线性空间

- 设 A、B 是线性空间 V 的线性子空间，则
  - 称 $A+B \triangleq \{x+y \mid x\in A \wedge y\in B\}$ 为 A 和 B 的和
  - 称 $A\bigcap B \triangleq \{x \mid x\in A \wedge x\in B\}$ 为 A 和 B 的交
  - 若 $A+B=V \wedge A\bigcap B=\{0\}$，则称 V 是 A 和 B 的直接求和，简称直和，记作 $V=A\oplus B$
- 定理：$A+B$ 和 $A\bigcap B$ 是 V 的线性子空间
- 推论：$A+B$ 是 V 中包含 A 和 B 的最小线性子空间，$A\bigcap B$ 是 V 中既属于 A 又属于 B 的最大线性子空间

内积空间

- 定义：设 U 是数域 $K(\mathbb{R} or \mathbb{C})$ 上的线性空间，若 $\forall x,y\in U$，存在唯一的数 $\langle x,y\rangle\in K$，满足下列三条内积公理：
  - 正定性：$\langle\alpha,\alpha\rangle\geq 0,\langle\alpha,\alpha\rangle=0\Leftrightarrow\alpha=0$
  - 对第一变元线性：$\langle k_1\alpha+k_2\beta,\gamma\rangle=k_1\langle\alpha,\gamma\rangle+k_2\langle\beta,\gamma\rangle$
  - 共轭对称性：$\langle\alpha,\beta\rangle=\overline{\langle\beta,\alpha\rangle}$

  则称 $\langle\xi,\eta\rangle$ 为 ξ,η 的内积，U 为内积空间。
- 推论：$\langle\alpha,k_1\beta+k_2\gamma\rangle=\overline{k_1}\langle\alpha,\beta\rangle+\overline{k_2}\langle\alpha,\gamma\rangle$
- 性质：
  - 内积的连续性
  - 内积空间的完备性：内积空间 V 中的任意柯西序列收敛于 V 内。
- 典范内积：n 阶复向量 x,y 之间的内积 $\langle\alpha,\beta\rangle\triangleq\alpha^H\beta=\sum_{i=1}^{n}\overline{x_i}y_i$，其中 $A^H\triangleq\overline{A}^T$
- 函数向量内积：$\langle\alpha(t),\beta(t)\rangle\triangleq\int_a^b\alpha^H(t)\beta(t)dt$

赋范空间

- 定义：在内积空间 U 中，若令 $\|\alpha\|=\sqrt{\langle\alpha,\alpha\rangle}$，即 $\|\alpha\|^2=\langle\alpha,\alpha\rangle$，则其满足范数公理：
  - $\|\alpha\|\geq 0,\|\alpha\|=0\Leftrightarrow\alpha=0$
  - $\|k\alpha\|=|k|\cdot\|\alpha\|$

- $\|\alpha+\beta\|\le\|\alpha\|+\|\beta\|$

称 U 是按内积导出的赋范线性空间。进一步可由范数导出距离 $\rho(\alpha,\beta)=\|\alpha-\beta\|$，则 U 也是距离空间

- 常见向量范数：
  - 0 范数：$\|\alpha\|_0 \triangleq$ 非零元素个数
  - 无穷范数、极大范数：$\|\alpha\|_\infty \triangleq \max\{|x_i|\}$
  - p 范数：$\|\alpha\|_p=(\sum_{i=1}^n \|x_i\|^p)^{1/p}, p\ge 1$
  - 函数向量范数：$\|\alpha(t)\|\triangleq(\int_a^b \alpha^H(t)\alpha(t)dt)^{1/2}$
- 2 范数 $\|\bullet\|_2$ 是酉不变的：$\|U\alpha\|_2=\|\alpha\|_2$，其中 U 是正交矩阵

