FDU 高等线性代数 Homework 02

Due: Sept. 14, 2024
姓名: 雍崔扬
学号: 21307140051

Problem 1

设 m,n 为正整数且 $m\geq n$，给定列满秩矩阵 $A\in\mathbb{R}^{m\times n}$ 和向量 $b\in\mathbb{R}^m$
试证明 $\mathbb{R}^n$ 上的函数 $f(x)=\|b-Ax\|_2^2$ 的全局最小点为 $x^\star=(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b$

Solution:

$$
\begin{aligned}
\nabla f(x)&=\nabla_x\{\|b-Ax\|_2^2\}\\
&=\nabla_x\{b-Ax\}\cdot 2(b-Ax)\\
&=-A^{\mathrm{T}}\cdot 2(b-Ax)\\
&=-2(A^{\mathrm{T}}b-A^{\mathrm{T}}Ax)\\
\nabla^2 f(x)&=-2\nabla_x\{A^{\mathrm{T}}b-A^{\mathrm{T}}Ax\}\\
&=-2(-A^{\mathrm{T}}A)^{\mathrm{T}}\\
&=2A^{\mathrm{T}}A
\end{aligned}
$$

注意到 $\operatorname{rank}(A)=n\leq m$ 保证了 $\nabla^2 f(x)=2A^{\mathrm{T}}A\succ 0\ (\forall\, x\in\mathbb{R}^n)$
因此 $f(x)$ 在 $\mathbb{R}^n$ 上是严格凸函数，其全局最小点即驻点 (即令梯度为零的点)，且若存在则必定唯一.
令 $\nabla f(x)=-2(A^{\mathrm{T}}b-A^{\mathrm{T}}A)x=0_n$ 即可解得全局最小点为 $x_\star=(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b$
($A^{\mathrm{T}}A$ 的正定性保证了它是非奇异的)

Problem 2

设 n 为正整数，$A\in\mathbb{R}^{n\times n}$ 是给定的对称阵.
试求下列 $\mathbb{R}^n\mapsto\mathbb{R}$ 的函数的梯度:

$$
\begin{aligned}
f(x)&=\operatorname{tr}\left((A+xx^{\mathrm{T}})^2\right)\\
g(x)&=\operatorname{tr}\left((A+xx^{\mathrm{T}})^3\right)
\end{aligned}
$$

Solution:
首先求解 f 的梯度:

$$
\begin{aligned}
f(x)&=\operatorname{tr}\left((A+xx^{\mathrm{T}})^2\right)\\
&=\operatorname{tr}\left(A^2+xx^{\mathrm{T}}A+Axx^{\mathrm{T}}+x^{\mathrm{T}}xxx^{\mathrm{T}}\right)\\
&=\operatorname{tr}\left(A^2\right)+\operatorname{tr}\left(xx^{\mathrm{T}}A\right)+\operatorname{tr}\left(Axx^{\mathrm{T}}\right)+x^{\mathrm{T}}x\operatorname{tr}\left(xx^{\mathrm{T}}\right)\\
&=\operatorname{tr}\left(A^2\right)+2\operatorname{tr}\left(x^{\mathrm{T}}Ax\right)+x^{\mathrm{T}}x\operatorname{tr}\left(x^{\mathrm{T}}x\right)\\
&=\operatorname{tr}\left(A^2\right)+2x^{\mathrm{T}}Ax+(x^{\mathrm{T}}x)^2\\
\hline
\nabla f(x)&=2\nabla_x\{x^{\mathrm{T}}Ax\}+\nabla_x\{(x^{\mathrm{T}}x)^2\}\\
&=2\cdot(A+A^{\mathrm{T}})x+\nabla_x\{x^{\mathrm{T}}x\}\cdot 2(x^{\mathrm{T}}x)\quad(\text{note that }A^{\mathrm{T}}=A)\\
&=4Ax+2x\cdot 2(x^{\mathrm{T}}x)\\
&=4(A+xx^{\mathrm{T}})x
\end{aligned}
$$

