3.5 正规矩阵

定理 （Schur 定理）(复习第一章)

任一复方阵均可酉相似于上三角矩阵.

证明：（只需把第一章中的证明所用的基底取成标准正交基即可.）

证法二：已知对任意 $A \in \mathbb{C}^{n\times n}$，存在非奇异矩阵 $P \in \mathbb{C}^{n\times n}$ 使得

$$P^{-1}AP = J$$

为上三角矩阵. 现对 P 的列向量组实施 Gram-schmidt 标准正交化，即对 P 作正交-三角分解如下

$$P = UT,$$

其中 U 为酉矩阵，T 是(正实数对角线的)上三角矩阵. 于是

Review.

Schur 定理:

Gram-Schmidt 正交化: $A = QR$, Q 是正交的, R 是上三角阵.

$A \in \mathbb{C}^{n\times n}$

$$AP = P\begin{bmatrix} \lambda_1 & \cdots & * \\ & \ddots & \vdots \\ 0 & & \lambda_2 \end{bmatrix}$$

$$P = QR$$

$$AQR = QR\begin{bmatrix} \lambda_1 & \cdots & * \\ \vdots & \ddots & \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

$$AQ = Q\boxed{R\begin{bmatrix} \lambda_1 & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}R^{-1}}$$

R 是上三角 $\Rightarrow$ R^{-1} 是上三角 (证明? 考题)

矩阵可以用一组基(特征向量)相似变换成对角阵

$\Rightarrow$ 如何用酉矩阵相似变换

定义（正规矩阵）一个复方阵若与它的共轭转置矩阵可交换，则称为正规矩阵．即 $A\in\mathbb{C}^{n\times n}$ 称为正规矩阵，若

$$AA^{\mathrm{H}} = A^{\mathrm{H}}A\,.$$

例（正规矩阵的典型例子）

Hermit 矩阵（对称矩阵）；反 Hermit 矩阵（反对称矩阵）；酉矩阵（正交矩阵）

矩阵正规性的酉相似不变性：设 A 是正规矩阵，且 A 酉相似于 B，则 B 也是正规矩阵．

证：设 $B=U^HAU$，且 $AA^H=A^HA$，易得 $BB^H=B^HB$

上三角的正规矩阵是对角矩阵：上三角矩阵若还是正规矩阵，则必为对角矩阵．

证 $R^HR = \begin{bmatrix} \ddots & 0 \\ * & \ddots \end{bmatrix}\begin{bmatrix} \ddots & * \\ 0 & \ddots \end{bmatrix}$，(1,1) 位置．

$|a_{11}|^2=|a_{11}|^2+|a_{12}|^2+\cdots+|a_{1n}|^2$，∴ 后面 $n-1$ 个均为 0

定理 $A\in\mathbb{C}^{n\times n}$ 可酉相似对角化，当且仅当 A 是正规矩阵．

回忆，此定理等价于说，n 阶复方阵有 n 个相互正交的特征向量的充要条件，是该矩阵为正规矩阵.

定理（三种特殊正规矩阵的特征值）设 A 是正规矩阵，则

（1） A 是 Hermit 矩阵的充要条件是 A 的特征值是实数.

（2） A 是反 Hermit 矩阵的充要条件是 A 的特征值的实部为零.

（2） A 是酉矩阵的充要条件是 A 的特征值的模等于 1.

