数值线性代数与算法

第一章 数值线性代数理论基础

本章介绍后面章节中需要用到的一些矩阵代数基础知识: 首先介绍一些基本概念和几种常用的矩阵分解; 其次介绍作为数值误差分析度量工具的向量和矩阵范数及其性质; 接着介绍矩阵的广义逆以及几种特殊的矩阵类型; 最后介绍数值代数中常用的模型问题.

§1.1 一些概念和记号

本书引用下列记号. 用 $\mathbb{C}^n\,(\mathbb{R}^n)$ 表示 n 维复 (实) 向量空间, $\mathbb{C}^{m\times n}$ $(\mathbb{R}^{m\times n})$ 表示 $m\times n$ 阶复 (实) 矩阵空间, $\mathbb{C}_r^{m\times n}$ $(\mathbb{R}_r^{m\times n})$ 表示秩为 r 的 $m\times n$ 阶复 (实) 矩阵集合. 对于任意的矩阵 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$, $\boldsymbol{A}^{\mathrm{T}},\boldsymbol{A}^{\mathrm{H}},\boldsymbol{A}^{-1}$ 分别表示矩阵 $\boldsymbol{A}$ 的转置矩阵、共轭转置矩阵和逆矩阵. 对于任意的向量 $\boldsymbol{x},\boldsymbol{y}\in\mathbb{C}^n$, 用 $(\boldsymbol{x},\boldsymbol{y}):=\boldsymbol{y}^{\mathrm{H}}\boldsymbol{x}$ 表示 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的欧几里得内积.

定义 1.1 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$. 若存在数 $\lambda \in \mathbb{C}$ 和非零向量 $\boldsymbol{x} \in \mathbb{C}^n$ 使得 $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, 则称 λ 为 $\boldsymbol{A}$ 的特征值, $\boldsymbol{x}$ 为 $\boldsymbol{A}$ 属于 λ 的特征向量.

由定义 1.1 可知, λ 是 $\boldsymbol{A}$ 的特征值当且仅当 $\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = 0$. 称 $p(\lambda) = \det(\lambda\boldsymbol{I} - \boldsymbol{A})$ 为 $\boldsymbol{A}$ 的特征多项式.

注意到 $\det(\lambda\boldsymbol{I} - \boldsymbol{A}^{\mathrm{T}}) = \det(\lambda\boldsymbol{I} - \boldsymbol{A})$, 此即 $\boldsymbol{A}$ 和 $\boldsymbol{A}^{\mathrm{T}}$ 具有相同的特征值. 故存在非零向量 $\boldsymbol{y} \in \mathbb{C}^n$ 满足 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} = \lambda\boldsymbol{y}$, 即 $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A} = \lambda\boldsymbol{y}^{\mathrm{T}}$, 称 $\boldsymbol{y}$ 为 $\boldsymbol{A}$ 属于 λ 的左特征向量. 相应地, 将 $\boldsymbol{x}$ 称为 $\boldsymbol{A}$ 属于 λ 的右特征向量. 一般来说, 左右特征向量不相等.

如果 $\boldsymbol{A}$ 有 r 个互不相同的特征值 $\lambda_1, \lambda_2, \cdots, \lambda_r$, 则 $\boldsymbol{A}$ 的特征多项式可表示为

$$\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2}\cdots(\lambda - \lambda_r)^{n_r},$$

称 n_i 为 λ_i 的代数重数. 记 $\gamma_i = n - \operatorname{rank}(\lambda_i\boldsymbol{I} - \boldsymbol{A})$, 称 γ_i 为

λ_i 的几何重数, 它表示属于 λ_i 的线性无关特征向量的个数, 满足 $1 \leqslant \gamma_i \leqslant n_i$. 若 $n_i = 1$, 则称 λ_i 为 $\boldsymbol{A}$ 的简单特征值. 若 $\gamma_i = n_i$, 则称 λ_i 为 $\boldsymbol{A}$ 的半简单特征值. 显然, 简单特征值必为半简单特征值. 满足 $p(\boldsymbol{A}) = \boldsymbol{O}$ 的首项系数为 1 且次数最低的多项式 p 称为 $\boldsymbol{A}$ 的最小多项式. 可以证明, $\boldsymbol{A}$ 的最小多项式具有下列形式, 即

$$p(\lambda) = (\lambda - \lambda_1)^{l_1}(\lambda - \lambda_2)^{l_2} \cdots (\lambda - \lambda_r)^{l_r}, \ 1 \leqslant l_i \leqslant n_i, \ i = 1, 2, \cdots, r.$$

定义 1.2 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$. 若 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{I}$, 则称 $\boldsymbol{A}$ 为酉矩阵. 实的酉矩阵称为正交矩阵, 即 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}$.

由定义 1.2 显然有, 酉矩阵 $\boldsymbol{A}$ 的逆矩阵为 $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{H}}$, 正交矩阵 $\boldsymbol{A}$ 的逆矩阵为 $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{T}}$.

由酉矩阵的定义, 容易证明酉矩阵还具有下列性质.

性质 1.1 设 $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{C}^{n\times n}$ 是酉矩阵, 则 $|\det(\boldsymbol{U})| = 1$, 且 $\boldsymbol{U}^{-1}$ 和 $\boldsymbol{UV}$ 也是酉矩阵.

性质 1.2 $\boldsymbol{U} \in \mathbb{C}^{n\times n}$ 是酉矩阵的充分必要条件是 $\boldsymbol{U}$ 的列向量是两两正交的单位向量.

定义 1.3 设 $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{n\times n}$. 若存在非奇异矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{BP}$, 则称 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相似. 若 $\boldsymbol{P}$ 为酉矩阵, 则称为酉相似. 若 $\boldsymbol{P}$ 为正交矩阵, 则称为正交相似.

显然, 相似矩阵有相同的特征多项式, 因而有相同的特征值.

定义 1.4 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$. 若 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$, 则称 $\boldsymbol{A}$ 为正规矩阵. 若 $\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}$, 则称 $\boldsymbol{A}$ 为 Hermite 矩阵. 实的 Hermite 矩阵称为实对称矩阵, 即 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}$. 若 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^{\mathrm{H}} = -\boldsymbol{A}$, 则称 $\boldsymbol{A}$ 为反 Hermite 矩阵. 实的反 Hermite 矩阵称为反对称矩阵, 即 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^{\mathrm{T}} = -\boldsymbol{A}$.

Back
Close

容易验证, 酉矩阵、正交矩阵、Hermite 矩阵、反 Hermite 矩阵、实对称矩阵、反对称矩阵都是正规矩阵.

根据反对称矩阵的定义容易证明反对称矩阵的下列性质.

性质 1.3 设 $\boldsymbol{A}=(a_{ij})\in\mathbb{R}^{n\times n}$ 是反对称矩阵, 即满足 $\boldsymbol{A}^{\mathrm{T}}=-\boldsymbol{A}$. 则

(1) $a_{ii}=0,\ i=1,2,\cdots,n$.

(2) 不存在奇数阶非奇异反对称矩阵.

(3) $\boldsymbol{A}$ 的特征值只能是 0 或纯虚数.

注 1.1 若 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$ 为 Hermite 矩阵, 则其对角元为 0 或纯虚数, 其特征值也只能是 0 或纯虚数.

定义 1.5 设 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$. 若对任意的非零向量 $\boldsymbol{x}\in\mathbb{R}^n$ 有 $(\boldsymbol{A}\boldsymbol{x},\boldsymbol{x})>0\ (\geqslant 0)$, 则称 $\boldsymbol{A}$ 为 (实) 正定矩阵 (半正定矩阵). 若 $\boldsymbol{A}$ 还是对称的, 则称为 (实) 对称正定矩阵 (对称半正定矩阵).

Back
Close

由定义 1.5 不难推得, 对称矩阵的特征值均为实数. 对称正定矩阵 (对称半正定矩阵) 的特征值均为正数 (非负数). 此外, 若记

$$\boldsymbol{H} = \frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}}), \quad \boldsymbol{S} = \frac{1}{2}(\boldsymbol{A} - \boldsymbol{A}^{\mathrm{T}}),$$

分别称为 $\boldsymbol{A}$ 的对称部分和反对称部分, 则显然对任意的矩阵 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 都可唯一地分裂为

$$\boldsymbol{A} = \boldsymbol{H} + \boldsymbol{S},$$

称为矩阵 $\boldsymbol{A}$ 的对称–反对称分裂.

性质 1.4 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 正定 (或半正定) 的充分必要条件是其对称部分 $\boldsymbol{H}$ 对称正定 (或对称半正定).

定义 1.6 设 $i_1, i_2, \cdots, i_n$ 是 $1, 2, \cdots, n$ 的一个排列, 以 n 阶单位矩阵 $\boldsymbol{I}_n$ 的 n 个列向量

$$\boldsymbol{e}_1 = (1, 0, \cdots, 0)^{\mathrm{T}}, \boldsymbol{e}_2 = (0, 1, \cdots, 0)^{\mathrm{T}}, \cdots, \boldsymbol{e}_n = (0, 0, \cdots, 1)^{\mathrm{T}}$$

为列构成的 n 阶矩阵 $\boldsymbol{P}=[\boldsymbol{e}_{i_1},\boldsymbol{e}_{i_2},\cdots,\boldsymbol{e}_{i_n}]$, 称为置换矩阵或排列矩阵.

例如

$$\boldsymbol{P}=[\boldsymbol{e}_3,\boldsymbol{e}_1,\boldsymbol{e}_2]=\begin{bmatrix}0&1&0\\0&0&1\\1&0&0\end{bmatrix}$$

是一个 3 阶置换矩阵.

性质 1.5 置换矩阵具有如下性质:

(1) 置换矩阵的转置仍是置换矩阵.

(2) 置换矩阵是正交矩阵.

(3) 设 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$, $\boldsymbol{P}=[\boldsymbol{e}_{i_1},\boldsymbol{e}_{i_2},\cdots,\boldsymbol{e}_{i_n}]$, 则 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}$ 是将 $\boldsymbol{A}$ 按 $i_1,i_2,\cdots,i_n$ 行重新排列所得到的矩阵, $\boldsymbol{AP}$ 是将 $\boldsymbol{A}$ 按 $i_1,i_2,\cdots,i_n$ 列重新排列得到的矩阵.

Back
Close

定义 1.7 设 $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{n\times n}$. 如果存在 n 阶非奇异矩阵 $\boldsymbol{C}$, 使得

(1) $\boldsymbol{B} = \boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}$, 则称 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 为 T–合同.

(2) $\boldsymbol{B} = \boldsymbol{C}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{C}$, 则称 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 为 H–合同.

显然, 这两个合同概念具有密切的联系. 当 $\boldsymbol{C}$ 是实矩阵时, T–合同和 H–合同是一致的. 此外, 容易证明, T–合同和 H–合同都是等价关系.

利用合同的定义还可以得到:

(1) 如果 $\boldsymbol{A}$ 是 Hermite 矩阵, 则 $\boldsymbol{C}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{C}$ 也是 Hermite 矩阵 (即使 $\boldsymbol{C}$ 是奇异矩阵).

(2) 如果 $\boldsymbol{A}$ 是对称矩阵 (不一定是实矩阵), 则 $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}$ 也是对称矩阵.

(3) 如果 $\boldsymbol{A}$ 是 Hermite 正定 (半正定) 矩阵, 则 $\boldsymbol{C}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{C}$ 也是

Hermite 正定 (半正定) 矩阵.

(4) 如果 $\boldsymbol{A}$ 是对称正定 (半正定) 矩阵, 则 $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{C}$ 也是对称正定 (半正定) 矩阵.

定义 1.8 对于 2×2 分块矩阵

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}, \tag{1.1}$$

当 $\boldsymbol{A}_{11}$ 可逆时, 称 $\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$ 为 $\boldsymbol{A}$ 关于 $\boldsymbol{A}_{11}$ 的 Schur 补, 记为 $\boldsymbol{A}/\boldsymbol{A}_{11}$. 当 $\boldsymbol{A}_{22}$ 可逆时, 称 $\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$ 为 $\boldsymbol{A}$ 关于 $\boldsymbol{A}_{22}$ 的 Schur 补, 记为 $\boldsymbol{A}/\boldsymbol{A}_{22}$.

定理 1.1 设 $\boldsymbol{A}$ 具有式 (1.1) 的分块形式, 且 $\boldsymbol{A}_{11}$ 可逆. 则

(1) $\det(\boldsymbol{A})=\det(\boldsymbol{A}_{11})\det(\boldsymbol{A}/\boldsymbol{A}_{11})$.

(2) $\operatorname{rank}(\boldsymbol{A})=\operatorname{rank}(\boldsymbol{A}_{11})+\operatorname{rank}(\boldsymbol{A}/\boldsymbol{A}_{11})$.