接下来求解 g 的梯度:

$$
\begin{aligned}
g(x) &= \operatorname{tr}\left((A+xx^{\mathrm{T}})^3\right) \\
&= \operatorname{tr}\left(A^3 + xx^{\mathrm{T}}A^2 + Axx^{\mathrm{T}}A + x^{\mathrm{T}}xxx^{\mathrm{T}}A + A^2xx^{\mathrm{T}} + xx^{\mathrm{T}}Axx^{\mathrm{T}} + Axx^{\mathrm{T}}xx^{\mathrm{T}} + (x^{\mathrm{T}}x)^2xx^{\mathrm{T}}\right) \\
&= \operatorname{tr}\left(A^3\right) + \operatorname{tr}\left(xx^{\mathrm{T}}A^2\right) + \operatorname{tr}\left(Axx^{\mathrm{T}}A\right) + \operatorname{tr}\left(x^{\mathrm{T}}xxx^{\mathrm{T}}A\right) + \operatorname{tr}\left(A^2xx^{\mathrm{T}}\right) + \operatorname{tr}\left(xx^{\mathrm{T}}Axx^{\mathrm{T}}\right) \\
&\quad + \operatorname{tr}\left(Axx^{\mathrm{T}}xx^{\mathrm{T}}\right) + \operatorname{tr}\left((x^{\mathrm{T}}x)^2xx^{\mathrm{T}}\right) \\
&= \operatorname{tr}\left(A^3\right) + 3\operatorname{tr}\left(x^{\mathrm{T}}A^2x\right) + 3(x^{\mathrm{T}}x)\operatorname{tr}\left(x^{\mathrm{T}}Ax\right) + (x^{\mathrm{T}}x)^2\operatorname{tr}\left(x^{\mathrm{T}}x\right) \\
&= \operatorname{tr}\left(A^3\right) + 3x^{\mathrm{T}}A^2x + 3(x^{\mathrm{T}}x)(x^{\mathrm{T}}Ax) + (x^{\mathrm{T}}x)^3
\end{aligned}
$$

$$
\begin{aligned}
\nabla g(x) &= 3\nabla_x\{x^{\mathrm{T}}A^2x\} + 3\nabla_x\{(x^{\mathrm{T}}x)(x^{\mathrm{T}}Ax)\} + \nabla_x\{(x^{\mathrm{T}}x)^3\} \\
&= 3(A^2 + (A^2)^{\mathrm{T}})x + 3[(x^{\mathrm{T}}x)\cdot\nabla_x\{x^{\mathrm{T}}Ax\} + \nabla_x\{x^{\mathrm{T}}x\}\cdot(x^{\mathrm{T}}Ax)] + 3(x^{\mathrm{T}}x)^2\nabla_x\{x^{\mathrm{T}}x\} \\
&= 3(A^2 + (A^2)^{\mathrm{T}})x + 3[(x^{\mathrm{T}}x)\cdot(A + A^{\mathrm{T}})x + 2x\cdot(x^{\mathrm{T}}Ax)] + 3(x^{\mathrm{T}}x)^2\cdot 2x \quad (\text{note that } A^{\mathrm{T}} = A) \\
&= 6A^2x + 6(x^{\mathrm{T}}x)Ax + 6(x^{\mathrm{T}}Ax)x + 6(x^{\mathrm{T}}x)^2x \\
&= 6(A^2 + Axx^{\mathrm{T}} + xx^{\mathrm{T}}A + xx^{\mathrm{T}}xx^{\mathrm{T}})x \\
&= 6(A + xx^{\mathrm{T}})^2x
\end{aligned}
$$

Problem 3

设 n 为正整数，$a, b \in \mathbb{R}^n$ 为给定的向量.
试求下列 $\mathbb{R}^{n\times n} \mapsto \mathbb{R}$ 的函数的梯度:

$$f(X) = a^{\mathrm{T}}X^{\mathrm{T}}Xb$$

邵老师提供的解法:

> $\nabla_X \operatorname{tr}(AX) = \nabla_X \operatorname{tr}(XA) = \nabla_X \operatorname{tr}\left(A^{\mathrm{T}}X^{\mathrm{T}}\right) = \nabla_X \operatorname{tr}\left(X^{\mathrm{T}}A^{\mathrm{T}}\right) = A^{\mathrm{T}}$ (特殊地我们有 $\nabla_X \operatorname{tr}(X) = I$)
> 据此可推出 $\nabla_X a^{\mathrm{T}}Xb = \nabla_X \operatorname{tr}\left(a^{\mathrm{T}}Xb\right) = \nabla_X \operatorname{tr}\left(ba^{\mathrm{T}}X\right) = \nabla_X \operatorname{tr}\left(X^{\mathrm{T}}ab^{\mathrm{T}}\right) = \nabla_X b^{\mathrm{T}}X^{\mathrm{T}}a = ab^{\mathrm{T}}$
> 一般来说我们有 $\nabla_X \operatorname{tr}\left(AX^k\right) = \nabla_X \operatorname{tr}\left(X^kA\right) = \sum_{i=0}^{k-1}(X^iAX^{k-1-i})^{\mathrm{T}}$

现在我们计算 $f(X)$ 的梯度:

$$
\begin{aligned}
\nabla_X f(X) &= \nabla_X\{a^{\mathrm{T}}X^{\mathrm{T}}Xb\} \\
&= \nabla_X\{\operatorname{tr}\left(a^{\mathrm{T}}X^{\mathrm{T}}Xb\right)\} \\
&= \nabla_X\{\operatorname{tr}\left(X^{\mathrm{T}}Xba^{\mathrm{T}}\right)\} \\
&= \left[\nabla_{X_1}\{\operatorname{tr}\left(X_1^{\mathrm{T}}X_2ba^{\mathrm{T}}\right)\} + \nabla_{X_2}\{\operatorname{tr}\left(ba^{\mathrm{T}}X_1^{\mathrm{T}}X_2\right)\}\right]\Big|_{X_1=X_2=X} \\
&= \left[X_2ba^{\mathrm{T}} + (ba^{\mathrm{T}}X_1^{\mathrm{T}})^{\mathrm{T}}\right]\Big|_{X_1=X_2=X} \\
&= X(ba^{\mathrm{T}} + ab^{\mathrm{T}})
\end{aligned}
$$

My Solution:
首先我们证明一个引理:

> **Lemma:** $\nabla_Y\{a^{\mathrm{T}}Yb\} = ab^{\mathrm{T}}$ $(\forall\, Y = [y_{ij}] \in \mathbb{C}^{n\times n})$
> **Proof:**
>
> $$
> \begin{aligned}
> \nabla_Y\{a^{\mathrm{T}}Yb\} &= \left[\frac{\partial(a^{\mathrm{T}}Yb)}{\partial y_{ij}}\right]_{i,j=1}^n \\
> &= \left[\frac{\partial}{\partial y_{ij}}\sum_{k,l=1}^n a_ky_{kl}b_l\right]_{i,j=1}^n \\
> &= [a_ib_j]_{i,j=1}^n \\
> &= ab^{\mathrm{T}}
> \end{aligned}
> $$

现在我们计算 $f(X)$ 的梯度:

$$
\begin{aligned}
\nabla_X f(x) &= \nabla_X\{a^{\mathrm{T}}X^{\mathrm{T}}Xb\} \\
&= \left[\nabla_{X_1}\{a^{\mathrm{T}}X_1^{\mathrm{T}}X_2b\} + \nabla_{X_2}\{a^{\mathrm{T}}X_1^{\mathrm{T}}X_2b\}\right]\Big|_{X_1=X_2=X} \\
&= \left[\nabla_{X_1}\{(X_2b)^{\mathrm{T}}X_1a\} + \nabla_{X_2}\{(X_1a)X_2b\}\right]\Big|_{X_1=X_2=X} \quad (\text{utilizing Lemma:}\nabla_Y\{a^{\mathrm{T}}Yb\} = ab^{\mathrm{T}}) \\
&= \left[(X_2b)a^{\mathrm{T}}\} + \{(X_1a)b^{\mathrm{T}}\}\right]\Big|_{X_1=X_2=X} \\
&= Xba^{\mathrm{T}} + Xab^{\mathrm{T}} \\
&= X(ba^{\mathrm{T}} + ab^{\mathrm{T}})
\end{aligned}
$$

