矩阵论备忘录

杜磊

大连理工大学 数学科学学院

2022 年 10 月 12 日

定义 (正规矩阵)

设 $A \in \mathbb{C}^{n\times n}$, 若 $A^HA = AA^H$, 则称 A 为正规矩阵.

定理 (谱分解定理)

每个正规矩阵都酉相似于一个对角矩阵, 即若 $A \in \mathbb{C}^{n\times n}$ 正规, 则存在酉矩阵 U 满足

$$A = U\mathrm{diag}(\lambda_1, \cdots, \lambda_n)U^H.$$

证明.

利用 Schur 分解. □

特殊矩阵

定义 (Toeplitz 矩阵)

$T \in \mathbb{C}^{n\times n}$ 称为 Toeplitz 矩阵，若存在数 $t_{-n+1}, \cdots, t_{-1}, a_0, a_1, \cdots, t_{n-1}$ 使得 $t_{ij} = t_{j-i}$. Toeplitz 矩阵有如下形式

$$T = \begin{bmatrix} t_0 & t_1 & t_2 & \cdots & t_{n-1} \\ t_{-1} & t_0 & t_1 & \cdots & t_{n-2} \\ t_{-2} & t_{-1} & t_0 & \cdots & t_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{-n+1} & t_{-n+2} & t_{-n+3} & \cdots & t_0 \end{bmatrix}.$$

求解 Toeplitz 线性方程组的经典算法，见参考文献 (Matrix Computations, p. 208).

特殊矩阵

定义 (Hankel 矩阵)

$H \in \mathbb{C}^{n\times n}$ 称为 Hankel 矩阵，若存在数 $h_1, \cdots, h_{hn-1}$ 使得 $h_{ij} = h_{i+j-1}$. Hankel 矩阵有如下形式

$$H = \begin{bmatrix} h_1 & h_2 & h_3 & \cdots & h_n \\ h_2 & h_3 & h_4 & \cdots & h_{n+1} \\ h_3 & h_4 & \ddots & \cdots & h_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_n & h_{n+1} & h_{n+2} & \cdots & h_{2n-1} \end{bmatrix}.$$

Reversing the rows (or columns) of a Toeplitz matrix gives a Hankel matrix. That is, if $J = (e_n, e_{n-1}, \ldots, e_1)$, then JT (and TJ) are Hankel matrices, and vice versa.

特殊矩阵

定义 (Circulant 矩阵)

$A \in \mathbb{C}^{n\times n}$ 称为 Circulant 矩阵，若存在数 $a_1, \cdots, a_n$ 使得矩阵 A 有如下形式

$$A = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ a_n & a_1 & a_2 & \cdots & a_{n-1} \\ a_{n-1} & a_n & a_1 & \cdots & a_{n-22} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ a_2 & a_3 & a_4 & \cdots & a_1 \end{bmatrix}.$$

记 $A = \mathrm{Circ}(a_1, a_2, \cdots, a_n)$. 称置换矩阵 $P = \mathrm{Circ}(0, 1, 0, \cdots, 0)$ 为基本循环矩阵. 有

$$A = \sum_{k=0}^{n-1} a_{k+1} P^k.$$

对基本循环矩阵

$$P = \mathrm{Circ}(0,1,0,\cdots,0) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

易知 $\det(\lambda I - P) = \lambda^n - 1$, 故 P 的特征值 λ_j 和对应特征向量 x_j 分别为

$$\lambda_j = z^j, \quad j = 0, 1, \cdots, n-1, z = e^{\frac{2\pi}{n}i},$$

$$x_j = \frac{1}{\sqrt{n}}(1, z^j, \cdots, z^{(n-1)j})^T.$$

易证, $x_0, x_1, \cdots, x_{n-1}$ 为 P 的一组标准正交特征向量. 记 $U = [x_0, x_1, \cdots, x_{n-1}]$, 则 U 为酉矩阵, 且

$$P = U\mathrm{diag}(1, z, z^2, \cdots, z^{n-1})U^H.$$

记

$$f(t) = \sum_{k=0}^{n-1} a_{k+1} t^k.$$

易得循环矩阵的谱分解形式为

$$A = U \mathrm{diag}(f(1), f(z), f(z^2), \cdots, f(z^{n-1}))\, U^H.$$

Hilbert 矩阵

A famous class of ill-conditioned matrices is the Hilbert matrix $H_n(p) \in \mathbb{R}^{n\times n}$ with $p \neq -1, -2, \ldots, -(2n-1)$. A Hilbert matrix, for a given value of p, is defined as

$$(H_n(p))_{i,j} = h_{i,j}^{(p)} = (p+i+j-1)^{-1} \, \mathrm{for all} i, j = 1, 2, \ldots, n.$$