Hilbert 空间
定义：完备的赋范线性空间称为 Banach 空间
定义：完备的内积空间 U 称为 Hilbert 空间，记作 H

矩阵

- 向量化：令 $m\times n$ 复矩阵 $A=[\alpha_1\cdots\alpha_n]$，$vec(A)\triangleq[\alpha_1{}^T\cdots\alpha_n{}^T]^T$
- 内积：$\langle A,B\rangle:\mathbb{C}^{m\times n}\times\mathbb{C}^{m\times n}\to\mathbb{C}$，$\langle A,B\rangle\triangleq\langle vec(A),vec(B)\rangle=\sum_{i=1}^n \alpha_i{}^H\beta_i=tr(A^HB)$
- 范数：K 是一个数域，$K^{m\times n}$ 是 $m\times n$ 矩阵集合，则矩阵的范数 $\|A\|$ 除了要满足范数公理外，还需满足：
  - $\|AB\|\le\|A\|\|B\|$
- 诱导范数、算子范数：$\|A\|\triangleq\max\{\|Ax\|:x\in K^n,\|x\|=1\}=\max\{\frac{\|Ax\|}{\|x\|}:x\in K^n,x\ne 0\}$
  - 1 范数、绝对列和范数：$\|A\|_1=\max\limits_{1\le j\le n}\sum_{i=1}^m|a_{ij}|$
  - 无穷范数、绝对行和范数：$\|A\|_\infty=\max\limits_{1\le i\le m}\sum_{j=1}^n|a_{ij}|$
  - 2 范数、谱范数：$\|A\|_{spec}=\|A\|_2=\sigma_1$（最大奇异值）

- 1 范数证明:

$$\|Ax\|_1=\sum_{i=1}^{m}\left|\sum_{j=1}^{n}a_{ij}x_j\right|\le\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}||x_j|=\sum_{j=1}^{n}|x_j|\sum_{i=1}^{m}|a_{ij}|\le\left(\max_j\sum_{i=1}^{m}|a_{ij}|\right)\sum_{j=1}^{n}|x_j|$$

$$\therefore\|A\|_1=\max_{x\ne0}\frac{\|Ax\|_1}{\|x\|_1}\le\max_j\sum_{i=1}^{m}|a_{ij}|$$

又取 k 使 $\sum_{i=1}^{m}|a_{ik}|=\max_j\sum_{i=1}^{m}|a_{ij}|$，取 $x_0=e_k=(0,\cdots,0,1,0,\cdots,0)^T$

则有 $\|A\|_1=\max_{x\ne0}\frac{\|Ax\|_1}{\|x\|_1}\ge\|Ax_0\|_1=\sum_{i=1}^{m}|a_{ik}|=\max_j\sum_{i=1}^{m}|a_{ij}|$，证毕。

无穷范数证明:

$$\|Ax\|_\infty=\max_i\left|\sum_{j=1}^{n}a_{ij}x_j\right|\le\max_i\sum_{j=1}^{n}|a_{ij}||x_j|\le\max_i\sum_{j=1}^{n}|a_{ij}|\max_j|x_j|$$

$$\therefore\|A\|_\infty=\max_{x\ne0}\frac{\|Ax\|_\infty}{\|x\|_\infty}\le\max_i\sum_{j=1}^{n}|a_{ij}|$$

又取 k 使 $\sum_{j=1}^{n}|a_{kj}|=\max_i\sum_{j=1}^{n}|a_{ij}|$，取 $x_0=(\mathrm{sgn}(a_{k1}),\cdots,\mathrm{sgn}(a_{kn}))^T$

则有 $\|A\|_\infty=\max_{x\ne0}\frac{\|Ax\|_\infty}{\|x\|_\infty}\ge\|Ax_0\|_\infty=\sum_{j=1}^{n}|a_{kj}|=\max_i\sum_{j=1}^{n}|a_{ij}|$，证毕。

- 元素形式的 p 范数: $\|A\|_p\triangleq\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^p\right)^{1/p}=\|vec(A)\|_p$

伪逆矩阵

- 定义：对 $A\in\mathbb{C}^{m\times n}$，满足 $LA=E$ 但不满足 $AL=E$ 时，L 称为左逆矩阵；满足 $AR=E$ 但不满足 $RA=E$ 时，R 称为右逆矩阵。
  - A 列满秩时， $L=(A^HA)^{-1}A^H$ 且唯一，称为左伪逆矩阵
  - A 行满秩时， $R=A^H(AA^H)^{-1}$ 且唯一，称为右伪逆矩阵