下面是另一种更繁难的解法:
(不幸的是，这是我最开始想到的解法)

记 X 的行向量为 $x_1^{\mathrm{T}},\dots,x_n^{\mathrm{T}}$，记 X 的 (i,j) 位置元素为 x_{ij}
定义 Kronecker δ-函数 $\delta_{ij} := \begin{cases} 1 & \text{if } i=j \\ 0 & \text{otherwise} \end{cases}$

$$\begin{aligned}
\frac{\partial (X^{\mathrm{T}}X)_{kl}}{\partial x_{ij}} &= \frac{\partial}{\partial x_{ij}}\{x_k^{\mathrm{T}}x_l\} \\
&= \frac{\partial}{\partial x_{ij}}\left\{\sum_{m=1}^n x_{km}x_{lm}\right\} \\
&= \delta_{ik}x_{lj} + \delta_{il}x_{kj}
\end{aligned}$$

定义单位基矩阵 E_{ij} (它在 (i,j) 位置上为 1，在其余位置为零)
注意到 $(E_{ij})_{kl} = \delta_{ik}\delta_{jl}$)，因此我们有:

$$\begin{aligned}
\nabla_{x_{ij}}\{X^{\mathrm{T}}X\} &= \left[\frac{\partial (X^{\mathrm{T}}X)_{kl}}{\partial x_{ij}}\right]_{k,l=1}^n \\
&= [\delta_{ik}x_{lj} + \delta_{il}x_{kj}]_{k,l=1}^n \\
&= [\delta_{ik}x_{lj}]_{k,l=1}^n + [\delta_{il}x_{kj}]_{k,l=1}^n \\
&= X^{\mathrm{T}}E_{ij} + E_{ji}X
\end{aligned}$$

记 e_i 为 $\mathbb{C}^n$ 的标准单位基向量，记 $a=[a_1,\dots,a_n]^{\mathrm{T}}, b=[b_1,\dots,b_n]^{\mathrm{T}}$，回忆起 X 的行向量为 $x_1^{\mathrm{T}},\dots,x_n^{\mathrm{T}}$
则我们有:

$$\begin{aligned}
\nabla f(X) &= \nabla_X\{a^{\mathrm{T}}X^{\mathrm{T}}Xb\} && (\text{utilizing Lemma:}\nabla_Y\{a^{\mathrm{T}}Yb\} = ab^{\mathrm{T}}) \\
&= \nabla_X\{X^{\mathrm{T}}X\}\cdot(ab^{\mathrm{T}}) && (\text{note that } \nabla_X\{X^{\mathrm{T}}X\} \in \mathbb{R}^{n\times n\times n\times n} \text{ is a four dimensional tensor}) \\
&= [a^{\mathrm{T}}\nabla_{x_{ij}}\{X^{\mathrm{T}}X\}b]_{i,j=1}^n \\
&= [a^{\mathrm{T}}(X^{\mathrm{T}}E_{ij} + E_{ji}X)b]_{i,j=1}^n \\
&= [a^{\mathrm{T}}X^{\mathrm{T}}E_{ij}b]_{i,j=1}^n + [a^{\mathrm{T}}E_{ji}Xb]_{i,j=1}^n \\
&= [a^{\mathrm{T}}X^{\mathrm{T}}b_je_i]_{i,j=1}^n + [a_je_i^{\mathrm{T}}Xb]_{i,j=1}^n \\
&= [b_ja^{\mathrm{T}}x_i]_{i,j=1}^n + [a_jx_i^{\mathrm{T}}b]_{i,j=1}^n \\
&= [(x_i^{\mathrm{T}}a)b_j]_{i,j=1}^n + [(x_i^{\mathrm{T}}b)a_j]_{i,j=1}^n \\
&= (Xa)b^{\mathrm{T}} + (Xb)a^{\mathrm{T}} \\
&= X(ab^{\mathrm{T}} + ba^{\mathrm{T}})
\end{aligned}$$

Problem 4

定义 $\mathbb{C}^2$ 上的实值函数 $\|\cdot\|_{1/2}$ 为 $\|x\|_{1/2} := (|x_1|^{\frac{1}{2}} + |x_2|^{\frac{1}{2}})^2\ (\forall\, x\in\mathbb{C}^2)$
试证明 $\|\cdot\|_{1/2}$ 不是 $\mathbb{C}^2$ 上的范数.