正规矩阵酉相似对角化的具体求法

设 A 是 n 阶正规矩阵. 由上定理知

（1） A 的代数重数等于 A 的几何重数. 也就是说, 特征根 c 是特征多项式 $|\lambda I_n - A|$ 的 r 重根, 等价于齐次线性方程组

$$(cI_n - A)x = 0$$

的解空间 $\ker(cI_n - A)$（称为 c 的特征子空间）的维数是 r； 或者说相应于 r 重特征根 c， 恰有 r 个线性无关的特征向量. 更简洁地

$$r = \dim(\ker(cI_n - A)) = n - \operatorname{rank}(cI_n - A).$$

（2） 相应于不同特征值的特征子空间相互正交. 设 λ_1， λ_2 是两个不同的特征值, 则

$$\ker(\lambda_1 I_n - A) \perp \ker(\lambda_2 I_n - A).$$

问题： 已知 n 阶 A Λ U

$U^{\mathrm{H}} A U = \Lambda$

$|\lambda I_n - A|$ $\lambda_1, \ldots, \lambda_p$

λ_i $(\lambda_i I_n - A)x = 0$

$v_1^i, \ldots, v_{r_i}^i$

$v_1^i, \ldots, v_{r_i}^i$

$u_1^i, \ldots, u_{r_i}^i$

Λ U

$$\Lambda = \mathrm{diag}\{\lambda_i I_{r_i}, i = 1, \ldots, p\}$$

$$U = \begin{bmatrix} v_1^1 & \cdots & v_{r_1}^1 & \cdots & v_1^i & \cdots & v_{r_i}^i & \cdots & v_1^p & \cdots & v_{r_p}^p \end{bmatrix}$$

3.6 Hermit 矩阵与 Hermit 二次型

定理（Hermit 矩阵的内积刻画） 设 $A \in \mathbb{C}^{n\times n}$. 则下列条件等价

（1） $A^{\mathrm{H}} = A$， 即 A 为 Hermit 矩阵.

（2） 对任意的 $x, y \in \mathbb{C}^n$， $\langle Ax, y\rangle = \langle x, Ay\rangle$.

定义（Hermit 二次型）

给定 n 阶 Hermit 矩阵 A. 映射 $x \mapsto \langle Ax, x\rangle = x^{\mathrm{H}} Ax$ 确定从 $\mathbb{C}^n$ 到 $\mathbb{R}$ 的函数， 称为 Hermit 二次型. 写成分量形式， 有

$$x^{\mathrm{H}} Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\bar{x}_i x_j .$$

定义（Hermit 矩阵的合同）

设 A 和 B 是两个 n 阶 Hermit 矩阵，P 是 n 阶非奇异矩阵．若有

$$P^{\mathrm{H}}AP = B,$$

则称 A 和 B 关于 P 合同．此时也简称 A 与 B 合同，并称 P 为相应的合同矩阵或合同变换矩阵．

注（Hermit 矩阵的酉相似也是合同）

注（Hermit 二次型在坐标变换下的动态） Hermit 二次型与坐标变换两个概念的结合，导致 Hermit 矩阵的合同概念．向量 $x \in \mathbb{C}^n$ 在 $\mathbb{C}^n$ 的标准基 I_n 下的坐标就是 x 自己：$x = I_n x$．它在 $\mathbb{C}^n$ 的新的基底 P 下的坐标 $\tilde{x}$ 满足：$x = P\tilde{x}$．于是

$$x^{\mathrm{H}}Ax = (P\tilde{x})^{\mathrm{H}} A(P\tilde{x}) = \tilde{x}^{\mathrm{H}}\underbrace{\left(P^{\mathrm{H}}AP\right)}_{=:B}\tilde{x}.$$

由 Hermit 是正规矩阵，可找到酉矩阵 U，将 Hermit 二次型 $x^{\mathrm{H}}Ax$ 通过正交坐标变换 $x=U\tilde{x}$ 化为“平方和”形式

$$
\begin{aligned}
x^{\mathrm{H}}Ax &= \tilde{x}^{\mathrm{H}}(U^{\mathrm{H}}AU)\tilde{x} \\
&= \lambda_1\bar{x}_1x_1+\lambda_2\bar{x}_2x_2+\cdots+\lambda_n\bar{x}_nx_n\,, \\
&= \lambda_1|x_1|^2+\lambda_2|x_2|^2+\cdots+\lambda_n|x_n|^2
\end{aligned}
$$

这里，$\lambda_1,\lambda_2,\ldots,\lambda_n$ 是 A 的 n 个特征值.

Hermit 正定矩阵与 Hermit 正定二次型

定义（正定矩阵与正定二次型）

给定 Hermit 矩阵 A，如果对任意 $x \in \mathbb{C}^n$，$x \neq 0$，有 $x^{\mathrm{H}} A x > 0$ (≥ 0)，则称 $x^{\mathrm{H}} A x$ 为 Hermit 正定(半正定)二次型，A 为 Hermit 正定(半正定)矩阵.

这里，我们只补充大学阶段通常的线性代数课程未包含的内容.

定理（正定性的 Sylvester 行列式判别法）n 阶 Hermit 矩阵 A 是正定的，当且仅当它的 n 个顺序主子式都大于零.

注（半正定性的 Sylvester 行列式判别法）

Hermit 矩阵 A 是半正定的，当且仅当它的所有各阶主子式都非负.