Moore-penrose 逆矩阵

- 定义：设 $A\in\mathbb{C}^{m\times n}$，若 $\exists G\in\mathbb{C}^{n\times m}$，满足:
  - $AGA=A$

- $GAG = G$
- $(AG)^H = AG$
- $(GA)^H = GA$

称 G 是 A 的 M-P 广义逆，记作 $A^\dagger$

- 性质：
  - 唯一性
  - $A^{H\dagger} = A^{\dagger H}$
  - $(A^\dagger)^\dagger = A$
  - $(cA)^\dagger = \frac{1}{c} A^\dagger$
  - ${O_{m\times n}}^\dagger = O_{n\times m}$
  - 若 $A = BC$，B 列满秩，C 行满秩，则有 $A^\dagger = (BC)^\dagger = C^H (CC^H)^{-1} (B^H B)^{-1} B^H$
- 计算：
  - 矩阵的满秩分解：A 可以分解为 $A = F_{m\times r} G_{r\times n}$，将 A 化为简化的阶梯型，取主元列对应的原矩阵的列构成 F，主元行构成 G。

向量化与矩阵化

- 向量化：矩阵 $A \in \mathbb{R}^{m\times n}$ 的向量化是一个线性变换，将矩阵 A 按列堆栈排列成 $mn \times 1$ 矩阵，记为 $vec(A)$；按行堆栈排列成 $1\times mn$ 矩阵，记为 $rvec(A)$
- 存在唯一 $mn \times mn$ 交换阵满足 $K_{mn} vec(A) = vec(A^T)$
- 矩阵化

*稀疏表示与压缩感知

矩阵分解

Cholesky 分解：

- 定义：任意一个正定矩阵 A 都可以分解为 $A = KK^T$，其中 K 是对角线元素都大于 0 的上三角矩阵

- 计算：待定系数法，通过一定顺序可得递推式 $k_{ij}=\begin{cases}\sqrt{a_{ij}-\sum_{r=j+1}^{n}k_{jr}^{\ 2}}, j=i\\ \frac{1}{k_{jj}}\left(a_{ji}-\sum_{r=j+1}^{n}k_{jr}k_{ir}\right), j>i\\ 0, j<i\end{cases}$

奇异值分解：

- 奇异值：$A\in\mathbb{C}_r^{m\times n}$，$A^TA$ 的特征值为 $\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_r>\lambda_{r+1}>\cdots>\lambda_n=0$，则称 $\sigma_i=\sqrt{\lambda_i}$ 为 A 的奇异值。
  - A^TA 或 AA^T 产生的非零奇异值相同，零奇异值个数不同：$A^TA\alpha=\lambda\alpha, AA^T(A\alpha)=\lambda(A\alpha)$
- 奇异值分解（SVD）：$A\in\mathbb{R}_r^{m\times n}$，则有 $A=USV^T$，其中 U 是 m 阶正交阵，V 是 n 阶正交阵，$S_{m\times n}=\begin{pmatrix}\Sigma & 0\\ 0 & 0\end{pmatrix},\Sigma=diag(\sigma_1,\sigma_2,\cdots,\sigma_r)$，且 V 是 A^TA 特征值分解中的 P，U 是 AA^T 特征值分解中的 P
  - 民间简约形式：$A=U_{m\times r}\Sigma_r V_{n\times r}{}^T$，U、V 分别是 AA^T、A^TA 的非零特征值对应的特征向量形成的矩阵。且 $U_{m\times r}=A_{m\times n}V_{n\times r}\Sigma^{-1}$。
  - 定理证明（不完善）：

    对 A^TA 特征值分解：$V^TA^TAV=\Lambda=\begin{pmatrix}\Sigma^2 & 0\\ 0 & 0\end{pmatrix}$，由奇异值的定义，后一个等号成立

    令 $V=(V_1,V_2),V_1:n\times r,V_2:n\times(n-r)$

    代入得 $V^TA^TAV=\begin{pmatrix}V_1^TA^TAV_1 & V_1^TA^TAV_2\\ V_2{}^TA^TAV_1 & V_2{}^TA^TAV_2\end{pmatrix}$