Solution:
非负性、正定性、齐次性和连续性 $\|\cdot\|_{1/2}$ 显然是满足的，因此它是一个**准范数** (pre-norm)

> **(Matrix Analysis 定理 $5.4.4$ 中的定义)**
> 设 f 是域 $\mathbb{F}=\mathbb{R}$ or $\mathbb{C}$ 上的 n 维向量空间 V 上的实值函数
> 若 f 满足:
>
> - 非负性: $f(x)\geq 0\ (\forall\, x\in V)$
> - 正定性: $f(x)=0$ 当且仅当 $x=0_V$
> - 齐次性: $f(\alpha x) = |\alpha| f(x)\ (\forall\, x\in V, \alpha\in\mathbb{F})$
> - 连续性: $f(x)$ 在 $\mathbb{F}^n$ 上关于 Euclid 范数是连续的
>
> 则我们称满足上述性质的 f 称为**准范数** (pre-norm)
> 满足三角不等式的准范数就是范数.

下面我们只要举例说明 $\|\cdot\|_{1/2}$ 不满足三角不等式即可.
假设 $x,y\in\mathbb{R}_+^2$ (即它们的元素都是非负实数)，则我们有:

$$
\begin{aligned}
&\|x+y\|_{1/2}-\|x\|_{1/2}-\|y\|_{1/2} \\
&=\left[(x_1+y_1)^{\frac{1}{2}}+(x_2+y_2)^{\frac{1}{2}}\right]^2-(x_1^{\frac{1}{2}}+x_2^{\frac{1}{2}})^2-(y_1^{\frac{1}{2}}+y_2^{\frac{1}{2}})^2 \\
&=2\left[(\frac{1}{2}x_1+\frac{1}{2}y_1)^{\frac{1}{2}}+(\frac{1}{2}x_2+\frac{1}{2}y_2)^{\frac{1}{2}}\right]^2-(x_1^{\frac{1}{2}}+x_2^{\frac{1}{2}})^2-(y_1^{\frac{1}{2}}+y_2^{\frac{1}{2}})^2 \quad (\text{use concavity of } f(x)=\sqrt{x}\ (x\geq 0)) \\
&\geq 2\left[(\frac{1}{2}x_1^{\frac{1}{2}}+\frac{1}{2}y_1^{\frac{1}{2}})+(\frac{1}{2}x_2^{\frac{1}{2}}+\frac{1}{2}y_2^{\frac{1}{2}})\right]^2-(x_1^{\frac{1}{2}}+x_2^{\frac{1}{2}})^2-(y_1^{\frac{1}{2}}+y_2^{\frac{1}{2}})^2 \\
&=\frac{1}{2}\left[(x_1^{\frac{1}{2}}+x_2^{\frac{1}{2}})+(y_1^{\frac{1}{2}}+y_2^{\frac{1}{2}})\right]^2-(x_1^{\frac{1}{2}}+x_2^{\frac{1}{2}})^2-(y_1^{\frac{1}{2}}+y_2^{\frac{1}{2}})^2
\end{aligned}
$$

注意到最末端的式子在 $x_1^{\frac{1}{2}}+x_2^{\frac{1}{2}}=y_1^{\frac{1}{2}}+y_2^{\frac{1}{2}}$ 时等于零
为要求凹性保证的不等式严格成立，我们还要求 $x\neq y$
例如我们可以取:

$$
x=\begin{bmatrix}1\\2\end{bmatrix}\quad y=\begin{bmatrix}2\\1\end{bmatrix}\quad x+y=\begin{bmatrix}3\\3\end{bmatrix}
$$

$$
\begin{aligned}
\|x+y\|_{1/2}-\|x\|_{1/2}-\|y\|_{1/2}&=(\sqrt{3}+\sqrt{3})^2-(1+\sqrt{2})^2-(\sqrt{2}+1)^2 \\
&=12-2(3+2\sqrt{2}) \\
&=2(3-2\sqrt{2}) \\
&=2(\sqrt{9}-\sqrt{8}) \\
&>0
\end{aligned}
$$