注（正定性的判别不须求特征根）
虽然 Hermit 矩阵正定的充要条件是它的特征值都是正数，但 Sylvester 行列式判别法说明正定性的判定不需求特征多项式的根，只需计算矩阵的行列式.

定理（正定矩阵开平方）
设矩阵 A 是 Hermit 正定矩阵. 则存在唯一的 Hermit 正定矩阵 B 使得

$$A = B^2 .$$

定理（Gram 矩阵与 Hermit 半正定矩阵）
Hermit 矩阵是半正定的充要条件是它可以写成一个复内积空间中某个向量组的 Gram 矩阵.

Hermit 矩阵特征值的极值刻画

事实：Hermit 矩阵 A 的 n 个特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 都是实数，不妨设

$$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n .$$

又 $\langle Ax, x\rangle = x^{\mathrm{H}} Ax$ 总是实数.

定理（Hermit 矩阵特征值的极值刻画）

$$\lambda_{\min}(A) = \lambda_1 = \min_{x\in\mathbb{C}^n, x\neq 0} \frac{x^{\mathrm{H}} Ax}{x^{\mathrm{H}} x} = \min_{x\in\mathbb{C}^n, \|x\|=1} x^{\mathrm{H}} Ax ,$$

$$\lambda_{\max}(A) = \lambda_n = \max_{x\in\mathbb{C}^n, x\neq 0} \frac{x^{\mathrm{H}} Ax}{x^{\mathrm{H}} x} = \max_{x\in\mathbb{C}^n, \|x\|=1} x^{\mathrm{H}} Ax .$$

解 $x^H A x$ 的最值.

①

$$f(x_1\ x_2 \cdots x_n)$$

$$= [x_1\ x_2 \cdots x_n]\ A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \sum_{i=1}^{n} a_{ii} x_i + 2\sum_{i<j} a_{ij} x_i x_j \qquad s.t.\ x_1^2 + x_2^2 + \cdots x_n^2 = 1$$

用拉格朗日乘数法求解(思考)

② A是Hermite阵，则∃U阵，使 $U^H A U = \Lambda$，记 $xU = \tilde{x}$

即 $x^H A x \xlongequal{U} (U\tilde{x})^H A (U\tilde{x})$

$$= \tilde{x}^H [U^H A U] \tilde{x}$$

$$= \lambda_1 \bar{\tilde{x}}_1 \tilde{x}_1 + \lambda_2 \bar{\tilde{x}}_2 \tilde{x}_2 + \cdots + \lambda_n \bar{\tilde{x}}_n \tilde{x}_n$$，记为 L，λ按升序排列

则 $\lambda_1 \sum_{i=1}^{n} \bar{x}_i x_i \le L \le \lambda_n \sum_{i=1}^{n} \bar{x}_i x_i$，由于 $\bar{x}_i x_i = \|x_i\|_2^2$

$\therefore \sum_{i=1}^{n} \bar{x}_i x_i = \sum_{i=1}^{n} \|x_i\|_2^2 = 1$ （酉变换保长度）

$\therefore \lambda_{max} = \lambda_n,\quad \lambda_{min} = \lambda_1$.

界是可以取到的，只要令 x_i 其它分量为0，如 $x = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$ 可取到 λ_n

注 (Hermit 矩阵特征值与 Rayleigh 商)

从 $\mathbb{C}^n \setminus \{0\}$ 到 $\mathbb{R}$ 的函数

$$x \mapsto R(x) = \frac{x^{\mathrm{H}} A x}{x^{\mathrm{H}} x}$$

称为Hermit矩阵 A 的Rayleigh商. A 的其他特征值 $\lambda_2, \ldots, \lambda_{n-1}$ 也可通过 A 的 Rayleigh 商用极值来刻画. 对此我们不再详述.

注 (极值刻画将特征多项式求根与二次型求极值联系起来)

矩阵的奇异值分解

引理（标准酉空间中的 Gram 矩阵）

设 $A \in \mathbb{C}^{m\times n}$， $\mathrm{rank}(A) = r$． 则

A^T的···列···

（1） $A^{\mathrm{H}} A$ 和 AA^{H} 分别是 n 阶和 m 阶 Hermit 半正定矩阵.

A的列向量组的Gram阵

（2） $\mathrm{rank}(A^{\mathrm{H}} A) = \mathrm{rank}(AA^{\mathrm{H}}) = r$.