证明 注意到

$$\boldsymbol{B}=\begin{bmatrix}\boldsymbol{I} & \boldsymbol{O}\\ -\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\\ \boldsymbol{O} & \boldsymbol{I}\end{bmatrix}$$

$$=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{A}/\boldsymbol{A}_{11}\end{bmatrix}.$$

由此立即可得定理的结论. 证毕. □

由定理 1.1 立即可得下面的结论.

推论 1.1 设 $\boldsymbol{A}$ 具有式 (1.1) 的分块形式, 且 $\boldsymbol{A}_{22}$ 可逆. 则

(1) $\det(\boldsymbol{A})=\det(\boldsymbol{A}_{22})\det(\boldsymbol{A}/\boldsymbol{A}_{22})$.

(2) $\operatorname{rank}(\boldsymbol{A})=\operatorname{rank}(\boldsymbol{A}_{22})+\operatorname{rank}(\boldsymbol{A}/\boldsymbol{A}_{22})$.

Back
Close

推论 1.2 设
$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}$$
是 Hermite 矩阵, $\boldsymbol{A}$ 关于 $\boldsymbol{A}_{11}$ 的 Schur 补 $\boldsymbol{A}/\boldsymbol{A}_{11}=\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$. 则

(1) $\boldsymbol{A}$ 正定当且仅当 $\boldsymbol{A}_{11}$ 及 $\boldsymbol{A}/\boldsymbol{A}_{11}$ 均正定.

(2) 若 $\boldsymbol{A}_{11}$ 正定, 则 $\boldsymbol{A}$ 正定当且仅当 $\boldsymbol{A}/\boldsymbol{A}_{11}$ 正定.

§1.2 几种常用的矩阵分解

所谓矩阵分解, 就是将一个矩阵分解为 (从某种意义上讲) 比较简单或对其性质比较熟悉的若干个矩阵的乘积. 下面介绍几种常用的矩阵分解.

§1.2.1 矩阵的特征分解

特征分解, 又称谱分解, 是将矩阵分解为由其特征值和特征向量表示的矩阵之积的方法. 注意: 只有对可对角化矩阵才可以实施特征分解.

设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ 有 n 个线性无关的特征向量 $\boldsymbol{q}_i\,(i=1,2,\cdots,n)$. 则 $\boldsymbol{A}$ 可以被分解为

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}, \tag{1.2}$$

Back
Close

式中: $\boldsymbol{\varLambda}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$, $\lambda_i\ (i=1,2,\cdots,n)$ 为 $\boldsymbol{A}$ 的特征值; $\boldsymbol{Q}$ 为 n 阶方阵, 且其第 i 列为 $\boldsymbol{A}$ 的特征向量 $\boldsymbol{q}_i$.

一般来说, 特征向量 $\boldsymbol{q}_i\,(i=1,2,\cdots,n)$ 通常被正交单位化 (但这不是必须的). 未被正交单位化的特征向量组 $\boldsymbol{v}_i\,(i=1,2,\cdots,n)$ 也可以作为 $\boldsymbol{Q}$ 的列向量.

通常可以通过特征分解来求矩阵的逆. 若矩阵 $\boldsymbol{A}$ 可被特征分解并特征值中不含零, 则矩阵 $\boldsymbol{A}$ 为非奇异矩阵, 且其逆矩阵可以由下式给出:

$$\boldsymbol{A}^{-1}=\boldsymbol{Q}\boldsymbol{\varLambda}^{-1}\boldsymbol{Q}^{-1}=\boldsymbol{Q}\boldsymbol{\varLambda}^{-1}\boldsymbol{Q}^{\mathrm{T}}, \tag{1.3}$$

此处假设 $\boldsymbol{Q}$ 的列向量被正交单位化. 因为 $\boldsymbol{\varLambda}$ 为对角矩阵, 其逆矩阵容易计算出:

$$\boldsymbol{\varLambda}^{-1}=\mathrm{diag}(\lambda_1^{-1},\lambda_2^{-1},\cdots,\lambda_n^{-1}).$$

由于任意的 n 阶实对称矩阵 $\boldsymbol{A}$ 都有 n 个线性无关的特征向

量, 并且这些特征向量都可以正交单位化而得到一组正交且模为 1 的向量. 故实对称矩阵 $\boldsymbol{A}$ 可被分解成

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}},$$

式中: $\boldsymbol{Q}$ 为正交矩阵; $\boldsymbol{\Lambda}$ 为实对角矩阵.

类似地, 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ 为正规矩阵, 由于正规矩阵是可对角化的 (见定理 1.5), 故其具有一组标准正交特征向量基, 因此可被分解成

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

式中: $\boldsymbol{U}$ 为酉矩阵. 进一步地, 若 $\boldsymbol{A}$ 是 Hermite 矩阵，那么对角矩阵 $\boldsymbol{\Lambda}$ 的对角元全为实数. 若 $\boldsymbol{A}$ 是酉矩阵, 则 $\boldsymbol{\Lambda}$ 的所有对角元在复平面的单位圆上取得.

下面再给出矩阵特征分解的一个应用.

定理 1.2 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ 为 Hermite 正定 (半正定) 矩阵, 则存在唯一的 Hermite 正定 (半正定) 矩阵 $\boldsymbol{B}$ 使得

$$\boldsymbol{A} = \boldsymbol{B}^2. \tag{1.4}$$

称这样定义的矩阵 $\boldsymbol{B}$ 为矩阵 $\boldsymbol{A}$ 的平方根矩阵, 常记为 $\boldsymbol{A}^{1/2}$.

证明 只证明正定的情形. 存在性. 由于 $\boldsymbol{A}$ 是 Hermite 正定矩阵, 故存在酉矩阵 $\boldsymbol{P}$ 使得

$$\boldsymbol{A} = \boldsymbol{P}\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)\boldsymbol{P}^{\mathrm{H}},$$

式中: $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n > 0$ 为 $\boldsymbol{A}$ 的特征值. 令

$$\boldsymbol{B} = \boldsymbol{P}\mathrm{diag}(\lambda_1^{1/2}, \lambda_2^{1/2}, \cdots, \lambda_n^{1/2})\boldsymbol{P}^{\mathrm{H}},$$

则显然 $\boldsymbol{B}$ 为 Hermite 正定矩阵, 且 $\boldsymbol{B}^2 = \boldsymbol{A}$.

Back
Close

唯一性. 假设另有 Hermite 正定矩阵 $\boldsymbol{C}$ 满足 $\boldsymbol{A}=\boldsymbol{C}^2$. 则存在酉矩阵 $\boldsymbol{Q}$ 使得

$$\boldsymbol{C}=\boldsymbol{Q}\mathrm{diag}(\mu_1,\mu_2,\cdots,\mu_n)\boldsymbol{Q}^{\mathrm{H}},$$

式中: $\mu_1\geqslant\mu_2\geqslant\cdots\geqslant\mu_n>0$ 为 $\boldsymbol{C}$ 的特征值. 由于

$$\boldsymbol{C}^2=\boldsymbol{Q}\mathrm{diag}(\mu_1^2,\mu_2^2,\cdots,\mu_n^2)\boldsymbol{Q}^{\mathrm{H}}=\boldsymbol{A},$$

故 $\mu_i^2=\lambda_i\,(i=1,2,\cdots,n)$, 即 $\mu_i=\lambda_i^{1/2}$. 于是

$$\boldsymbol{C}=\boldsymbol{Q}\mathrm{diag}(\lambda_1^{1/2},\lambda_2^{1/2},\cdots,\lambda_n^{1/2})\boldsymbol{Q}^{\mathrm{H}}.$$

再由 $\boldsymbol{A}=\boldsymbol{B}^2=\boldsymbol{C}^2$, 得

$$\boldsymbol{P}\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\boldsymbol{P}^{\mathrm{H}}=\boldsymbol{Q}\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\boldsymbol{Q}^{\mathrm{H}},$$

或等价地, 有

$$\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{P}\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{P}.$$

记 $\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{P}=(t_{ij})$, 则有

$$t_{ij}\lambda_j=\lambda_i t_{ij},\ \ i,j=1,2,\cdots,n.$$

因此, 当 $\lambda_i\neq\lambda_j$ 时, 必有 $t_{ij}=0$. 故无论 λ_i 与 λ_j 是否相等, 都有

$$t_{ij}\lambda_j^{1/2}=t_{ij}\lambda_i^{1/2},\ \ i,j=1,2,\cdots,n.$$

即

$$\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{P}\mathrm{diag}(\lambda_1^{1/2},\lambda_2^{1/2},\cdots,\lambda_n^{1/2})=\mathrm{diag}(\lambda_1^{1/2},\lambda_2^{1/2},\cdots,\lambda_n^{1/2})\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{P},$$

或者等价地, 有

$$\boldsymbol{P}\mathrm{diag}(\lambda_1^{1/2},\lambda_2^{1/2},\cdots,\lambda_n^{1/2})\boldsymbol{P}^{\mathrm{H}}=\boldsymbol{Q}\mathrm{diag}(\lambda_1^{1/2},\lambda_2^{1/2},\cdots,\lambda_n^{1/2})\boldsymbol{Q}^{\mathrm{H}},$$

即 $\boldsymbol{B}=\boldsymbol{C}$. 证毕. $\square$

§1.2.2 矩阵的 Schur 分解

矩阵的 Schur 分解在理论上十分重要, 它是许多重要定理证明的出发点. 如矩阵论中极为重要的 Hamilton–Cayley (哈密顿–凯莱) 定理, 就可以利用矩阵的 Schur 分解定理进行简洁而优美的证明.

定理 1.3 (Schur 分解定理) 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, 则存在酉矩阵 $\boldsymbol{P} \in \mathbb{C}^{n\times n}$ 使得

$$\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{T},$$

式中: $\boldsymbol{T}$ 为上三角矩阵, 其对角元素是 $\boldsymbol{A}$ 的特征值, 而且可以选取 $\boldsymbol{P}$ 使得 $\boldsymbol{T}$ 的对角元可以任意排列.

证明 对矩阵的阶数 n 用数学归纳法. 当 $n = 1$ 时, 结论显然成立. 假设结论对 $n-1$ 阶矩阵成立, 下证对 n 阶矩阵成立.

设 $\boldsymbol{A}$ 的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, $\boldsymbol{x}_1$ 为对应于特征值 λ_1 的

Back
Close

单位特征向量. 将 $\boldsymbol{x}_1$ 扩充为向量空间 $\mathbb{C}^n$ 的一组标准正交基 $\boldsymbol{x}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_n$. 定义酉矩阵 $\boldsymbol{P}_0 = [\boldsymbol{x}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_n]$, 则

$$\boldsymbol{P}_0^{\mathrm{H}} \boldsymbol{A} \boldsymbol{P}_0 = \begin{bmatrix} \lambda_1 & * \\ \boldsymbol{0} & \boldsymbol{A}_1 \end{bmatrix},$$

式中: $\boldsymbol{A}_1$ 为 $n-1$ 阶方阵, 其特征值为 $\lambda_2, \cdots, \lambda_n$.

由归纳法假设, 存在 $n-1$ 阶酉矩阵 $\boldsymbol{P}_1$ 使得

$$\boldsymbol{P}_1^{\mathrm{H}} \boldsymbol{A}_1 \boldsymbol{P}_1 = \boldsymbol{T}_1,$$

式中: $\boldsymbol{T}_1$ 为 $n-1$ 阶上三角矩阵, 其对角元为 $\boldsymbol{A}_1$ 的特征值 $\lambda_2, \cdots, \lambda_n$. 令

$$\boldsymbol{P} = \boldsymbol{P}_0 \begin{bmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}_1 \end{bmatrix}, \quad \boldsymbol{T} = \begin{bmatrix} \lambda_1 & * \\ \boldsymbol{0} & \boldsymbol{T}_1 \end{bmatrix}.$$

则有

$$
\begin{aligned}
\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P} &= \begin{bmatrix} 1 & \mathbf{0} \\ 0 & \boldsymbol{P}_1^{\mathrm{H}} \end{bmatrix} \boldsymbol{P}_0^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}_0 \begin{bmatrix} 1 & \mathbf{0} \\ 0 & \boldsymbol{P}_1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & \mathbf{0} \\ 0 & \boldsymbol{P}_1^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} \lambda_1 & * \\ 0 & \boldsymbol{A}_1 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ 0 & \boldsymbol{P}_1 \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 & * \\ 0 & \boldsymbol{P}_1^{\mathrm{H}}\boldsymbol{A}_1\boldsymbol{P}_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & * \\ 0 & \boldsymbol{T}_1 \end{bmatrix} = \boldsymbol{T}.
\end{aligned}
$$

由归纳法假设, $\boldsymbol{T}_1$ 的对角元可以任意排列, 故 $\boldsymbol{T}$ 的对角元也可以任意排列. 证毕. □

推论 1.3 设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 的特征值都是实数, 则 $\boldsymbol{A}$ 正交相似于上三角矩阵.