因此三角不等式 $\|x+y\|_{1/2}-\|x\|_{1/2}-\|y\|_{1/2}\leq 0$ 并不对任意 $x,y\in\mathbb{C}^2$ 成立.
这表明 $\|\cdot\|_{1/2}$ 不是 $\mathbb{C}^2$ 上的范数.

Problem 5

($\mathbb{C}^n$ 上范数的等价性)
设 n 为正整数，$\|\cdot\|_\alpha,\|\cdot\|_\beta$ 为 $\mathbb{C}^n$ 上的范数.
试证明存在正实数 $C_{\min}$ 和 $C_{\max}$ 使得:

$$
C_{\min}\|x\|_\alpha\leq\|x\|_\beta\leq C_{\max}\|x\|_\alpha\ (\forall\ x\in\mathbb{C}^n)
$$

> **Lemma: (Weierstrass 定理, Nonlinear Programming 命题 A.8)**
> 设 $\mathcal{X}$ 为 $\mathbb{F}=\mathbb{R}$ or $\mathbb{C}$ 上的有限维赋范空间 V 的非空子集
> 且 $f:\mathcal{X}\mapsto\mathbb{R}$ 在 $\mathcal{X}$ 处下半连续
> 即对于满足 $\lim\limits_{k\to\infty}x^{(k)}=x$ 的每一个序列 $\{x^{(k)}\}\subset\mathcal{X}$ 都有 $f(x)\leq\lim\limits_{k\to\infty}\inf f(x^{(k)})$
> (特殊地，连续函数一定下半连续)
>
> 若下列条件有一个成立:
>
> - ① $\mathcal{X}$ 是紧集 (即有界闭集)
> - ② $\mathcal{X}$ 是闭集，且 f 在 $x\in\mathcal{X}$ 上是强制的 (coercive)
>   即对于满足 $\lim\limits_{k\to\infty}\|x^{(k)}\|=\infty$ 的每一个序列 $\{x^{(k)}\}\subset\mathcal{X}$ 都有 $\lim\limits_{k\to\infty}f(x^{(k)})=\infty$
> - ③ 存在 $\alpha\in\mathbb{R}$ 使得下水平集 $\{x\in\mathcal{X}:f(x)\leq\alpha\}$ 是紧集
>
> 则 f 在 $\mathcal{X}$ 上的全局最小点的集合 $\arg\min\limits_{x\in\mathcal{X}}f(x)$ 为非空紧集.

Solution:
在 Euclid 单位球面 $S:=\{x\in\mathbb{C}^n:\|x\|_2=1\}$ (它是一个紧集) 上定义 $h(x)=\frac{\|x\|_\beta}{\|x\|_\alpha}$
根据范数的连续性可知 h 是紧集 S 上的连续函数
根据 **Weierstrass 定理**可知 h 在 S 上存在有限的最小值 $C_{\min}$ 和最大值 $C_{\max}$
即对于任意 $x\in S$ (即 $x\in\mathbb{C}^n$ 且 $\|x\|_2=1$) 我们都有 $C_{\min}\leq h(x)=\frac{\|x\|_\beta}{\|x\|_\alpha}\leq C_{\max}$
根据范数的非负性和正定性，并结合 $0_n\notin S$ 可知 $C_{\min},C_{\max}>0$