    可以验证分块矩阵维数相同，比较得 $V_1^TA^TAV_1=\Sigma^2$

    左右各乘 Σ^{-1}，即 $(AV_1\Sigma^{-1})^T(AV_1\Sigma^{-1})=I$

    令 $U_1=AV_1\Sigma^{-1}$，即 $U_1{}^TU_1=I$

    令 $U=(U_1,U_2),U_2:m\times(m-r)$

则$USV^T=(U_1,U_2)\begin{pmatrix}\Sigma & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}V_1^T\\ V_2^T\end{pmatrix}=U_1\Sigma V_1^T=AV\Sigma^{-1}\Sigma V_1^T=\ldots A$

其中U_2可由下述方法求得：$U_1^T x=0$

- SVD 分解求广义逆：

$A=USV^T=U\begin{pmatrix}\Sigma & 0\\ 0 & 0\end{pmatrix}_{m\times n}V^T$，则$A^\dagger=VS^\dagger U^T=V\begin{pmatrix}\Sigma^{-1} & 0\\ 0 & 0\end{pmatrix}_{n\times m}U^T$

最小二乘问题：

- 定义：求 x，使$\|Ax-b\|_2=\min\{\|Av-b\|,v\in\mathbb{R}^n\}$，所有满足条件的 x 的集合记为$X_{LS}$，

  最小范数解x_{LS}满足$\|x_{LS}\|_2=\min\{\|x\|_2,x\in X_{LS}\}$

- $x\in X_{LS}\Leftrightarrow A^TAx=A^Tb$
- 解的形式：
  - 一般解$x=A^\dagger b+(I-A^\dagger A)z,z\in\mathbb{R}^n$
  - 最小范数解：$x_{LS}=A^\dagger b$
- 求解方法：A 列满秩，即$r(A)=n$时，称为满秩最小二乘问题，有唯一解。$r(A)<n$时，称为亏秩最小二乘问题，有无穷解，这时一般要求最小范数解。
  - $r(A)=n$时，Cholesky 分解：$A^TA=CC^T$，解$Cy=A^Tb,C^Tx=y$
  - $r(A)=n$时，QR 分解：$A=Q\begin{pmatrix}R\\ 0\end{pmatrix}=Q_1R$，解$Rx=Q_1^Tb$
  - $r(A)=n$时，SVD 分解：$A=U\begin{pmatrix}\Sigma_n\\ 0\end{pmatrix}V^T$，于是$A^\dagger=VS^\dagger U^T=V\left(\Sigma^{-1},0\right)U^T$

    $x=A^\dagger b=V\left(\Sigma^{-1},0\right)U^Tb=\sum_{i=1}^{n}\frac{u_i^Tb}{\sigma_i}v_i$

  - $r(A)<n$时，SVD 分解：$A=U\begin{pmatrix}\Sigma & 0\\ 0 & 0\end{pmatrix}V^T$则可以直接得到最小范数解或一般解

    $x_{LS}=A^\dagger b=V\begin{pmatrix}\Sigma^{-1} & 0\\ 0 & 0\end{pmatrix}U^Tb=\sum_{i=1}^{r}\frac{u_i^Tb}{\sigma_i}v_i$

函数矩阵：若矩阵 A 的元素 a_{ij} 均为实变量 t 的函数，即 $A(t)=(a_{ij}(t))_{m\times n}$

- 导数：若矩阵 A 中每个元素在 $t=t_0$ 处可导，则 A 在 t_0 处可导，记为 $\frac{dA(t)}{dt}$
- 性质：
  - $A(t),B(t)$ 可导，则 $\frac{d}{dt}(A(t)\pm B(t))=\frac{dA(t)}{dt}\pm\frac{dB(t)}{dt}$
  - $\varphi(t)$ 是可导实值函数， $A(t)$ 可导，则 $\frac{d}{dt}(\varphi(t)A(t))=\frac{d\varphi(t)}{dt}A(t)+\frac{dA(t)}{dt}\varphi(t)$
  - $A(t),B(t)$ 可导可乘，则 $\frac{d}{dt}(A(t)B(t))=\frac{dA(t)}{dt}B(t)+A(t)\frac{dB(t)}{dt}$
  - $A(t)$ 可导可逆，则 $\frac{dA^{-1}(t)}{dt}=-A^{-1}\frac{dA(t)}{dt}A^{-1}$
  - 链式法则
  - $\frac{d}{dt}(trA)=tr\frac{dA}{dt}$
- 函数关于向量的导数：
  - 定义： $x=(x_1,\cdots x_n)^T$ ，n 元函数 $g=g(x_1,\cdots,x_n)$ ，则 $\frac{dg}{dx}=(\frac{\partial g}{\partial x_1},\cdots,\frac{\partial g}{\partial x_n})^T$ ，