Back
Close

设多项式

$$f(\lambda)=b_0\lambda^m+b_1\lambda^{m-1}+\cdots+b_{m-1}\lambda+b_m,$$

对于 n 阶矩阵 $\boldsymbol{A}$, 定义矩阵多项式

$$f(\boldsymbol{A})=b_0\boldsymbol{A}^m+b_1\boldsymbol{A}^{m-1}+\cdots+b_{m-1}\boldsymbol{A}+b_m\boldsymbol{I},$$

式中: $\boldsymbol{I}$ 为单位矩阵. 如果 $f(\boldsymbol{A})=\boldsymbol{O}$, 那么称 $\boldsymbol{A}$ 为 $f(\lambda)$ 的矩阵根.

下面利用矩阵的 Schur 分解来证明著名的 Hamilton–Cayley 定理.

定理 1.4 (Hamilton–Cayley 定理) 设 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$, $p(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})$, 则 $p(\boldsymbol{A})=\boldsymbol{O}$. 即矩阵 $\boldsymbol{A}$ 的特征多项式是它的零化多项式.

Back
Close

证明 设 n 阶矩阵 $\boldsymbol{A}$ 的特征多项式为

$$p(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})=\lambda^n+a_1\lambda^{n-1}+\cdots+a_{n-1}\lambda+a_n,$$

记 $\boldsymbol{A}$ 的 n 个特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_n$, 则有

$$p(\lambda)=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n),$$

由定理 1.3 可知, 存在酉矩阵 $\boldsymbol{U}$, 使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\begin{bmatrix}\lambda_1 & * & \cdots & * \\ & \lambda_2 & \ddots & \vdots \\ & & \ddots & * \\ & & & \lambda_n\end{bmatrix},$$

于是,

$$p(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}) = (\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} - \lambda_1\boldsymbol{I})(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} - \lambda_2\boldsymbol{I})\cdots(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} - \lambda_n\boldsymbol{I})$$

$$= \begin{bmatrix} 0 & * & \cdots & * \\ & \lambda_2-\lambda_1 & \ddots & \vdots \\ & & \ddots & * \\ & & & \lambda_n-\lambda_1 \end{bmatrix} \times \begin{bmatrix} \lambda_1-\lambda_2 & * & \cdots & * \\ & 0 & \ddots & \vdots \\ & & \ddots & * \\ & & & \lambda_n-\lambda_2 \end{bmatrix}$$

$$\times \begin{bmatrix} \lambda_1-\lambda_3 & * & * & \cdots & * \\ & \lambda_2-\lambda_3 & * & \cdots & * \\ & & 0 & \ddots & \vdots \\ & & & \ddots & * \\ & & & & \lambda_n-\lambda_3 \end{bmatrix} \times\cdots\times \begin{bmatrix} \lambda_1-\lambda_n & * & * & \cdots & * \\ & \lambda_2-\lambda_n & * & \cdots & * \\ & & \ddots & \ddots & \vdots \\ & & & \lambda_{n-1}-\lambda_n & * \\ & & & & 0 \end{bmatrix}$$

$$
= \begin{bmatrix} 0 & 0 & * & \cdots & * \\ 0 & 0 & * & \cdots & * \\ 0 & 0 & * & \cdots & * \\ 0 & 0 & & \ddots & \vdots \\ 0 & 0 & & & * \end{bmatrix} \times \begin{bmatrix} \lambda_1 - \lambda_3 & * & * & \cdots & * \\ & \lambda_2 - \lambda_3 & * & \cdots & * \\ & & 0 & \ddots & \vdots \\ & & & \ddots & * \\ & & & & \lambda_n - \lambda_3 \end{bmatrix}
$$

$$
\times \cdots \times \begin{bmatrix} \lambda_1 - \lambda_n & * & * & \cdots & * \\ & \lambda_2 - \lambda_n & * & \cdots & * \\ & & \ddots & \ddots & \vdots \\ & & & \lambda_{n-1} - \lambda_n & * \\ & & & & 0 \end{bmatrix} = \boldsymbol{O},
$$

即 $\boldsymbol{U}^{\mathrm{H}} p(\boldsymbol{A}) \boldsymbol{U} = \boldsymbol{O}$, 也就是 $p(\boldsymbol{A}) = \boldsymbol{O}$. 证毕. $\square$

Back

Close

在计算数学领域, 人们颇为关心的是, 通过相似变换可以把一个矩阵变换成何种最简单的形状. 利用定理 1.3 和推论 1.3, 可以导出矩阵酉 (正交) 相似于对角矩阵的充要条件.

定理 1.5 矩阵 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ 酉相似于对角矩阵当且仅当 $\boldsymbol{A}$ 是正规矩阵, 即 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$. 换言之, 正规矩阵一定可对角化.

证明 必要性. 设酉矩阵 $\boldsymbol{U}$, 使得 $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{\Lambda}$, 其中 $\boldsymbol{\Lambda}$ 为对角矩阵, 则有

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}, \quad \boldsymbol{A}^{\mathrm{H}} = \boldsymbol{U}\bar{\boldsymbol{\Lambda}}\boldsymbol{U}^{\mathrm{H}},$$

$$\begin{aligned}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} &= \boldsymbol{U}\bar{\boldsymbol{\Lambda}}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}} = \boldsymbol{U}\bar{\boldsymbol{\Lambda}}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}} = \boldsymbol{U}\boldsymbol{\Lambda}\bar{\boldsymbol{\Lambda}}\boldsymbol{U}^{\mathrm{H}} \\ &= \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U}\bar{\boldsymbol{\Lambda}}\boldsymbol{U}^{\mathrm{H}} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}.\end{aligned}$$

充分性. 由 Schur 分解定理可知, 存在酉矩阵 $\boldsymbol{U}$, 使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ & t_{22} & \cdots & t_{2n} \\ & & \ddots & \vdots \\ & & & t_{nn} \end{bmatrix}:=\boldsymbol{T}.$$

利用条件 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$, 得

$$\begin{aligned}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{T}&=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}\\&=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}=\boldsymbol{T}\boldsymbol{T}^{\mathrm{H}}.\end{aligned}$$

比较上式两端矩阵的对应元素, 可得 $t_{ij}=0\ (i<j)$, 也就是 $\boldsymbol{T}=\mathrm{diag}(t_{11},t_{22},\cdots,t_{nn})$, 即 $\boldsymbol{A}$ 酉相似于对角矩阵. 证毕. □

推论 1.4 设 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ 的 n 个特征值都是实数, 则 $\boldsymbol{A}$ 正交相似于对角矩阵的充要条件是 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$.

推论 1.5 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ 为正规矩阵, λ 是 $\boldsymbol{A}$ 的特征值, $\boldsymbol{x}$ 是对应的特征向量. 则 $\overline{\lambda}$ 是 $\boldsymbol{A}^{\mathrm{H}}$ 的特征值, $\overline{\lambda}$ 对应的特征向量仍然是 $\boldsymbol{x}$.

证明 由定理 1.5, 存在 n 阶酉矩阵 $\boldsymbol{U}$ 使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n).$$

于是有 $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U} = \mathrm{diag}(\overline{\lambda}_1, \overline{\lambda}_2, \cdots, \overline{\lambda}_n)$. 设 $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n]$, 则上面两式可写为

$$\boldsymbol{A}\boldsymbol{u}_i = \lambda_i\boldsymbol{u}_i, \quad \boldsymbol{A}^{\mathrm{H}}\boldsymbol{u}_i = \overline{\lambda}_i\boldsymbol{u}_i, \quad i = 1, 2, \cdots, n.$$

由此可见, 若 λ_i 是 $\boldsymbol{A}$ 的特征值且 $\boldsymbol{u}_i$ 是相应的特征向量时, $\overline{\lambda}_i$ 是 $\boldsymbol{A}^{\mathrm{H}}$ 的特征值且相应的特征向量仍是 $\boldsymbol{u}_i$. 证毕. $\square$

推论 1.6 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ 为正规矩阵, λ, μ 是 $\boldsymbol{A}$ 的特征值, $\boldsymbol{x}, \boldsymbol{y}$ 是对应的特征向量. 如果 $\lambda \neq \mu$, 则 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 正交, 即 $\boldsymbol{y}^{\mathrm{H}}\boldsymbol{x} = 0$.

证明 因为 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$, $\boldsymbol{A}\boldsymbol{y}=\mu\boldsymbol{y}$. 由推论 1.5 得 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}=\overline{\lambda}\boldsymbol{x}$. 故

$$\overline{\mu}\boldsymbol{y}^{\mathrm{H}}\boldsymbol{x}=(\mu\boldsymbol{y})^{\mathrm{H}}\boldsymbol{x}=(\boldsymbol{A}\boldsymbol{y})^{\mathrm{H}}\boldsymbol{x}=\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}=\overline{\lambda}\boldsymbol{y}^{\mathrm{H}}\boldsymbol{x},$$

即 $\overline{(\lambda-\mu)}\boldsymbol{y}^{\mathrm{H}}\boldsymbol{x}=0$. 由题设 $\lambda\neq\mu$, 故必有 $\boldsymbol{y}^{\mathrm{H}}\boldsymbol{x}=0$. 证毕. $\square$

推论 1.7 设 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$, 则 $\boldsymbol{A}$ 酉相似于实对角矩阵的充要条件是 $\boldsymbol{A}$ 为 Hermite 矩阵.

证明 必要性. 设存在 n 阶酉矩阵 $\boldsymbol{U}$ 使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)=\boldsymbol{\Lambda},$$

式中: $\lambda_i\,(i=1,2,\cdots,n)$ 为实数, 则有

$$\boldsymbol{A}^{\mathrm{H}}=(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}})^{\mathrm{H}}=\boldsymbol{U}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}=\boldsymbol{A},$$

即 $\boldsymbol{A}$ 是 Hermite 矩阵.

充分性. 若 $\boldsymbol{A}$ 是 Hermite 矩阵, 则 $\boldsymbol{A}$ 为正规矩阵, 于是存在酉矩阵 $\boldsymbol{U}$ 使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) = \boldsymbol{\Lambda},$$

故

$$\boldsymbol{\Lambda}^{\mathrm{H}} = (\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U})^{\mathrm{H}} = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U} = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{\Lambda}.$$

从而 $\overline{\lambda}_i = \lambda_i\,(i = 1, 2, \cdots, n)$, 即 λ_i 都是实数. 证毕. $\square$

推论 1.8 *设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ 是反 Hermite 矩阵, 则存在 n 阶酉矩阵 $\boldsymbol{U}$ 使得*

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U} = \mathrm{diag}(\mathrm{i}b_1, \mathrm{i}b_2, \cdots, \mathrm{i}b_n),$$

式中: $b_i\,(i = 1, 2, \cdots, n)$ 为实数.

证明 由于反 Hermite 矩阵是正规矩阵且其特征值为 0 或纯虚数, 由定理 1.5 即得结论. 证毕. $\square$

下面介绍两个 n 阶实对称矩阵“同时”相似于对角矩阵的问题.

定理 1.6 设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都是 n 阶实对称矩阵, 则存在正交矩阵 $\boldsymbol{P}$, 使得 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$ 和 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{P}$ 都是对角矩阵的充要条件是 $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A}$.

证明 必要性. 设 n 阶正交矩阵 $\boldsymbol{P}$ 使得

$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{\Lambda}, \quad \boldsymbol{P}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{P} = \boldsymbol{\Sigma},$$

式中: $\boldsymbol{\Lambda}$ 和 $\boldsymbol{\Sigma}$ 为对角矩阵. 则有

$$\begin{aligned}\boldsymbol{A}\boldsymbol{B} &= \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{\Sigma}\boldsymbol{P}^{\mathrm{T}} \\ &= \boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{\Lambda}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{B}\boldsymbol{A}.\end{aligned}$$

充分性. 设 $\boldsymbol{A}$ 的全体互异特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_r$, 相应的重数为 $n_1, n_2, \cdots, n_r$ $(n_1 + n_2 + \cdots + n_r = n)$. 由 $\boldsymbol{A}$ 实对称知, 存在

正交矩阵 $\boldsymbol{Q}$ 使得

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}=\begin{bmatrix}\lambda_1\boldsymbol{I}_{n_1} & & & \\ & \lambda_2\boldsymbol{I}_{n_2} & & \\ & & \ddots & \\ & & & \lambda_r\boldsymbol{I}_{n_r}\end{bmatrix},$$

将矩阵 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Q}$ 划分为如下分块形式

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Q}=\begin{bmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \cdots & \boldsymbol{B}_{1r}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} & \cdots & \boldsymbol{B}_{2r}\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{B}_{r1} & \boldsymbol{B}_{r2} & \cdots & \boldsymbol{B}_{rr}\end{bmatrix},$$

式中: $\boldsymbol{B}_{ij}$ 为 $n_i\times n_j$ 阶矩阵 $(i,j=1,2,\cdots,r)$.