证(1) 用二次型：$x^H A^H A x = \langle Ax, Ax \rangle \geq 0$

证(2) 因为齐次方程组 $Ax=0$ 解空间维数为 $n-r$

首先证：$Ax=0 \Leftrightarrow A^H A x = 0$

$\Rightarrow$ 显然

$\Leftarrow$ $x^H A^H A x = x^H \cdot 0 = 0$，$\therefore (Ax)^H (Ax) = 0 \Rightarrow$ 内积，仅当 $Ax=0$ 为 0

则 $Ax=0$ 与 $A^H A x = 0$ 的解空间一致，

$\therefore n - r(A) = n - r(A^H A)$

定理 （矩阵的奇异值分解）

设 $A \in \mathbb{C}^{m\times n}$． 则存在 m 阶酉矩阵 U 和 n 阶酉矩阵 V， 使得

$$U^{\mathrm{H}} A V = \begin{bmatrix} \Sigma_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix},$$

$\Rightarrow AV = U\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}$

入口基！ 出口基！

其中 $r = \mathrm{rank}(A)$， r 阶对角矩阵

$$\Sigma_r = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}, \quad \sigma_i > 0, \quad i = 1,\ldots,r.$$

（标准正交基）

推论 $A^{\mathrm{H}}A$ 与 AA^{H} 有相同的非零（正实数）特征值

$$\lambda_i = \lambda_i(A^{\mathrm{H}}A) = \lambda_i(AA^{\mathrm{H}}), \quad i = 1,\ldots,r.$$

且

$$\sigma_i = \sqrt{\lambda_i} > 0, \quad i = 1,\ldots,r.$$

Review.

$A \in \mathbb{C}^{m\times n}$，则 A 是从 n 维向量到 m

矩阵等价： $AP = QB$， P、Q 非奇异

若 PQ 可任取，则最简形可以得到 $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$

这里 P、Q 是标准正交的，可以得到

$$AP = Q\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}, \quad \Sigma_r = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$$

证明：

A 的 Gram 阵 A^HA 是 n 阶半正定的（Hermite 阵）

由 Gram 定理，存在标准正交阵 V，使 $V^HA^HAV = \begin{bmatrix} \lambda_1 & & & & & \\ & \lambda_2 & & & & \\ & & \ddots & & & \\ & & & \lambda_r & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}$

分块，得 $(AV)^H(AV)$

$$= (A[V_1\ V_2])^H(A[V_1\ V_2])$$

$$= \begin{bmatrix} (AV_1)^H \\ (AV_2)^H \end{bmatrix}[AV_1\ AV_2] = \begin{bmatrix} (AV_1)^H(AV_1) & (AV_1)^H(AV_2) \\ (AV_2)^H(AV_1) & (AV_2)^H(AV_2) \end{bmatrix} = \left[\begin{array}{c|c} \begin{matrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{matrix} & 0 \\ \hline 0 & 0 \end{array}\right]$$

① $(AV_1)^H(AV_1) = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{bmatrix}$，$\Rightarrow$ AV_1 的列向量组是正交向量组.

$AV_1 = \{Av_1\ Av_2 \cdots Av_r\}$ 是正交向量组.　　前 r 个列向量的像仍正交.

$\|Av_1\| = \sqrt{\lambda_1} = \sigma_1$， $\cdots$ $\|Av_r\| = \sqrt{\lambda_r} = \sigma_r$

单位化 $\{Av_1 \cdot \frac{1}{\sigma_1}, Av_2 \frac{1}{\sigma_2}, \cdots Av_r \cdot \frac{1}{\sigma_r}\}$，记 U_1，是 $m\times r$ 的，扩充成 $[U_1\ U_2]^{m\times m}$

② $(AV_2)^H(AV_2)=\vec{0} \Rightarrow AV_2=0$. 后 n-r 个向量构成核

注（奇异值分解的几何意义）

由 $A \in \mathbb{C}^{m\times n}$ 决定的从 $\mathbb{C}^n$ 到 $\mathbb{C}^m$ 的线性映射 $x \mapsto y = Ax$，在 $\mathbb{C}^n$ 中的标准正交基 V（入口基）和 $\mathbb{C}^m$ 中的标准正交基 U（出口基）下的矩阵表示为

$$\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}.$$

以下讨论奇异值的几何意义和极值刻画.