These matrices naturally arise in various situations, e.g., if we want to approximate a continuous function $f(x)$ in the interval $[0, 1]$ by a polynomial of degree $n-1$ in x such that the error

$$E = \int_0^1 \left(p_{n-1}(x) - f(x)\right)^2 dx, \quad p_{n-1}(x) = \sum_{j=1}^{n} \alpha_j x^{j-1}$$

is minimized. By differentiating E within respect to the coefficients α_j and imposing the result to be 0 gives its minimum; we obtain the following system of n equation in the α_js for $i = 1, 2, \ldots n$:

$$\sum_{j=1}^{n}\left(\int_0^1 x^{i+j-2}dx\right)\alpha_j = \left(\equiv \sum_{j=1}^{n} h_{i,j}^{(0)} =\right)\int_0^1 f(x)x^{i-1}dx \equiv b_i,$$

i.e., $H_n(0)\alpha = b$. For this particular class of matrices we can express, for nonnegative integer values of p, the inverse explicitly as

$$\left(H_n^{-1}(p)\right)_{i,j} = \frac{(-1)^{i+j}}{p+i+j-1}\left[\frac{\prod_{k=0}^{n-1}(p+i+k)(p+j+k)}{(i-1)!(n-i)!(j-1)!(n-j)!}\right].$$

见文献中具体算例: ($f(x)=\sin(2\pi x)$, $n=15$, p. 7).

Daniele Bertaccini & Fabio Durastante
Iterative Methods and Preconditioning for Large and Sparse Linear Systems with Applications, Chapman and Hall/CRC, 2017.

Cauchy 矩阵

The Hilbert matrix is a special case of a Cauchy matrix $C_n \in \mathbb{R}^{n\times n}$, whose elements are $c_{ij} = 1/\left(x_i + y_j\right)$, where $x, y \in \mathbb{R}^n$ are given n-vectors (take $x_i = y_i = i - 0.5$ for the Hilbert matrix). The following formulae give the inverse and determinant of C_n, and therefore generalize those for the Hilbert matrix. The (i, j) element of C_n^{-1} is

$$\frac{\prod_{1\le k\le n}\left(x_j + y_k\right)\left(x_k + y_i\right)}{\left(x_j + y_i\right)\prod_{\substack{1\le k\le n\\ k\ne j}}\left(x_j - x_k\right)\prod_{\substack{1\le k\le n\\ k\ne i}}\left(y_i - y_k\right)}$$

and

$$\det\left(C_n\right) = \frac{\prod_{1\le i<j\le n}\left(x_j - x_i\right)\left(y_j - y_i\right)}{\prod_{1\le i,j\le n}\left(x_i + y_j\right)}$$

the latter formula having been published by Cauchy in 1841.

The LU factors of C_n were found explicitly by Cho:

$$l_{ij} = \frac{x_j + y_j}{x_i + y_i} \prod_{k=1}^{j-1} \frac{(x_j + y_k)(x_i - x_k)}{(x_i + y_k)(x_j - x_k)}, \quad 1 \le j < i \le n,$$

$$u_{ij} = \frac{1}{x_i + y_j} \prod_{k=1}^{i-1} \frac{(x_i - x_k)(y_j - y_k)}{(x_k + y_j)(x_i + y_k)}, \quad 1 \le i \le j \le n.$$

It is known that C_n is totally positive if $0 < x_1 < \cdots < x_n$ and $0 < y_1 < \cdots < y_n$ (as is true for the Hilbert matrix). Interestingly, the sum of all the elements of C_n^{-1} is $\sum_{i=1}^{n} (x_i + y_i)$. By exploiting their structure, singular value decompositions of Cauchy matrices can be computed to high accuracy.

Kronecker Products (or tensor product)

定义

设 $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{p\times q}$. 则称 $mp \times nq$ 的矩阵

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}$$

为矩阵 A, B 的 Tensor product 或 Kronecker product.

基本性质:

1. $(A + B) \otimes C = (A \otimes C) + (B \otimes C)$,
2. $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$,
3. $A \otimes (B \otimes C) = (A \otimes B) \otimes C$,
4. $(A \otimes B)^T = A^T \otimes B^T$.

Kronecker Products (or tensor product)

引理

设 $A \in \mathbb{R}^{m\times n}, C \in \mathbb{R}^{n\times p}, B \in \mathbb{R}^{r\times s}, D \in \mathbb{R}^{s\times q}$. 则有

$$(A \otimes B)(C \otimes D) = AC \otimes BD.$$

基本性质:

1. 若 $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{p\times p}$ 可逆, 则 $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
2. 若 $A = X^{-1}\Lambda X, B = Y^{-1}\Gamma Y$, 则 $A \otimes B = (X^{-1}\Lambda X) \otimes (Y^{-1}\Gamma Y) = (X \otimes Y)^{-1}(\Lambda \otimes \Gamma)(X \otimes Y)$
3. 若 $Ax_i = \lambda_i x_i, i = 1:n, By_j = \mu_j y_j, j = 1:m$, 则

$$\begin{aligned}\Big[(I_m \otimes A) + (B \otimes I_n)\Big](y_j \otimes x_i) &= y_j \otimes (Ax_i) + (By_j) \otimes x_i \\ &= (\lambda_i + \mu_j)(y_j \otimes x_i)\end{aligned}$$

引理

设 $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{p\times q}, X \in \mathbb{R}^{q\times n}$. 则有

$$(A \otimes B)\mathrm{vec}(X) = \mathrm{vec}(BXA^T).$$

定理

设 $A\in\mathbb{C}^{m\times m}, B\in\mathbb{C}^{n\times n}$.