由 $\boldsymbol{A}\boldsymbol{B}=\boldsymbol{B}\boldsymbol{A}$ 及 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\boldsymbol{I}$, 得

$$(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q})(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Q})=(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Q})(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}),$$

即

$$\begin{bmatrix} \lambda_1\boldsymbol{B}_{11} & \lambda_1\boldsymbol{B}_{12} & \cdots & \lambda_1\boldsymbol{B}_{1r} \\ \lambda_2\boldsymbol{B}_{21} & \lambda_2\boldsymbol{B}_{22} & \cdots & \lambda_2\boldsymbol{B}_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_r\boldsymbol{B}_{r1} & \lambda_r\boldsymbol{B}_{r2} & \cdots & \lambda_r\boldsymbol{B}_{rr} \end{bmatrix} = \begin{bmatrix} \lambda_1\boldsymbol{B}_{11} & \lambda_2\boldsymbol{B}_{12} & \cdots & \lambda_r\boldsymbol{B}_{1r} \\ \lambda_1\boldsymbol{B}_{21} & \lambda_2\boldsymbol{B}_{22} & \cdots & \lambda_r\boldsymbol{B}_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1\boldsymbol{B}_{r1} & \lambda_2\boldsymbol{B}_{r2} & \cdots & \lambda_r\boldsymbol{B}_{rr} \end{bmatrix},$$

也就是

$$\lambda_i\boldsymbol{B}_{ij} = \lambda_j\boldsymbol{B}_{ij} \quad (i,j=1,2,\cdots,r).$$

当 $i \neq j$ 时, 由 $\lambda_i \neq \lambda_j$ 可得 $\boldsymbol{B}_{ij} = \boldsymbol{O}\ (i \neq j;\ i,j=1,2,\cdots,r)$. 于是

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Q} = \mathrm{diag}(\boldsymbol{B}_{11}, \boldsymbol{B}_{22}, \cdots, \boldsymbol{B}_{rr}),$$

因为 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Q}$ 实对称, 所以 $\boldsymbol{B}_{ii}\,(i=1,2,\cdots,r)$ 实对称, 那么存在正交矩阵 $\boldsymbol{U}_i$, 使得 $\boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{B}_{ii}\boldsymbol{U}_i = \boldsymbol{\Sigma}_i\,(i=1,2,\cdots,r)$, 其中 $\boldsymbol{\Sigma}_i$ 为对角矩阵. 令 $\boldsymbol{U} = \mathrm{diag}(\boldsymbol{U}_1, \boldsymbol{U}_2, \cdots, \boldsymbol{U}_r)$, 则 $\boldsymbol{U}$ 为正交矩阵, 从而

$\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{U}$ 也为正交矩阵, 且有

$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{P} = \boldsymbol{U}^{\mathrm{T}}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Q})\boldsymbol{U} = \mathrm{diag}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \cdots, \boldsymbol{\Sigma}_r),$$
$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{U}^{\mathrm{T}}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q})\boldsymbol{U} = \mathrm{diag}(\lambda_1\boldsymbol{I}_{n_1}, \lambda_2\boldsymbol{I}_{n_2}, \cdots, \lambda_r\boldsymbol{I}_{n_r}).$$

证毕. □

定理 1.6 给出了两个实对称矩阵“同时”正交相似于对角矩阵的充要条件, 下面再给出两个 Hermite 矩阵“同时”合同于对角矩阵的充分条件.

定理 1.7 设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都是 n 阶 Hermite 矩阵, 且 $\boldsymbol{B}$ 为正定矩阵, 则存在可逆矩阵 $\boldsymbol{P}$, 使得 $\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}$ 和 $\boldsymbol{P}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{P}$ 都是对角矩阵.

证明 由 $\boldsymbol{B}$ 正定知, 存在酉矩阵 $\boldsymbol{Q}_1$, 使得

$$\boldsymbol{Q}_1^{\mathrm{H}}\boldsymbol{B}\boldsymbol{Q}_1=\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}=\boldsymbol{\Lambda}_1,\quad (\lambda_i>0),$$

令

$$\boldsymbol{Q}_2=\boldsymbol{Q}_1\begin{bmatrix}1/\sqrt{\lambda_1} & & \\ & \ddots & \\ & & 1/\sqrt{\lambda_n}\end{bmatrix},$$

则 $\boldsymbol{Q}_2$ 可逆, 且有 $\boldsymbol{Q}_2^{\mathrm{H}}\boldsymbol{B}\boldsymbol{Q}_2=\boldsymbol{I}$. 对于 Hermite 矩阵 $\boldsymbol{Q}_2^{\mathrm{H}}\boldsymbol{A}\boldsymbol{Q}_2$, 存在酉矩阵 $\boldsymbol{Q}_3$, 使得

$$\boldsymbol{Q}_3^{\mathrm{H}}(\boldsymbol{Q}_2^{\mathrm{H}}\boldsymbol{A}\boldsymbol{Q}_2)\boldsymbol{Q}_3=\begin{bmatrix}\mu_1 & & \\ & \ddots & \\ & & \mu_n\end{bmatrix}=\boldsymbol{\Lambda}_2.$$

令 $\boldsymbol{P}=\boldsymbol{Q}_2\boldsymbol{Q}_3$, 则 $\boldsymbol{P}$ 可逆, 且有

$$\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{\Lambda}_2,\quad \boldsymbol{P}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{P}=\boldsymbol{I},$$

式中: $\boldsymbol{\Lambda}_2$ 和 $\boldsymbol{I}$ 为对角矩阵. 证毕. □

§1.2.3 矩阵的奇异值分解

矩阵的奇异值分解在理论上和实际应用中都十分重要. 特别地, 它已成为信息处理、多元统计分析等工程技术领域中不可缺少的工具. 由 1.2.2 节的 Schur 分解定理可知, 任一方阵用酉相似变换只能约化为上三角矩阵, 不能约化为对角矩阵. 而奇异值分解定理表明: 用两个酉矩阵乘到一个矩阵 $\boldsymbol{A}$ 的两边就可以变为对角矩阵. 奇异值分解常用于奇异的或数值上非常接近奇异的矩阵计算. 它不仅能判断矩阵是否接近奇异, 而且也用于数值求解.

对于任意的 $\boldsymbol{A}\in\mathbb{C}^{m\times n}$, 容易验证:

(1) $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 是 Hermite (半) 正定矩阵.

(2) 齐次线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 与 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 同解.

(3) $\mathrm{rank}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})=\mathrm{rank}(\boldsymbol{A})$.

(4) $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{O}\Longleftrightarrow\boldsymbol{A}=\boldsymbol{O}$.

定义 1.9 设 $\boldsymbol{A}\in\mathbb{C}_r^{m\times n}(r\geqslant 1)$, 记 Hermite 矩阵 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的 n 个特征值为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\lambda_r>\lambda_{r+1}=\cdots=\lambda_n=0$, 则称

$$\sigma_i=\sqrt{\lambda_i},\quad i=1,2,\cdots,n$$

为 $\boldsymbol{A}$ 的奇异值.

容易看出, $\boldsymbol{A}$ 的奇异值的个数与 $\boldsymbol{A}$ 的列数相同, $\boldsymbol{A}$ 的正奇异值的个数与 $\boldsymbol{A}$ 的秩相同.

定理 1.8 (奇异值分解定理) 设 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}$ 的正奇异值为 $\sigma_1, \sigma_2, \cdots, \sigma_r$. 则存在 m 阶酉矩阵 $\boldsymbol{U}$ 和 n 阶酉矩阵 $\boldsymbol{V}$, 使得

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}, \tag{1.5}$$

式中: $\boldsymbol{\Sigma}$ 为 $m \times n$ 阶对角矩阵, 且

$$\boldsymbol{\Sigma} = \begin{array}{c} \left[\begin{array}{cc} \boldsymbol{\Sigma}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{array}\right] \begin{array}{l} r \\ m-r \end{array} \\ \begin{array}{cc} r & n-r \end{array} \end{array}, \tag{1.6}$$

这里 $\boldsymbol{\Sigma}_r = \operatorname{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$, $\sigma_i > 0$, $1 \leqslant i \leqslant r$, $r \leqslant \min\{m, n\}$.

证明 对任意的 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}$, $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 是 n 阶 Hermite 半正定矩阵, 它的 n 个非负特征值记为 $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$. 不妨设 $\sigma_1^2 \geqslant \sigma_2^2 \geqslant \cdots \geqslant \sigma_r^2 > 0$, $\sigma_{r+1}^2 = \sigma_{r+2}^2 = \cdots = \sigma_n^2 = 0$, 这里 $r \leqslant \min\{m, n\}$. $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的

特征分解为

$$\boldsymbol{V}^{\mathrm{H}}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})\boldsymbol{V}=\mathrm{diag}(\sigma_1^2,\sigma_2^2,\cdots,\sigma_n^2), \tag{1.7}$$

式中: $\boldsymbol{V}=[\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n]$ 为 n 阶酉矩阵.

划分 $\boldsymbol{V}=[\boldsymbol{V}_1,\ \boldsymbol{V}_2]$, 其中 $\boldsymbol{V}_1=[\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_r]$, $\boldsymbol{V}_2=[\boldsymbol{v}_{r+1},\boldsymbol{v}_{r+2},\cdots,\boldsymbol{v}_n]$, 令 $\boldsymbol{\Sigma}_r=\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$, 则有

$$\begin{aligned}\boldsymbol{V}^{\mathrm{H}}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})\boldsymbol{V}&=\begin{bmatrix}\boldsymbol{V}_1^{\mathrm{H}}\\ \boldsymbol{V}_2^{\mathrm{H}}\end{bmatrix}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})[\boldsymbol{V}_1,\boldsymbol{V}_2]\\&=\begin{bmatrix}\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_1 & \boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_2\\ \boldsymbol{V}_2^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_1 & \boldsymbol{V}_2^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_2\end{bmatrix}.\end{aligned} \tag{1.8}$$

对比式 (1.7) 和式 (1.8), 得

Back
Close

$$\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_1 = \boldsymbol{\Sigma}_r^2, \quad \boldsymbol{V}_2^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_2 = \boldsymbol{O},$$
$$\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_2 = \boldsymbol{O}, \quad \boldsymbol{V}_2^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_1 = \boldsymbol{O}. \tag{1.9}$$

由式 (1.9) 的第 1, 2 个等式可知 $\boldsymbol{A}\boldsymbol{V}_1$ 的列互相正交, $\boldsymbol{A}\boldsymbol{V}_2$ 的列为零向量. 取 $m \times r$ 阶矩阵

$$\boldsymbol{U}_1 = \boldsymbol{A}\boldsymbol{V}_1\boldsymbol{\Sigma}_r^{-1}, \tag{1.10}$$

则 $\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{U}_1 = \boldsymbol{I}_r$. 再取 $\boldsymbol{U}_2 \in \mathbb{C}^{m\times(n-r)}$, 使得 $\boldsymbol{U} = [\boldsymbol{U}_1, \boldsymbol{U}_2]$ 为酉矩阵. 则由 $\boldsymbol{A}\boldsymbol{V}_2 = \boldsymbol{O}$ 及 $\boldsymbol{U}_2^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_1 = \boldsymbol{U}_2^{\mathrm{H}}(\boldsymbol{U}_1\boldsymbol{\Sigma}_r) = (\boldsymbol{U}_2^{\mathrm{H}}\boldsymbol{U}_1)\boldsymbol{\Sigma}_r = \boldsymbol{O}$, 立即推得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V} = \begin{bmatrix} \boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_1 & \boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_2 \\ \boldsymbol{U}_2^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_1 & \boldsymbol{U}_2^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} = \boldsymbol{\Sigma}.$$

证毕. $\square$

在奇异值分解式 (1.5) 中, $\boldsymbol{U}$ 的第 i 列是 $\boldsymbol{A}$ 的对应于奇异值 σ_i 的左奇异向量, $\boldsymbol{V}$ 的第 i 列是 $\boldsymbol{A}$ 的对应于奇异值 σ_i 的右奇异向量. 从定理的证明过程不难看出, $\boldsymbol{A}$ 的奇异值由 $\boldsymbol{A}$ 唯一确定, 但对应于每个奇异值的奇异向量一般不是唯一的.

注 1.2 *若 $\boldsymbol{A} \in \mathbb{R}_r^{m\times n}$, 则定理 1.8 中的 $\boldsymbol{U}, \boldsymbol{V}$ 都可以取为实矩阵, 即 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 分别为 m 阶和 n 阶的正交矩阵.*

两个与矩阵 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ 有关的重要子空间是其列空间和零空间.

定义 1.10 *设矩阵 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$, 分别称*

$$\mathcal{R}(\boldsymbol{A}) = \{\boldsymbol{y} : \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \in \mathbb{C}^n\} \subset \mathbb{C}^m,$$

$$\mathcal{N}(\boldsymbol{A}) = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}, \boldsymbol{x} \in \mathbb{C}^n\} \subset \mathbb{C}^n$$

为 $\boldsymbol{A}$ 的列空间 (值域) 和零空间 (核).

