j = \min_{\dim S=1}\Big\{\max_{x\in S, x\neq 0} \frac{\langle x, Mx\rangle}{\langle x, x\rangle}\Big\}$ for $x \in S$ and $x \neq 0$ and H $a_1 \leq a_2 \leq \cdots \leq a_n$.

Foundation for game theory.