根据范数的齐次性我们可知: 对于任意非零向量 $x\in\mathbb{C}^n$ 都有:

$$
C_{\min}\leq h\left(\frac{x}{\|x\|_2}\right)=\frac{\|\frac{x}{\|x\|_2}\|_\beta}{\|\frac{x}{\|x\|_2}\|_\alpha}=\frac{\frac{1}{\|x\|_2}\|x\|_\beta}{\frac{1}{\|x\|_2}\|x\|_\alpha}=\frac{\|x\|_\beta}{\|x\|_\alpha}\leq C_{\max}
$$

$$
\Leftrightarrow
$$

$$
C_{\min}\|x\|_\alpha\leq\|x\|_\beta\leq C_{\max}\|x\|_\alpha
$$

显然 $x=0_n$ 也满足不等式 $C_{\min}\|x\|_\alpha \le \|x\|_\beta \le C_{\max}\|x\|_\alpha$
因此我们有:

$$C_{\min}\|x\|_\alpha \le \|x\|_\beta \le C_{\max}\|x\|_\alpha \ \ (\forall\, x \in \mathbb{C}^n)$$

命题得证.

Problem 6 (optional)

设 n 为正整数.
试求下列函数的梯度:

$$f(X) = \log\left(\det\left(X\right)\right) \text{ where } \operatorname{dom}(f) = \{X \in \mathbb{R}^{n\times n} : \det\left(X\right) > 0\}$$

Solution:

注意到伴随矩阵的定义保证了 $\nabla_A \det\left(A\right) = \left(\frac{\partial \det(A)}{\partial A}\right)^{\mathrm{T}} = \operatorname{adj}(A)^{\mathrm{T}}$

$$\operatorname{adj}(A) := \left(\left[\frac{\partial}{\partial a_{ij}}\det\left(A\right)\right]_{i,j=1}^n\right)^{\mathrm{T}} = \{[(-1)^{i+j}\det\left(A[\{i\}^c,\{j\}^c]\right)]_{i,j=1}^n\}^{\mathrm{T}} = [(-1)^{i+j}\det\left(A[\{j\}^c,\{i\}^c]\right)]_{i,j=1}^n$$

其中 $A[\{i\}^c,\{j\}^c]$ 代表 A 删去第 i 行和第 j 列所有元素构成的余子阵.

容易验证 $[\operatorname{adj}(A)A]_{i,j} = \begin{cases} \det\left(A\right) & \text{if } i=j \\ 0 & \text{otherwise} \end{cases}$
这表明 $\operatorname{adj}(A)A = \det\left(A\right)I_n$
当 A 非奇异时我们就有 $\operatorname{adj}(A) = \det\left(A\right)A^{-1}$

因此对于任意满足 $\det\left(X\right) > 0$ 的 $X \in \mathbb{R}^{n\times n}$ 我们都有:

$$\begin{aligned}\nabla f(X) &= \nabla_X\{\log\left(\det\left(X\right)\right)\} \\ &= \frac{1}{\det\left(X\right)}\nabla_X\{\det\left(X\right)\} \\ &= \frac{1}{\det\left(X\right)}\operatorname{adj}(X)^{\mathrm{T}} \quad (\text{note that } \operatorname{adj}(X) = \det\left(X\right)X^{-1}) \\ &= X^{-\mathrm{T}}\end{aligned}$$

Problem 7 (optional)

考虑闭区间 $[0,1]$ 上的实值连续函数构成的空间 $C_{[0,1]}$
试证明下列定义的 $\mathbb{C}_{[0,1]} \mapsto \mathbb{R}$ 的泛函 $\|\cdot\|_2$ 是范数:

$$\|f\|_2 := \left(\int_0^1 |f(x)|^2 \mathrm{d}x\right)^{\frac{1}{2}}$$

Solution:

- ① 非负性显然成立.
- ② 正定性:

$$\|f\|_2 = \left(\int_0^1 |f(x)|^2 \mathrm{d}x\right)^{\frac{1}{2}} = 0 \quad \Leftrightarrow \quad f(x) = 0 \text{ almost everywhere on } [0,1]$$

  因此如果我们将 $C_{[0,1]}$ 空间中的几乎处处相等的函数视为同一个函数，那么泛函 $\|\cdot\|_2$ 满足正定性.