容易验证: 若将 $\boldsymbol{A}$ 按列划分为 $\boldsymbol{A}=\left[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\right]$, 则

$$\mathcal{R}(\boldsymbol{A})=\operatorname{span}\left\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n\right\}.$$

给定两个向量子空间 $\mathcal{S}_1$ 和 $\mathcal{S}_2$, 它们的和 $\mathcal{S}$ 是一个子空间, 其每一个向量都是 $\mathcal{S}_1$ 的一个向量与 $\mathcal{S}_2$ 的一个向量之和. 两个子空间的交也是一个子空间. 如果 $\mathcal{S}_1$ 和 $\mathcal{S}_2$ 的交退化为 $\{\mathbf{0}\}$, 则 $\mathcal{S}_1$ 和 $\mathcal{S}_2$ 之和称为直和, 表示为 $\mathcal{S}=\mathcal{S}_1\oplus\mathcal{S}_2$. 当 $\mathcal{S}$ 等于 $\mathbb{C}^n$ 时, 则 $\mathbb{C}^n$ 的每一个向量 $\boldsymbol{x}$ 可被唯一地写成 $\boldsymbol{x}=\boldsymbol{x}_1+\boldsymbol{x}_2$, 其中 $\boldsymbol{x}_1\in\mathcal{S}_1,\boldsymbol{x}_2\in\mathcal{S}_2$.

关于矩阵 $\boldsymbol{A}\in\mathbb{C}^{m\times n}$ 的列空间和零空间有下面的正交分解:

$$\mathcal{R}(\boldsymbol{A})\oplus\mathcal{N}(\boldsymbol{A}^{\mathrm{H}})=\mathbb{C}^m,\quad \mathcal{N}(\boldsymbol{A})\oplus\mathcal{R}(\boldsymbol{A}^{\mathrm{H}})=\mathbb{C}^n.$$

此外, 当 $\boldsymbol{A}\mathcal{S}\subset\mathcal{S}$ 时, 称子空间 $\mathcal{S}$ *在矩阵* $\boldsymbol{A}$ *之下是不变的*. 易见, 矩阵 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$ 的列空间 $\mathcal{R}(\boldsymbol{A})$ 和零空间 $\mathcal{N}(\boldsymbol{A})$ 都是 $\boldsymbol{A}$ 的不变子空间. 特别地, 对 $\boldsymbol{A}$ 的任意特征值 λ, 子空间 $\mathcal{N}(\boldsymbol{A}-\lambda\boldsymbol{I})$ 在 $\boldsymbol{A}$ 下

不变. 子空间 $\mathcal{N}(\boldsymbol{A}-\lambda\boldsymbol{I})$ 称为相应于 λ 的特征子空间, 它包含零向量及 $\boldsymbol{A}$ 的所有相应于 λ 的特征向量.

定理 1.9 在 $\boldsymbol{A}$ 的奇异值分解式 (1.5) 中, 记 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的列向量分别为 $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_m$ 和 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$, 则

(1) $\boldsymbol{A}$ 的秩等于其非零奇异值的个数.

(2) $\|\boldsymbol{A}\|_2 = \max\limits_{1\leqslant i\leqslant n} \sigma_i$. 设 σ_1 为最大奇异值, 其对应的右奇异向量为 $\boldsymbol{v}_1$, 则 $\|\boldsymbol{A}\|_2 = \|\boldsymbol{A}\boldsymbol{v}_1\|_2$.

(3) $\mathcal{N}(\boldsymbol{A}) = \operatorname{span}\{\boldsymbol{v}_{r+1}, \boldsymbol{v}_{r+2}, \cdots, \boldsymbol{v}_n\}$.

(4) $\mathcal{R}(\boldsymbol{A}) = \operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_r\}$.

(5) $\boldsymbol{A} = \sum\limits_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{H}}$.

证明 沿用定理 1.8 中的记号, 式 (1.5) 可写为

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}} = [\boldsymbol{U}_1, \boldsymbol{U}_2] \begin{bmatrix} \boldsymbol{\Sigma}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1^{\mathrm{H}} \\ \boldsymbol{V}_2^{\mathrm{H}} \end{bmatrix} = \boldsymbol{U}_1 \boldsymbol{\Sigma}_r \boldsymbol{V}_1^{\mathrm{H}}. \tag{1.11}$$

(1) 由于 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 都是非奇异矩阵, 故立即有 $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{\Sigma}) = r$.

(2) 注意到 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 都是酉矩阵, 故有

$$\begin{aligned}
\|\boldsymbol{A}\|_2 &= \max_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{A}\boldsymbol{x}\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{x}\|_2 = \max_{\|\boldsymbol{y}\|_2=1} \|\boldsymbol{\Sigma}\boldsymbol{y}\|_2 \\
&= \max\left\{ \left[(\sigma_1 y_1)^2 + (\sigma_2 y_2)^2 + \cdots + (\sigma_n y_n)^2\right]^{1/2} : \|\boldsymbol{y}\|_2 = 1 \right\} \\
&= \max_{1\leqslant i\leqslant n} \sigma_i.
\end{aligned}$$

设 σ_1 为最大奇异值, 其对应的右奇异向量为 $\boldsymbol{v}_1$, 则

$$\|\boldsymbol{A}\boldsymbol{v}_1\|_2 = \|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}}\boldsymbol{v}_1\|_2 = \|\boldsymbol{\Sigma}\boldsymbol{e}_1\|_2 = \sigma_1 = \|\boldsymbol{A}\|_2.$$

(3) 有

$$\begin{aligned}\mathcal{N}(\boldsymbol{A}) &= \left\{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\right\} = \left\{\boldsymbol{x} : \boldsymbol{U}_1\boldsymbol{\Sigma}_r\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{x} = \boldsymbol{0}\right\} = \left\{\boldsymbol{x} : \boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{x} = \boldsymbol{0}\right\}\\ &= \left\{\boldsymbol{x} : \boldsymbol{x} = k_{r+1}\boldsymbol{v}_{r+1} + \cdots + k_n\boldsymbol{v}_n, k_i \in \mathbb{C}\right\}\\ &= \operatorname{span}\left\{\boldsymbol{v}_{r+1}, \boldsymbol{v}_{r+2}, \cdots, \boldsymbol{v}_n\right\}.\end{aligned}$$

(4) 由于

$$\begin{aligned}\mathcal{R}(\boldsymbol{A}) &= \left\{\boldsymbol{y} : \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}\right\} = \left\{\boldsymbol{y} : \boldsymbol{y} = \boldsymbol{U}_1(\boldsymbol{\Sigma}_r\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{x})\right\} \subset \mathcal{R}(\boldsymbol{U}_1),\\ \mathcal{R}(\boldsymbol{U}_1) &= \left\{\boldsymbol{y} : \boldsymbol{y} = \boldsymbol{U}_1\boldsymbol{z}\right\} = \left\{\boldsymbol{y} : \boldsymbol{y} = (\boldsymbol{U}_1\boldsymbol{\Sigma}_r\boldsymbol{V}_1^{\mathrm{T}})(\boldsymbol{V}_1\boldsymbol{\Sigma}_r^{-1}\boldsymbol{z})\right\}\\ &= \left\{\boldsymbol{y} : \boldsymbol{y} = \boldsymbol{A}(\boldsymbol{V}_1\boldsymbol{\Sigma}_r^{-1}\boldsymbol{z})\right\} \subset \mathcal{R}(\boldsymbol{A}),\end{aligned}$$

故 $\mathcal{R}(\boldsymbol{A}) = \mathcal{R}(\boldsymbol{U}_1) = \operatorname{span}\left\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_r\right\}$.

(5) 直接计算式 (1.11), 得

$$
\begin{aligned}
\boldsymbol{A} &= [\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_r] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{H}} \\ \vdots \\ \boldsymbol{v}_r^{\mathrm{H}} \end{bmatrix} \\
&= [\sigma_1\boldsymbol{u}_1, \sigma_2\boldsymbol{u}_2, \cdots, \sigma_r\boldsymbol{u}_r] \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{H}} \\ \vdots \\ \boldsymbol{v}_r^{\mathrm{H}} \end{bmatrix} \\
&= \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{H}}.
\end{aligned}
$$

证毕. $\square$

下面的定理表明奇异值可以刻画一个矩阵与更低秩的矩阵之间的距离.

定理 1.10 设 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}$, 它的奇异值分解由定理 1.8 给出,

其中 $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ 为其非零奇异值, 则有

$$\min_{\operatorname{rank}(\boldsymbol{B})=k} \|\boldsymbol{A}-\boldsymbol{B}\|_2 = \|\boldsymbol{A}-\boldsymbol{A}_k\|_2 = \sigma_{k+1}, \quad 0 \leqslant k \leqslant r-1, \tag{1.12}$$

式中:

$$\boldsymbol{A}_k = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{H}}, \quad k>0, \ \boldsymbol{A}_0 = \boldsymbol{O}. \tag{1.13}$$

证明 当 $k=0$ 时, $\|\boldsymbol{A}\|_2 = \sigma_1$, 结论显然成立. 设 $k>0$, 由 $\boldsymbol{A}_k$ 的定义及 $\boldsymbol{A}$ 的奇异值分解, 有

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}_k\boldsymbol{V} = \operatorname{diag}(\sigma_1, \cdots, \sigma_k, 0, \cdots, 0),$$

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V} = \operatorname{diag}(\sigma_1, \cdots, \sigma_k, \sigma_{k+1}, \cdots, \sigma_n).$$

于是有

$$\boldsymbol{U}^{\mathrm{H}}(\boldsymbol{A}-\boldsymbol{A}_k)\boldsymbol{V} = \operatorname{diag}(0, \cdots, 0, \sigma_{k+1}, \cdots, \sigma_n).$$

注意到 σ_{k+1} 是 $\boldsymbol{A}-\boldsymbol{A}_k$ 的最大奇异值, 立即有 $\|\boldsymbol{A}-\boldsymbol{A}_k\|_2=\sigma_{k+1}$. 由于 $\operatorname{rank}(\boldsymbol{A}_k)=k$, 故

$$\min_{\operatorname{rank}(\boldsymbol{B})=k}\|\boldsymbol{A}-\boldsymbol{B}\|_2\leqslant\|\boldsymbol{A}-\boldsymbol{A}_k\|_2=\sigma_{k+1}.$$

下证任意秩为 k 的 $m\times n$ 阶矩阵 $\boldsymbol{B}$, 均有 $\|\boldsymbol{A}-\boldsymbol{B}\|_2\geqslant\sigma_{k+1}$. 由于 $\operatorname{rank}(\boldsymbol{B})=k$, 故 $\boldsymbol{B}$ 的零空间 $\mathcal{N}(\boldsymbol{B})=\{\boldsymbol{x}:\boldsymbol{B}\boldsymbol{x}=\mathbf{0}\}$ 的维数为 $n-k$, 故存在单位正交向量 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{n-k}$ 使得 $\mathcal{N}(\boldsymbol{B})=\operatorname{span}\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{n-k}\}$. 利用维数定理可知

$$\mathcal{S}:=\operatorname{span}\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{n-k}\}\cap\operatorname{span}\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{k+1}\}\neq\{\mathbf{0}\}.$$

在 $\mathcal{S}$ 中取一非零单位向量 $\boldsymbol{z}$, 即 $\boldsymbol{z}$ 满足 $\|\boldsymbol{z}\|_2=1$, $\boldsymbol{B}\boldsymbol{z}=\mathbf{0}$ 及

$$\boldsymbol{z}=\sum_{i=1}^{k+1}\alpha_i\boldsymbol{v}_i,\quad \boldsymbol{A}\boldsymbol{z}=\sum_{i=1}^{k+1}\alpha_i\boldsymbol{A}\boldsymbol{v}_i=\sum_{i=1}^{k+1}\sigma_i\alpha_i\boldsymbol{u}_i.$$

Back
Close

注意到 $\|\boldsymbol{z}\|_2^2 = \sum\limits_{i=1}^{k+1} \alpha_i^2 = 1$, 立即有

$$\|\boldsymbol{A}-\boldsymbol{B}\|_2^2 \geqslant \|(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{z}\|_2^2 = \|\boldsymbol{A}\boldsymbol{z}\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 \alpha_i^2 \geqslant \sigma_{k+1}^2.$$

证毕. □

§1.2.4 矩阵的极分解和满秩分解

本节讨论矩阵的极分解和满秩分解. 极分解是指将一个复方阵分解为一个酉矩阵和 Hermite 半正定矩阵的乘积. 满秩分解是指将一个非零矩阵分解成列满秩矩阵和行满秩矩阵的乘积, 它是研究广义逆矩阵的重要工具之一. 对于极分解有下面的定理.