1. 若 A, B 可逆, 则 $A\otimes B$ 也可逆, 且 $(A\otimes B)^{-1}=A^{-1}\otimes B^{-1}$.
2. 若 A, B 为正规 (酉) 矩阵, 则 $A\otimes B$ 也为正规 (酉) 矩阵.
3. 若 $\lambda\in\lambda(A), x$ 是对应的特征向量, $\mu\in\lambda(B), y$ 是对应的特征向量, 则
$$\lambda\mu\in\lambda(A\otimes B), x\otimes y\text{是对应的特征向量}.$$
4. 若 $\lambda(A)=\{\lambda_1,\cdots,\lambda_m\}, \lambda(B)=\{\mu_1,\cdots,\mu_n\}$, 则
$$\lambda(A\otimes B)=\{\lambda_i\mu_j|i=1,\cdots,m, j=1,\cdots,n\}.$$
5. $\det(A\otimes B)=(\det A)^n(\det B)^m$.
6. 若 $\sigma(A)=\{\sigma_1,\cdots,\sigma_m\}, \sigma(B)=\{\tau_1,\cdots,\tau_n\}$, 则
$$\sigma(A\otimes B)=\{\sigma_i\tau_j|i=1,\cdots,m, j=1,\cdots,n\}.$$

引理

设 $A \in \mathbb{C}^{m\times n}, B \in \mathbb{C}^{n\times k}, C \in \mathbb{C}^{k\times t}$, 则

$$\mathrm{vec}(ABC) = (C^H \otimes A)\mathrm{vec}(B).$$

设 $A_i \in \mathbb{C}^{m\times n}, B_i \in \mathbb{C}^{s\times t}, i = 1, \cdots, k, C \in \mathbb{C}^{m\times t}$ 给定. 考虑关于未知矩阵 $X \in \mathbb{C}^{n\times s}$ 的线性方程

$$A_1XB_1 + A_2XB_2 + \cdots + A_kXB_k = C.$$

对上式两边作 vec 运算, 得到如下等价线性方程组

$$(B_1^H \otimes A_1 + B_2^H \otimes A_2 + \cdots + B_k^H \otimes A_k)\mathrm{vec}(X) = \mathrm{vec}(C).$$

定理

定理

设 $A \in \mathbb{C}^{m\times m}, B \in \mathbb{C}^{n\times n}, C \in \mathbb{C}^{m\times n}$ 给定. 关于未知矩阵 $X \in \mathbb{C}^{m\times n}$ 的方程

$$AX - XB = C \iff (I_n \otimes A - B^H \otimes I_m)\mathrm{vec}(X) = \mathrm{vec}(C)$$

称为 *Sylvester* 方程. 该方程有唯一解当且仅当 A 和 B 没有公共特征值.

证明.

设 $\lambda(A) = \{\lambda_1, \cdots, \lambda_m\}, \lambda(B) = \{\mu_1, \cdots, \mu_n\}$. 则 $\lambda(I_n \otimes A - B^H \otimes I_m) = \{\lambda_i - \mu_j | i = 1, \cdots, m, j = 1, \cdots, n\}$. $\square$

定义 (Khatri-Rao product)

设 $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{p\times n}$. 则称 $mp \times n$ 的矩阵

$$A \odot B = \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \cdots & \mathbf{a}_n \otimes \mathbf{b}_n \end{bmatrix}$$

为矩阵 A, B 的 Khatri-Rao product.

定义 (Hadamard Product)

设 $A, B \in \mathbb{R}^{m\times n}$. 则称 $m \times n$ 的矩阵

$$A * B = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1n}b_{1n} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2n}b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1}b_{m1} & a_{m2}b_{m2} & \cdots & a_{mn}b_{mn} \end{pmatrix}$$

为矩阵 A, B 的 Hadamard product.

基本性质:

1. $A \odot B \odot C = (A \odot B) \odot C = A \odot (B \odot C)$,
2. $(A \odot B)^{\mathrm{T}}(A \odot B) = A^{\mathrm{T}}A * B^{\mathrm{T}}B$,
3. $(A \odot B)^{\dagger} = ((A^{\mathrm{T}}A) * (B^{\mathrm{T}}B))^{\dagger}(A \odot B)^{\mathrm{T}}$.