- ③ 齐次性:
  对于任意 $\alpha \in \mathbb{R}$ 都有:

$$\begin{aligned}\|\alpha f\|_2 &= \left(\int_0^1 |\alpha f(x)|^2 \mathrm{d}x\right)^{\frac{1}{2}} \\ &= |\alpha|\left(\int_0^1 |f(x)|^2 \mathrm{d}x\right)^{\frac{1}{2}} \\ &= |\alpha|\|f\|_2\end{aligned}$$

- ④ 次可加性 (三角不等式): (这实际上是积分形式的 Minkowski 不等式 $p=2$ 的特殊情况)
  对于任意 $f,g\in C_{[0,1]}$ (假设 $f+g$ 在 $[0,1]$ 上不几乎处处为零)，我们都有:

$$\begin{aligned}\int_0^1|f(x)+g(x)|^2\mathrm{d}x&=\int_0^1|f(x)+g(x)|\cdot|f(x)+g(x)|\mathrm{d}x\\&\leq\int_0^1|f(x)+g(x)||f(x)|\mathrm{d}x+\int_0^1|f(x)+g(x)||g(x)|\mathrm{d}x\\&\leq\left(\int_0^1|f(x)+g(x)|^2\mathrm{d}x\right)^{\frac{1}{2}}\left(\int_0^1|f(x)|^2\mathrm{d}x\right)^{\frac{1}{2}}+\left(\int_0^1|f(x)+g(x)|^2\mathrm{d}x\right)^{\frac{1}{2}}\left(\int_0^1|g(x)|^2\mathrm{d}x\right)^{\frac{1}{2}}\\&=\left(\int_0^1|f(x)+g(x)|^2\mathrm{d}x\right)^{\frac{1}{2}}\left[\left(\int_0^1|f(x)|^2\mathrm{d}x\right)^{\frac{1}{2}}+\left(\int_0^1|g(x)|^2\mathrm{d}x\right)^{\frac{1}{2}}\right]\end{aligned}$$

  左右同除 $(\int_0^1|f(x)+g(x)|^2\mathrm{d}x)^{\frac{1}{2}}$ 可得:

$$\left(\int_0^1|f(x)+g(x)|^2\mathrm{d}x\right)^{\frac{1}{2}}\leq\left(\int_0^1|f(x)|^2\mathrm{d}x\right)^{\frac{1}{2}}+\left(\int_0^1|g(x)|^2\mathrm{d}x\right)^{\frac{1}{2}}$$
$$\Leftrightarrow$$
$$\|f+g\|_2\leq\|f\|_2+\|g\|_2$$

  事实上，当 $f+g$ 在 $[0,1]$ 上恒等于零 (即 $\|f+g\|_2=0$) 时，上述不等式仍然成立.
  因此对于任意 $f,g\in C_{[0,1]}$ 都有 $\|f+g\|_2\leq\|f\|_2+\|g\|_2$ 成立.

  邵老师提供了关于次可加性的另一种证法:

$$\begin{aligned}\|f+g\|_2^2&=\int_0^1|f(x)+g(x)|^2\mathrm{d}x\\&=\int_0^1\{|f(x)|^2+2|f(x)g(x)|+|g(x)|^2\}\mathrm{d}x\\&=\int_0^1|f(x)|^2\mathrm{d}x+\int_0^1|g(x)|^2\mathrm{d}x+2\int_0^1|f(x)g(x)|\mathrm{d}x\quad(\text{Cauchy-Schwarz})\\&\leq\int_0^1|f(x)|^2\mathrm{d}x+\int_0^1|g(x)|^2\mathrm{d}x+2\left(\int_0^1|f(x)|^2\mathrm{d}x\right)^{\frac{1}{2}}\left(\int_0^1|g(x)|^2\mathrm{d}x\right)^{\frac{1}{2}}\\&=\|f\|_2^2+\|g\|_2^2+2\|f\|_2\|g\|_2\\&=(\|f\|_2+\|g\|_2)^2\end{aligned}$$

  因此对于任意 $f,g\in C_{[0,1]}$ 都有 $\|f+g\|_2\leq\|f\|_2+\|g\|_2$ 成立.

综上所述，泛函 $\|\cdot\|_2$ 是 $C_{[0,1]}$ 空间上的范数.

The End