定理 1.11 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$. 则存在 n 阶酉矩阵 $\boldsymbol{U}$ 和 n 阶

Hermite 半正定矩阵 $\boldsymbol{H}$, 使得

$$\boldsymbol{A}=\boldsymbol{H}\boldsymbol{U}. \tag{1.14}$$

若 $\boldsymbol{A}$ 是非奇异矩阵, 则上述分解是唯一的, 此时, $\boldsymbol{H}$ 是 Hermite 正定矩阵.

证明 由奇异值分解定理可知, 存在酉矩阵 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 使得

$$\boldsymbol{A}=\boldsymbol{P}\begin{bmatrix}\boldsymbol{\Sigma}_r & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}\boldsymbol{Q}^{\mathrm{H}},$$

式中: $\boldsymbol{\Sigma}_r=\operatorname{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$, $\sigma_i>0$ 为 $\boldsymbol{A}$ 的奇异值; $r=\operatorname{rank}(\boldsymbol{A})$. 将上式改写为

$$\boldsymbol{A}=\boldsymbol{P}\begin{bmatrix}\boldsymbol{\Sigma}_r & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}\boldsymbol{P}^{\mathrm{H}}\boldsymbol{P}\boldsymbol{Q}^{\mathrm{H}},$$

令

$$\boldsymbol{H}=\boldsymbol{P}\begin{bmatrix}\boldsymbol{\Sigma}_r & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}\boldsymbol{P}^{\mathrm{H}},\quad \boldsymbol{U}=\boldsymbol{P}\boldsymbol{Q}^{\mathrm{H}},$$

容易验证 $\boldsymbol{H}$ 为 Hermite 半正定矩阵, 而 $\boldsymbol{U}$ 为酉矩阵.

下面证明当 $\boldsymbol{A}$ 非奇异时分解是唯一的. 事实上, 若 $\boldsymbol{A}$ 还有极分解 $\boldsymbol{A}=\boldsymbol{H}_1\boldsymbol{U}_1$, 即

$$\boldsymbol{A}=\boldsymbol{H}\boldsymbol{U}=\boldsymbol{H}_1\boldsymbol{U}_1,$$

式中: $\boldsymbol{U},\boldsymbol{U}_1$ 为酉矩阵; $\boldsymbol{H},\boldsymbol{H}_1$ 为 Hermite 正定矩阵. 注意到

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{H}^2=\boldsymbol{H}_1^2,$$

则由定理 1.2 可知 $\boldsymbol{H}_1=\boldsymbol{H}$. 于是, $\boldsymbol{U}_1=\boldsymbol{H}_1^{-1}\boldsymbol{A}=\boldsymbol{H}^{-1}\boldsymbol{A}=\boldsymbol{U}$. 证毕. $\square$

注 1.3 若定理 1.11 中的矩阵 $\boldsymbol{A}$ 是实方阵, 则分解式中的 $\boldsymbol{U}$ 和 $\boldsymbol{H}$ 分别为正交矩阵和对称半正定矩阵.

下面讨论满秩分解.

定义 1.11 设 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}(r > 0)$, 若存在列满秩矩阵 $\boldsymbol{F} \in \mathbb{C}_r^{m\times r}$ 和行满秩矩阵 $\boldsymbol{G} \in \mathbb{C}_r^{r\times n}$, 使得 $\boldsymbol{A} = \boldsymbol{FG}$, 则称 $\boldsymbol{FG}$ 为 $\boldsymbol{A}$ 的一个满秩分解.

定理 1.12 设 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}(r > 0)$, 则存在 $\boldsymbol{F} \in \mathbb{C}_r^{m\times r}$ 和 $\boldsymbol{G} \in \mathbb{C}_r^{r\times n}$, 使得 $\boldsymbol{A} = \boldsymbol{FG}$.

证明 利用初等变换, 可将 $\boldsymbol{A}$ 化为阶梯形矩阵

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{G} \\ \boldsymbol{O} \end{bmatrix}, \quad (\boldsymbol{G} \in \mathbb{C}_r^{r\times n}),$$

Back

Close

即存在有限个初等矩阵的乘积 $\boldsymbol{P} \in \mathbb{C}^{m\times m}$, 使得 $\boldsymbol{PA} = \boldsymbol{B}$. 由于 $\boldsymbol{P}$ 可逆, 划分

$$\boldsymbol{P}^{-1} = [\boldsymbol{F},\ \boldsymbol{S}], \quad (\boldsymbol{F} \in \mathbb{C}_r^{m\times r}, \boldsymbol{S} \in \mathbb{C}^{m\times(m-r)}).$$

于是有

$$\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{B} = [\boldsymbol{F},\ \boldsymbol{S}]\begin{bmatrix} \boldsymbol{G} \\ \boldsymbol{O} \end{bmatrix} = \boldsymbol{FG}.$$

证毕. □

例 1.1 求下列矩阵的满秩分解:

$$\boldsymbol{A} = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 2 & -1 & 1 \\ 2 & 2 & -2 & -1 \end{bmatrix}.$$

解 由

$$[\boldsymbol{A},\ \boldsymbol{I}] = \left[\begin{array}{cccc:ccc} -1 & 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 2 & -1 & 1 & 0 & 1 & 0 \\ 2 & 2 & -2 & -1 & 0 & 0 & 1 \end{array}\right] \xrightarrow{\text{行}} \left[\begin{array}{cccc:ccc} -1 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 3 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 1 \end{array}\right],$$

可知

$$\boldsymbol{P} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix}, \quad \boldsymbol{P}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & 1 & 1 \end{bmatrix},$$

于是有

$$\boldsymbol{F} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ -2 & 1 \end{bmatrix}, \quad \boldsymbol{G} = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 0 & 2 & 0 & 3 \end{bmatrix}, \quad \boldsymbol{A} = \boldsymbol{F}\boldsymbol{G}.$$

定理 1.12 提供的算法求矩阵的满秩分解时需要计算矩阵 $\boldsymbol{P}$ 及 $\boldsymbol{P}^{-1}$ 的前 r 列, 这在高维情形计算量是巨大的. 下面介绍一种避免求逆矩阵的方法.

定义 1.12 设 $\boldsymbol{B} \in \mathbb{C}_r^{m\times n}(r>0)$ 满足: ① $\boldsymbol{B}$ 的后 $m-r$ 行上的元素都是零; ② $\boldsymbol{B}$ 中有 r 个列 (c_1 列, $\cdots, c_r$ 列) 构成 $\boldsymbol{I}_m$ 的前 r 个列. 则称 $\boldsymbol{B}$ 为拟 Hermite 标准形.

定理 1.13 设 $\boldsymbol{A} \in \mathbb{C}_r^{m\times n}(r>0)$ 经过初等行变换化为拟 Hermite 标准形 $\boldsymbol{B}$, 那么 $\boldsymbol{A}$ 有满秩分解 $\boldsymbol{A}=\boldsymbol{F}\boldsymbol{G}$, 其中 $\boldsymbol{F}$ 是由 "$\boldsymbol{A}$ 的 c_1 列, $\cdots, c_r$ 列" 构成的矩阵, $\boldsymbol{G}$ 是由 "$\boldsymbol{B}$ 的前 r 行" 构成的矩阵.

证明 经过初等行变换将 $\boldsymbol{A}$ 化为拟 Hermite 标准形 $\boldsymbol{B}$, 等价于存在可逆矩阵 $\boldsymbol{P} \in \mathbb{C}^{m\times m}$, 使得 $\boldsymbol{P}\boldsymbol{A}=\boldsymbol{B}$, 或者 $\boldsymbol{A}=\boldsymbol{P}^{-1}\boldsymbol{B}$.

根据 $\boldsymbol{B}$ 中的列标 $c_1, \cdots, c_r$ 构造 n 阶置换矩阵

$$\boldsymbol{P}_1 = [\boldsymbol{e}_{c_1}, \cdots, \boldsymbol{e}_{c_r}, \boldsymbol{e}_{c_{r+1}}, \cdots, \boldsymbol{e}_{c_n}],$$

并对矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 按列分块:

$$\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n], \quad \boldsymbol{B} = [\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_n].$$

于是可得

$$\begin{aligned} \boldsymbol{A}\boldsymbol{P}_1 &= [\boldsymbol{a}_{c_1}, \cdots, \boldsymbol{a}_{c_r}, \boldsymbol{a}_{c_{r+1}}, \cdots, \boldsymbol{a}_{c_n}], \\ \boldsymbol{B}\boldsymbol{P}_1 &= [\boldsymbol{b}_{c_1}, \cdots, \boldsymbol{b}_{c_r}, \boldsymbol{b}_{c_{r+1}}, \cdots, \boldsymbol{b}_{c_n}] = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{B}_{12} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}. \end{aligned}$$

划分 $\boldsymbol{P}^{-1} = [\boldsymbol{F}, \boldsymbol{S}]$, 其中 $\boldsymbol{F} \in \mathbb{C}_r^{m\times r}$, $\boldsymbol{S} \in \mathbb{C}^{m\times(m-r)}$. 根据定理 1.12 可得满秩分解 $\boldsymbol{A} = \boldsymbol{F}\boldsymbol{G}$, 且 $\boldsymbol{G}$ 是由“$\boldsymbol{B}$ 的前 r 行”构成的矩阵.

Back
Close

由于

$$\boldsymbol{A}\boldsymbol{P}_1 = \boldsymbol{P}^{-1}(\boldsymbol{B}\boldsymbol{P}_1) = [\boldsymbol{F},\ \boldsymbol{S}]\begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{B}_{12} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} = [\boldsymbol{F},\ \boldsymbol{F}\boldsymbol{B}_{12}].$$

所以 $\boldsymbol{F}$ 是由“$\boldsymbol{A}\boldsymbol{P}_1$ 的前 r 列”构成的矩阵，也就是由“$\boldsymbol{A}$ 的 c_1 列，$\cdots$，c_r 列”构成的矩阵. 证毕. □

例 1.2 求下列矩阵的满秩分解:

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 0 & 2 \\ 0 & 0 & 1 & 2 \\ 2 & 1 & 1 & 4 \end{bmatrix}.$$

解 注意到

$$\boldsymbol{A} \xrightarrow{\text{行}} \begin{bmatrix} 2 & 1 & 0 & 2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

为拟 Hermite 标准形, 且 $c_1 = 2$, $c_2 = 3$, 故

$$\boldsymbol{F} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, \quad \boldsymbol{G} = \begin{bmatrix} 2 & 1 & 0 & 2 \\ 0 & 0 & 1 & 2 \end{bmatrix}, \quad \boldsymbol{A} = \boldsymbol{F}\boldsymbol{G}.$$

§1.3 向量和矩阵的范数

在许多实际问题中, 常需对同一线性空间中的向量 (或矩阵) 引入作为它们 “大小” 的一种度量, 进而比较两个向量 (或矩阵) 的 “接近” 程度. 引入这种体现其 “大小” 的量就是范数, 它们在理论与实际应用中都占有重要的地位.

§1.3.1 向量内积与向量范数

定义 1.13 设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in\mathbb{C}^n$, $\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}\in\mathbb{C}^n$, 称复数

$$(\boldsymbol{x},\boldsymbol{y}):=\boldsymbol{y}^{\mathrm{H}}\boldsymbol{x}=\sum_{i=1}^{n}x_i\bar{y}_i$$

Back
Close

为向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的 (欧几里得) 内积或数量积. 而称

$$\|\boldsymbol{x}\|_2 = (\boldsymbol{x}, \boldsymbol{x})^{\frac{1}{2}} = \left(\sum_{i=1}^{n} |x_i|^2\right)^{\frac{1}{2}} \tag{1.15}$$

为向量 $\boldsymbol{x}$ 的 (欧几里得) 范数或 2–范数.

内积的一个重要性质是

$$(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{y}),\ \boldsymbol{A} \in \mathbb{C}^{m\times n},\ \boldsymbol{x} \in \mathbb{C}^{n},\ \boldsymbol{y} \in \mathbb{C}^{m}. \tag{1.16}$$

$\mathbb{C}^n$ 中的欧几里得内积与欧几里得范数还具有下列性质.

性质 1.6 设 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$, 则

(1) $(\boldsymbol{x}, \boldsymbol{x}) = 0$ 当且仅当 $\boldsymbol{x} = \mathbf{0}$.

(2) $(\lambda\boldsymbol{x}, \boldsymbol{y}) = \lambda(\boldsymbol{x}, \boldsymbol{y})$, $(\boldsymbol{x}, \lambda\boldsymbol{y}) = \bar{\lambda}(\boldsymbol{x}, \boldsymbol{y})$, $\forall\, \lambda \in \mathbb{C}$.

(3) $(\boldsymbol{x}, \boldsymbol{y}) = \overline{(\boldsymbol{y}, \boldsymbol{x})}$.

(4) $(\boldsymbol{x}_1+\boldsymbol{x}_2,\boldsymbol{y})=(\boldsymbol{x}_1,\boldsymbol{y})+(\boldsymbol{x}_2,\boldsymbol{y})$.