    $\frac{dg}{dx^T}=(\frac{\partial g}{\partial x_1},\cdots,\frac{\partial g}{\partial x_n})$ ，称上式为 g 关于 x 的梯度
- 向量映射关于向量的导数：
  - 定义： $x=(x_1,\cdots x_n)^T$ ， $f(x)=(f_1(x),\cdots f_m(x))^T$ ，则

$$\frac{df(x)}{dx^T}=\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial f_m}{\partial x_1} & \cdots & & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

- 矩阵对矩阵求导：

- 定义： $X=(x_{ij})_{m\times n}$， $F(X)=(f_{ij}(X))_{r\times s}$，则 $\frac{dF}{dX}=\begin{pmatrix} \frac{\partial F}{\partial x_{11}} & \frac{\partial F}{\partial x_{12}} & \cdots & \frac{\partial F}{\partial x_{1n}} \\ \frac{\partial F}{\partial x_{21}} & \frac{\partial F}{\partial x_{22}} & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial F}{\partial x_{m1}} & \cdots & & \frac{\partial F}{\partial x_{mn}} \end{pmatrix}$，

  其中 $\frac{\partial F}{\partial x_{ij}}=\begin{pmatrix} \frac{\partial f_{11}}{\partial x_{ij}} & \frac{\partial f_{12}}{\partial x_{ij}} & \cdots & \frac{\partial f_{1s}}{\partial x_{ij}} \\ \frac{\partial f_{21}}{\partial x_{ij}} & \frac{\partial f_{22}}{\partial x_{ij}} & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial f_{r1}}{\partial x_{ij}} & \cdots & & \frac{\partial f_{rs}}{\partial x_{ij}} \end{pmatrix}$

- 常用公式：
  - $\frac{dx^T}{dx}=\frac{dx}{dx^T}=I$
  - $\frac{dx^T c}{dx}=\frac{dc^T x}{dx}=c$
  - $\frac{dx^T Ay}{dx}=\frac{dx^T}{dx}Ay=Ay$
  - $\frac{dy^T Ax}{dx}=\frac{dx^T A^T y}{dx}=A^T y$
  - $\frac{dx^T A}{dx}=A$
  - $\frac{dx^T Ax}{dx}=(A+A^T)x$
- 行列式的导数：
  - $\frac{d|A|}{dA}=(A_{ij})_{n\times n}=A^{*T}$
  - $\frac{d|A^{-1}|}{dA}=\frac{-\frac{d|A|}{dA}}{|A|^2}=-\frac{A^{-T}}{|A|}$

- 迹的导数:
  - $\frac{d}{dA}tr(AB^T)=\frac{d}{dA}tr(B^T A)=B$，其中 $A_{m\times n},B_{m\times n}$
  - $\frac{d}{dA}tr(A)=I_n$
  - $\frac{d}{dA}tr(A^T A)=\frac{d}{dA}tr(AA^T)=2A$
  - $\frac{d}{dA}tr(A^2)=2A^T$
  - $\frac{d}{dA}tr(ABA^T)=A(B+B^T)$
  - $\frac{d}{dA}tr(A^T BA)=(B+B^T)A$
  - $\frac{d}{dA}tr(BA^{-1})=-(A^{-1}BA^{-1})^T$
  - $\frac{d(x^T Ax)}{dA}=\frac{dtr(x^T Ax)}{dA}=\frac{dtr(Axx^T)}{dA}=xx^T$
- 矩阵直和: $A\oplus B=\begin{pmatrix} A & O \\ O & B \end{pmatrix}$
- Hadamard 积 $A*B$，$(A*B)_{ij}=a_{ij}b_{ij}$
- 右 Kronecker 积: $A_{m\times n},B_{p\times q}$，$A\otimes B=(a_{ij}B)_{m\times n}$