向量范数: l_p 范数 or Hölder 范数

定义

$\|x\|_p := (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}, \quad 1 \le p \le \infty$. 由此定义, 如下结论成立 $\|x\|_q \le \|x\|_p \le n^{(1/p-1/q)} \|x\|_q$, $1 \le p \le q \le \infty$.

证明.

证明三角不等式成立需要用到如下不等式: 如果 $p > 1, q$ 满足 $1/p + 1/q = 1$, 则

$$\alpha\beta \le \frac{\alpha^p}{p} + \frac{\beta^q}{q}.$$

设 $x, y \in \mathbb{R}, 0 < \lambda < 1$. 由指数函数凸性知

$$e^{\lambda x + (1-\lambda)y} \le \lambda e^x + (1-\lambda)e^y.$$

令 $\lambda = 1/p, x = p\log\alpha, y = q\log\beta$ 即证上述不等式. $\square$

定理

$|x^H y| \leq \|x\|_p \|y\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1,\ p \geq 1.$

证明.

证明可参考: Theorem 3.29, p. 77. Approximation Theory and Algorithms for Data Analysis. □

Since $\|A\|_2^2 = \lambda_{\max}(A^H A)$, we obtain an upper bound for the matrix 2-norm:

$$\|A\|_2 \leq (\|A^H\|_1 \|A\|_1)^{1/2} = (\|A\|_\infty \|A\|_1)^{1/2},$$

which is cheap to evaluate.

定理

设 V 是复或实的有限维空间, $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是 V 上的范数, 则存在正数 c 和 d 满足

$$c\|x\|_\beta \le \|x\|_\alpha \le d\|x\|_\beta, \quad \forall x \in V.$$

证明.

$S = \{x \in V \mid \|x\|_\beta = 1\}$ 是个紧集. 由 Weierstrass 定理, 实值连续函数 $f(x) = \|x\|_\alpha$ 在 S 上可以取到最小值 c 和最大值 d. 由范数正定性知 $c > 0, d > 0. \forall x \in V, x \neq 0$, 有 $x/\|x\|_\beta \in S$. 从而

$$c \le \left\| \frac{x}{\|x\|_\beta} \right\|_\alpha \le d,$$

即

$$c\|x\|_\beta \le \|x\|_\alpha \le d\|x\|_\beta.$$

Schatten norm

Perhaps the most important class of unitarily invariant matrix norms are the Schatten norms

$$\|A\| = \left(\sum_{i=1}^{r} \sigma_i^p \right)^{1/p}, \quad r = \min\{m, n\}, \quad 1 \leq p < \infty.$$

For $p = 2$ we get the Frobenius norm, and letting $p \to \infty$ gives the spectral norm. A norm of increasing importance in applications is the nuclear norm (or Ky Fan's norm), which corresponds to $p = 1$.

Wedderburn & Guttman Theorems

定理 (Wedderburn, 1934)

Suppose $A \in \mathbb{R}^{m\times n}, f \in \mathbb{R}^{n\times 1}$, *and* $g \in \mathbb{R}^{m\times 1}$. *Then*

$$\operatorname{rank}\left(A - \omega^{-1} A f g^T A\right) = \operatorname{rank}(A) - 1,$$

if and only if $\omega = g^T A f \neq 0$.

定理 (Guttman, 1957)

Suppose $A \in \mathbb{R}^{m\times n}, F \in \mathbb{R}^{n\times k}$, *and* $G \in \mathbb{R}^{m\times k}$. *Then*

$$\operatorname{rank}\left(\mathrm{A} - \mathrm{AFR}^{-1}\mathrm{G}^{\mathrm{T}}\mathrm{A}\right) = \operatorname{rank}(\mathrm{A}) - \operatorname{rank}\left(\mathrm{AFR}^{-1}\mathrm{G}^{\mathrm{T}}\mathrm{A}\right), \quad (1)$$

if and only if $R = G^T A F \in \mathbb{R}^{k\times k}$ *is nonsingular.*

It is shown that several standard matrix factorizations in numerical linear algebra are instances of the Wedderburn formula: Gram-Schmidt orthogonalization, singular value decomposition, QR and Cholesky decomposition, as well as the Lanczos procedure

AB 与 BA 有相同非零特征值

定理 (Flanders, 1951)

设 $A \in \mathbb{C}^{m\times n}, B \in \mathbb{C}^{n\times m}$, 则 AB 和 BA 的非零特征值 (包括重数) 相同.

证明.

利用

$$\begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix} \text{和} \begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix}$$

相似证明. $\square$

C.R. Johnson & E.A. Schreiner
The Relationship between AB and BA, The American Mathematical Monthly, 103:7(1996), 578-582.