(5) Cauchy–Schwarz 不等式: $|(\boldsymbol{x},\boldsymbol{y})|\leqslant\|\boldsymbol{x}\|_2\cdot\|\boldsymbol{y}\|_2$.

(6) 三角不等式: $\|\boldsymbol{x}+\boldsymbol{y}\|_2\leqslant\|\boldsymbol{x}\|_2+\|\boldsymbol{y}\|_2$.

由 Schwarz 不等式容易证明

$$\|\boldsymbol{x}\|_2=\max_{\|\boldsymbol{y}\|_2=1}|\boldsymbol{y}^{\mathrm{H}}\boldsymbol{x}|,\quad \boldsymbol{x}\in\mathbb{C}^n. \tag{1.17}$$

下面给出向量范数的一般定义.

定义 1.14 给定 $\mathbb{C}^n$ 中的某个实值函数 $\mathfrak{N}(\boldsymbol{x})=\|\boldsymbol{x}\|$. 若对任意的 $\boldsymbol{x},\boldsymbol{y}\in\mathbb{C}^n$ 有

(1) $\|\boldsymbol{x}\|\geqslant 0$, 且 $\|\boldsymbol{x}\|=0$ 当且仅当 $\boldsymbol{x}=\mathbf{0}$.

(2) $\|\lambda\boldsymbol{x}\|=|\lambda|\cdot\|\boldsymbol{x}\|$, $\forall\,\lambda\in\mathbb{C}$.

(3) $\|\boldsymbol{x}+\boldsymbol{y}\|\leqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|$.

则称 $\|\boldsymbol{x}\|$ 为向量 $\boldsymbol{x}$ 的范数. 定义了范数的线性空间称为赋范线性空间.

常用的向量范数有

(1) $\|\boldsymbol{x}\|_\infty = \max\limits_{1\leqslant i\leqslant n} |x_i| \quad (\infty$–范数).

(2) $\|\boldsymbol{x}\|_1 = \sum\limits_{i=1}^{n} |x_i| \quad (1$–范数).

(3) $\|\boldsymbol{x}\|_2 = \left(\sum\limits_{i=1}^{n} |x_i|^2\right)^{1/2} \quad (2$–范数).

(4) $\|\boldsymbol{x}\|_p = \left(\sum\limits_{i=1}^{n} |x_i|^p\right)^{1/p} \quad (p$–范数).

不难证明, 当 $p \to \infty$ 时, 对于任意的 $\boldsymbol{x} \in \mathbb{C}^n$, 都有 $\|\boldsymbol{x}\|_p \to \|\boldsymbol{x}\|_\infty$. 此外, 由于 $\big|\|\boldsymbol{x}\| - \|\boldsymbol{y}\|\big| \leqslant \|\boldsymbol{x} - \boldsymbol{y}\|$, 故向量范数是 $\mathbb{C}^n$ 中的连续函数.

性质 1.7 (向量范数的连续性) $\mathbb{C}^n$ 中的向量范数 $\|\boldsymbol{x}\|$ 是 $\mathbb{C}^n$ 中的连续函数.

性质 1.8 (向量范数的等价性) 设 $\|\boldsymbol{x}\|$ 和 $\|\boldsymbol{x}\|'$ 是 $\mathbb{C}^n$ 中的任意两种范数, 则成立

$$c_1\|\boldsymbol{x}\| \leqslant \|\boldsymbol{x}\|' \leqslant c_2\|\boldsymbol{x}\|, \quad \boldsymbol{x} \in \mathbb{C}^n, \tag{1.18}$$

式中: $c_1, c_2 > 0$ 为与向量 $\boldsymbol{x}$ 无关的常数.

定义 1.15 设 $\{\boldsymbol{x}^{(k)} = (x_1^{(k)}, x_2^{(k)}, \cdots, x_n^{(k)})^{\mathrm{T}}\}_{k=0}^{\infty}$ 是 $\mathbb{C}^n$ 中的向量序列, $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbb{C}^n$. 若

$$\lim_{k\to\infty} x_i^{(k)} = x_i, \quad i = 1, 2, \cdots, n,$$

则称序列 $\{\boldsymbol{x}^{(k)}\}$ 收敛于 $\boldsymbol{x}$, 记为 $\lim\limits_{k\to\infty} \boldsymbol{x}^{(k)} = \boldsymbol{x}$.

利用范数的等价性可以得到向量序列收敛的充分必要条件.

定理 1.14 $\lim\limits_{k\to\infty} \boldsymbol{x}^{(k)} = \boldsymbol{x}$ 当且仅当 $\lim\limits_{k\to\infty} \|\boldsymbol{x}^{(k)} - \boldsymbol{x}\| = 0$, 其中 $\|\cdot\|$ 是 $\mathbb{C}^n$ 中任意的向量范数.

§1.3.2 矩阵范数与内积

将 $m \times n$ 阶矩阵 $\boldsymbol{A}$ 看作线性空间 $\mathbb{C}^{m\times n}$ 中的元素, 则完全可以按照定义 1.14 的方式引入矩阵的范数. 其中最常用的是与向量 2–范数相对应的范数

$$\|\boldsymbol{A}\|_{\mathrm{F}} = \Big(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\Big)^{1/2} \tag{1.19}$$

称为矩阵 $\boldsymbol{A}$ 的 Frobenius 范数, 简称 F–范数.

利用奇异值分解定理, 容易证明

$$\|\boldsymbol{A}\|_{\mathrm{F}} = \Big(\sum_{i=1}^{n}\sigma_i^2\Big)^{1/2},$$

式中: $\sigma_1, \sigma_2, \cdots, \sigma_n$ 为矩阵 $\boldsymbol{A}$ 奇异值.

类似于向量范数的定义, 可以给出矩阵范数的一般定义.

定义 1.16 给定 $\mathbb{C}^{m\times n}$ 中的某个实值函数 $\mathfrak{N}(\boldsymbol{A}) = \|\boldsymbol{A}\|$. 若对任意的 $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{m\times n}$ 有

(1) $\|\boldsymbol{A}\| \geqslant 0$, 且 $\|\boldsymbol{A}\| = 0$ 当且仅当 $\boldsymbol{A} = \boldsymbol{O}$.

(2) $\|\lambda \boldsymbol{A}\| = |\lambda| \cdot \|\boldsymbol{A}\|, \forall\, \lambda \in \mathbb{C}$.

(3) $\|\boldsymbol{A} + \boldsymbol{B}\| \leqslant \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$.

则称 $\|\boldsymbol{A}\|$ 为矩阵 $\boldsymbol{A}$ 的范数. 若矩阵范数满足

$$\|\boldsymbol{A}\boldsymbol{x}\| \leqslant \|\boldsymbol{A}\| \cdot \|\boldsymbol{x}\|, \quad \forall\, \boldsymbol{x} \in \mathbb{C}^n,\ \boldsymbol{A} \in \mathbb{C}^{m\times n},$$

则称矩阵范数 $\|\cdot\|$ 和向量范数 $\|\cdot\|$ 是相容的, 其中 $\|\boldsymbol{A}\boldsymbol{x}\|$ 和 $\|\boldsymbol{x}\|$ 分别是 $\mathbb{C}^m$ 和 $\mathbb{C}^n$ 中的向量范数.

跟向量范数的等价性一样, 可以证明 $\mathbb{C}^{m\times n}$ 中的任意两个矩阵范数也等价, 即存在与矩阵 $\boldsymbol{A}$ 无关的常数 $c_1, c_2 > 0$ 使得

$$c_1\|\boldsymbol{A}\| \leqslant \|\boldsymbol{A}\|' \leqslant c_2\|\boldsymbol{A}\|.$$

定义 1.17 设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$, 给定一种向量范数 $\|\cdot\|$, 定义矩阵范数

$$\|\boldsymbol{A}\| = \max_{\|\boldsymbol{x}\|\neq 0} \frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \max_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\|. \tag{1.20}$$

称 $\|\boldsymbol{A}\|$ 为 $\mathbb{C}^{m\times n}$ 中矩阵 $\boldsymbol{A}$ 的算子范数, 或称由该向量范数诱导出来的矩阵范数.

容易证明, 矩阵的算子范数与相应的向量范数是相容的, 且

$$\|\boldsymbol{A}\boldsymbol{B}\| \leqslant \|\boldsymbol{A}\| \cdot \|\boldsymbol{B}\|, \quad \boldsymbol{A} \in \mathbb{C}^{m\times n}, \ \boldsymbol{B} \in \mathbb{C}^{n\times p},$$

其中矩阵范数都是由某个向量范数诱导出来的算子范数.

给定 $\mathbb{C}^n$ 中的向量 p–范数, 可以诱导出相应的矩阵 p–算子范数, 记为 $\|\cdot\|_p$. 下面是三种常用的矩阵算子范数.

定理 1.15 设 $\boldsymbol{A} = (a_{ij}) \in \mathbb{C}^{m\times n}$, 则

(1) $\|\boldsymbol{A}\|_\infty = \max\limits_{1\leqslant i\leqslant m} \sum\limits_{j=1}^{n} |a_{ij}|$, 称为行和范数或 ∞–范数.

(2) $\|\boldsymbol{A}\|_1 = \max\limits_{1\leqslant j\leqslant n} \sum\limits_{i=1}^{m} |a_{ij}|$, 称为列和范数或 1–范数.

(3) $\|\boldsymbol{A}\|_2 = \sigma_1 = \sqrt{\lambda_{\max}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})}$, 称为谱范数或 2–范数, 其中 σ_1 是矩阵 $\boldsymbol{A}$ 的最大奇异值, $\lambda_{\max}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$ 表示矩阵 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的最大特征值.

证明 **先证明** (1). **记** $\sigma = \max\limits_{1\leqslant i\leqslant m} \sum\limits_{j=1}^{n} |a_{ij}|$. **则对任意满足** $\|\boldsymbol{x}\|_\infty = 1$ **的向量** $\boldsymbol{x} \in \mathbb{C}^n$, **有**

$$\begin{aligned}\|\boldsymbol{A}\boldsymbol{x}\|_\infty &= \max_{1\leqslant i\leqslant m} \left|\sum_{j=1}^{n} a_{ij}x_j\right| \leqslant \max_{1\leqslant i\leqslant m} \sum_{j=1}^{n} |a_{ij}|\cdot|x_j| \\ &\leqslant \left(\max_{1\leqslant i\leqslant m} \sum_{j=1}^{n} |a_{ij}|\right)\|\boldsymbol{x}\|_\infty = \sigma. \qquad (1.21)\end{aligned}$$

另外, 不妨设 $\sigma = \sum\limits_{j=1}^{n} |a_{i_0 j}|$. 取

$$\widetilde{\boldsymbol{x}} = \big(\mathrm{sign}(a_{i_0 1}), \mathrm{sign}(a_{i_0 2}), \cdots, \mathrm{sign}(a_{i_0 n})\big)^{\mathrm{T}},$$

则 $\boldsymbol{A} \neq \boldsymbol{O}$ 蕴含着 $\|\widetilde{\boldsymbol{x}}\|_\infty = 1$. 由 $|a| = \mathrm{sign}(a)a$, 有

$$\|\boldsymbol{A}\widetilde{\boldsymbol{x}}\|_\infty = \max_{1\leqslant i\leqslant m} \left|\sum_{j=1}^{n} a_{ij}\widetilde{x}_j\right| \geqslant \left|\sum_{j=1}^{n} a_{i_0 j}\widetilde{x}_j\right| = \sum_{j=1}^{n} |a_{i_0 j}| = \sigma. \quad (1.22)$$

结论 (2) 的证明与结论 (1) 类似. 最后给出结论 (3) 的证明. 注意到 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} \in \mathbb{C}^{n\times n}$ 是 Hermite 半正定矩阵, 故其特征值为 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$, 相应的单位正交特征向量为 $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \cdots, \boldsymbol{\xi}_n \in \mathbb{C}^n$. 因此, 对任意满足 $\|\boldsymbol{x}\|_2 = 1$ 的向量 $\boldsymbol{x} \in \mathbb{C}^n$, 有

$$\boldsymbol{x} = \sum_{i=1}^{n} c_i \boldsymbol{\xi}_i, \quad \sum_{i=1}^{n} c_i^2 = 1.$$

于是有

$$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} = \left(\sum_{i=1}^{n} c_i\boldsymbol{\xi}_i\right)^{\mathrm{H}}\sum_{j=1}^{n} c_j\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{\xi}_j$$

$$= \left(\sum_{i=1}^{n} c_i\boldsymbol{\xi}_i\right)^{\mathrm{H}}\sum_{j=1}^{n} c_j\lambda_j\boldsymbol{\xi}_j$$

$$= \sum_{i=1}^{n} c_i^2\lambda_i\boldsymbol{\xi}_i^{\mathrm{H}}\boldsymbol{\xi}_i = \sum_{i=1}^{n} c_i^2\lambda_i \leqslant \lambda_1.$$

注意到

$$\|\boldsymbol{A}\|_2 = \max_{\|\boldsymbol{x}\|_2=1}\|\boldsymbol{A}\boldsymbol{x}\|_2 = \max_{\|\boldsymbol{x}\|_2=1}\left(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}\right)^{1/2},$$

故 $\|\boldsymbol{A}\|_2 \leqslant \sqrt{\lambda_1}$. 取 $\boldsymbol{x} = \boldsymbol{\xi}_1$, 则有

$$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{\xi}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{\xi}_1 = \lambda_1\boldsymbol{\xi}_1^{\mathrm{H}}\boldsymbol{\xi}_1 = \lambda_1,$$

故

$$\|\boldsymbol{A}\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{A}\boldsymbol{x}\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \left(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}\right)^{1/2}$$

$$\geqslant \left(\boldsymbol{\xi}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{\xi}_1\right)^{1/2} = \left(\lambda_1\boldsymbol{\xi}_1^{\mathrm{H}}\boldsymbol{\xi}_1\right)^{1/2} = \sqrt{\lambda_1}.$$

因此 $\|\boldsymbol{A}\|_2 = \sqrt{\lambda_1} = \sqrt{\lambda_{\max}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})}$. 证毕. □

与矩阵的行和范数及列和范数相比, 矩阵的谱范数显得不易计算, 但它有良好的分析性质.