定义 （高维空间中的椭球面）（复的也完全可以!）

设 S 是 n 阶 Hermit 正定矩阵， c 是大于 0 的实数. 则 $\mathbb{C}^n$ 中的点集

$$\left\{x \mid x \in \mathbb{C}^n, x^{\mathrm{H}} S x = c\right\}$$

称为 $\mathbb{C}^n$ 中的一个椭球面.

$$\left\{x \mid x \in \mathbb{C}^n, x^{\mathrm{H}} S x \le c\right\}$$

称为椭球体.

定理 （线性映射把椭球映射成椭球）

设 A 是 n 阶非奇异矩阵. 则任一椭球面在映射 $x \mapsto y = Ax$ 下的像也是椭球.

设 A 是 n 阶非奇异矩阵，记 A 的奇异值分解为

$$U^{\mathrm{H}}AV=\Sigma=\begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_n\end{bmatrix}.$$

现计算单位球

$\|x\|=1 \Downarrow |x_1|^2+|x_2|^2+\cdots+|x_n|^2=1$

$$\{x \mid x\in\mathbb{C}^n, \|x\|=1\}$$

单位球

$\downarrow A$

椭球

$\|A^{-1}y\|=1$

在 A 下的像. 由

$$\|x\|^2=\|A^{-1}y\|^2=\|(U\Sigma V^{\mathrm{H}})^{-1}y\|^2$$

$$=\|V\Sigma^{-1}(U^{\mathrm{H}}y)\|^2=\|\Sigma^{-1}\widetilde{y}\|^2$$

$$=\frac{|\widetilde{y}_1|^2}{\sigma_1^2}+\frac{|\widetilde{y}_2|^2}{\sigma_2^2}+\cdots+\frac{|\widetilde{y}_n|^2}{\sigma_n^2},$$

半轴长

像: $\|A^{-1}y\|=1$

$U^HAV=\Sigma$

$A=U\Sigma V^H$

$A^{-1}=V\Sigma^{-1}U^H$

$\|A^{-1}y\|=\|V\Sigma^{-1}U^Hy\|$

$=\|\Sigma^{-1}U^Hy\|$ (酉保长度)

记 $U^Hy=\widetilde{y}$，同时 $y=U\widetilde{y}$

这里 $\widetilde{y}=U^{\mathrm{H}}y$，即 $y=U\widetilde{y}$.

也就是说，在新的标准正交基 U 下，单位球的像的方程是标准的椭球面方程

$$\frac{|\widetilde{y}_1|^2}{\sigma_1^{\ 2}}+\frac{|\widetilde{y}_2|^2}{\sigma_2^{\ 2}}+\cdots+\frac{|\widetilde{y}_n|^2}{\sigma_n^{\ 2}}=1.$$

由于

$$U\Sigma=AV,\quad U=AV\Sigma^{-1}=\begin{bmatrix}Av_1\frac{1}{\sigma_1} & Av_2\frac{1}{\sigma_2} & \cdots & Av_n\frac{1}{\sigma_n}\end{bmatrix},$$

可知，

$$\sigma_i=\|Av_i\|,\quad i=1,2,\ldots,n.$$

换言之，标准正交出口基 $U=\begin{bmatrix}u_1 & u_2 & \cdots & u_n\end{bmatrix}$ 就是标准正交入口基 $V=\begin{bmatrix}v_1 & v_2 & \cdots & v_n\end{bmatrix}$ 在 A 下的像的单位化.

因此，若 σ_1 是最大奇异值，则 v_1 就是单位球面上被 A 映的最远的向量：映射成到椭球面最长的半轴 $Av_1=u_1\sigma_1$.

特征值 $\lambda_i = \lambda_i(A^HA) > 0$, $i=1,2,\cdots,r$, 奇异值 $\sigma_i = \sqrt{\lambda_i(A^HA)}$

定义大小顺序 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r$, 根据特征值的极值刻画,

最大奇异值 $\sigma_1 = \sqrt{\lambda_1} = \sqrt{\max_{x\neq 0} \frac{x^HA^HAx}{x^Hx}} = \sqrt{\max_{x\neq 0} \frac{\|Ax\|^2}{\|x\|^2}} = \max_{x\neq 0} \frac{\|Ax\|}{\|x\|}$

几何解释: $\frac{\text{像的长度}}{\text{原像的长度}}$ ⇒ 最大放大率

Summary:

入口基和出口基的找法(必考).