定理 1.16 设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$, 则

(1) $\|\boldsymbol{A}\|_2 = \max\limits_{\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2=1} |\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}|$.

(2) $\|\boldsymbol{A}^{\mathrm{H}}\|_2 = \|\boldsymbol{A}^{\mathrm{T}}\|_2 = \|\boldsymbol{A}\|_2$.

(3) $\|\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\|_2 = \|\boldsymbol{A}\|_2^2$.

(4) $\|\boldsymbol{A}\|_2^2 \leqslant \|\boldsymbol{A}\|_1 \cdot \|\boldsymbol{A}\|_\infty$.

(5) $\|\boldsymbol{A}\|_2 \leqslant \|\boldsymbol{A}\|_{\mathrm{F}}$.

(6) 设 $\boldsymbol{U} \in \mathbb{C}^{m\times m}$ 和 $\boldsymbol{V} \in \mathbb{C}^{n\times n}$ 满足 $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} = \boldsymbol{I}_m$ 和 $\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V} = \boldsymbol{I}_n$, 则 $\|\boldsymbol{U}\boldsymbol{A}\boldsymbol{V}\|_2 = \|\boldsymbol{A}\|_2$.

证明 由算子范数的定义, 存在 $\boldsymbol{x}_0 \in \mathbb{C}^n$ 满足 $\|\boldsymbol{x}_0\|_2 = 1$ 和 $\|\boldsymbol{A}\|_2 = \|\boldsymbol{A}\boldsymbol{x}_0\|_2$. 由式 (1.17) 知,

$$\|\boldsymbol{A}\boldsymbol{x}_0\|_2 = \max_{\|\boldsymbol{y}\|_2=1} |\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}_0|,$$

这意味着成立

$$\|\boldsymbol{A}\|_2 = \|\boldsymbol{A}\boldsymbol{x}_0\|_2 \leqslant \max_{\|\boldsymbol{x}\|_2=1,\, \|\boldsymbol{y}\|_2=1} |\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}|.$$

另外, 由 Schwarz 不等式及范数的相容性, 有

$$|\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}| \leqslant \|\boldsymbol{y}\|_2 \cdot \|\boldsymbol{A}\boldsymbol{x}\|_2 \leqslant \|\boldsymbol{y}\|_2 \cdot \|\boldsymbol{A}\|_2 \cdot \|\boldsymbol{x}\|_2,$$

故

$$\max_{\|\boldsymbol{x}\|_2=1,\,\|\boldsymbol{y}\|_2=1} |\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}| \leqslant \|\boldsymbol{A}\|_2,$$

这就证明了结论 (1) .

由结论 (1) 有

$$\|\boldsymbol{A}^{\mathrm{T}}\|_2 = \max_{\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2=1} |\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}| = \max_{\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2=1} |\bar{\boldsymbol{x}}^{\mathrm{H}}\boldsymbol{A}\bar{\boldsymbol{y}}|$$

$$= \max_{\|\bar{\boldsymbol{x}}\|_2=\|\bar{\boldsymbol{y}}\|_2=1} |\bar{\boldsymbol{x}}^{\mathrm{H}}\boldsymbol{A}\bar{\boldsymbol{y}}| = \|\boldsymbol{A}\|_2,$$

式中: $\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}$ 为 $\boldsymbol{x}, \boldsymbol{y}$ 的共轭向量, 这里用到了一个标量与其转置相等.

同理, 有

$$\|\boldsymbol{A}^{\mathrm{H}}\|_2 = \max_{\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2=1} |\boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}| = \max_{\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2=1} |\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{y}| = \|\boldsymbol{A}\|_2,$$

这里用到了一个复数的模与其共轭转置的模相等. 至此证明了结论 (2).

注意到 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 是 Hermite 矩阵, 故由谱范数的定义, 有

$$\begin{aligned}\|\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\|_2 &= \sqrt{\lambda_{\max}\big((\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{\mathrm{H}}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})\big)} = \sqrt{\lambda_{\max}\big((\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^2\big)}\\ &= \lambda_{\max}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = \|\boldsymbol{A}\|_2^2,\end{aligned}$$

结论 (3) 得证.

由于 $\|\boldsymbol{A}\|_2^2$ 是 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的最大特征值, 设其相应的特征向量为 $\boldsymbol{x} \neq \boldsymbol{0}$, 即 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} = \|\boldsymbol{A}\|_2^2\boldsymbol{x}$, 两边取 1–范数, 得

$$\begin{aligned}\|\boldsymbol{A}\|_2^2\|\boldsymbol{x}\|_1 &= \big\|\|\boldsymbol{A}\|_2^2\boldsymbol{x}\big\|_1 = \|\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}\|_1\\ &\leqslant \|\boldsymbol{A}^{\mathrm{H}}\|_1 \cdot \|\boldsymbol{A}\|_1 \cdot \|\boldsymbol{x}\|_1 = \|\boldsymbol{A}\|_\infty \cdot \|\boldsymbol{A}\|_1 \cdot \|\boldsymbol{x}\|_1,\end{aligned}$$

两边除以 $\|\boldsymbol{x}\|_1$ 即得结论 (4).

由于 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 是 Hermite 半正定矩阵, 故其特征值 $\lambda_s(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$, $s=1,2,\cdots,n$ 均非负, 且满足

$$\sum_{s=1}^{n}\lambda_s(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})=\mathrm{tr}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})=\sum_{s=1}^{n}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right)_{ss},$$

从而有

$$\begin{aligned}\|\boldsymbol{A}\|_2^2=\lambda_{\max}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})&\leqslant\sum_{s=1}^{n}\lambda_s(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})\\&=\sum_{s=1}^{n}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\right)_{ss}=\sum_{s=1}^{n}\left(\sum_{i=1}^{m}|a_{is}|^2\right)=\|\boldsymbol{A}\|_{\mathrm{F}}^2,\end{aligned}$$

两边开平方即得结论 (5).

最后证明结论 (6). 注意到对任意的 n 阶酉矩阵 $\boldsymbol{U}$ 和 n 维列向量 $\boldsymbol{x}$, 有

$$\|\boldsymbol{U}\boldsymbol{x}\|_2=\sqrt{(\boldsymbol{U}\boldsymbol{x})^{\mathrm{H}}(\boldsymbol{U}\boldsymbol{x})}=\sqrt{\boldsymbol{x}^{\mathrm{H}}(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U})\boldsymbol{x}}=\sqrt{\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}}=\|\boldsymbol{x}\|_2,$$

即酉矩阵保持列向量的 2–范数不变. 由此, 有

$$\begin{aligned}\|\boldsymbol{UAV}\|_2 &= \max_{\|\boldsymbol{x}\|_2=1}\|\boldsymbol{UAVx}\|_2 = \max_{\|\boldsymbol{y}\|_2=1}\|\boldsymbol{UAy}\|_2 \\ &= \max_{\|\boldsymbol{y}\|_2=1}\|\boldsymbol{Ay}\|_2 = \|\boldsymbol{A}\|_2. \qquad (1.23)\end{aligned}$$

证毕. $\square$

定义 1.18 设 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$, 其特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_n$, 则称

$$\rho(\boldsymbol{A}) = \max_{1\leqslant i\leqslant n}|\lambda_i|$$

为矩阵 $\boldsymbol{A}$ 的谱半径.

由上述定义, $\|\boldsymbol{A}\|_2$ 可定义为

$$\|\boldsymbol{A}\|_2 = \sqrt{\rho(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})}.$$

特别地, 当 $\boldsymbol{A}$ 为 Hermite 矩阵时, 有

$$\|\boldsymbol{A}\|_2 = \rho(\boldsymbol{A}).$$

对于一般情况, 有如下定理.

定理 1.17 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, 则

(1) $\boldsymbol{A}$ 的谱半径不超过 $\boldsymbol{A}$ 的任何范数, 即

$$\rho(\boldsymbol{A}) \leqslant \|\boldsymbol{A}\|.$$

(2) 对任意正数 ε, 存在 $\mathbb{C}^{n\times n}$ 上的某种算子范数 $\|\cdot\|_\varepsilon$, 使得

$$\|\boldsymbol{A}\|_\varepsilon \leqslant \rho(\boldsymbol{A}) + \varepsilon.$$

证明 (1) 设 λ 是 $\boldsymbol{A}$ 的一个特征值, 且 $\boldsymbol{A}$ 的谱半径 $\rho(\boldsymbol{A}) = |\lambda|$, $\boldsymbol{x}_0$ 是对应于 λ 的特征向量, 即 $\boldsymbol{A}\boldsymbol{x}_0 = \lambda\boldsymbol{x}_0$, 所以对任何一种向量

范数, 有 $\|\boldsymbol{A}\boldsymbol{x}_0\| = |\lambda| \cdot \|\boldsymbol{x}_0\|$, 故有

$$\rho(\boldsymbol{A}) = |\lambda| = \frac{\|\boldsymbol{A}\boldsymbol{x}_0\|}{\|\boldsymbol{x}_0\|} \leqslant \max_{\|\boldsymbol{x}\|\neq 0} \frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \|\boldsymbol{A}\|.$$

(2) 由 Jordan 分解定理知, 存在非奇异矩阵 $\boldsymbol{P}$ 使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \lambda_1 & \delta_1 & & & \\ & \lambda_2 & \delta_2 & & \\ & & \ddots & \ddots & \\ & & & \lambda_{n-1} & \delta_{n-1} \\ & & & & \lambda_n \end{bmatrix},$$

式中: δ_i 等于 1 或 0. 对于任意的正数 ε, 令

$$\boldsymbol{D}_\varepsilon = \begin{bmatrix} 1 & & & \\ & \varepsilon & & \\ & & \ddots & \\ & & & \varepsilon^{n-1} \end{bmatrix},$$

则

$$
\boldsymbol{D}_{\varepsilon}^{-1}\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\boldsymbol{D}_{\varepsilon}=\begin{bmatrix}\lambda_1 & \varepsilon\delta_1 & & & \\ & \lambda_2 & \varepsilon\delta_2 & & \\ & & \ddots & \ddots & \\ & & & \lambda_{n-1} & \varepsilon\delta_{n-1} \\ & & & & \lambda_n\end{bmatrix}.
$$

现定义一种新的矩阵范数

$$
\|\boldsymbol{A}\|_{\varepsilon}=\|\boldsymbol{D}_{\varepsilon}^{-1}\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\boldsymbol{D}_{\varepsilon}\|_{\infty},\quad \boldsymbol{A}\in\mathbb{C}^{n\times n},
$$

则容易验证这样定义的矩阵范数 $\|\cdot\|_{\varepsilon}$ 是由如下定义的向量范数诱导出来的算子范数:

$$
\|\boldsymbol{x}\|_{\boldsymbol{P}\boldsymbol{D}_{\varepsilon}}=\|(\boldsymbol{P}\boldsymbol{D}_{\varepsilon})^{-1}\boldsymbol{x}\|_{\infty},\quad \boldsymbol{x}\in\mathbb{C}^{n}.
$$

从而有

$$\|\boldsymbol{A}\|_\varepsilon = \|\boldsymbol{D}_\varepsilon^{-1}\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\boldsymbol{D}_\varepsilon\|_\infty = \max_{1\leqslant i\leqslant n}\big(|\lambda_i| + \varepsilon\delta_i\big) \leqslant \rho(\boldsymbol{A}) + \varepsilon,$$

这里假设 $\delta_n = 0$. 证毕. $\square$

类似于向量序列的极限定义, 有以下定义.

定义 1.19 设矩阵序列 $\{\boldsymbol{A}^{(k)} = (a_{ij}^{(k)}) \in \mathbb{C}^{m\times n}\}_{k=0}^{\infty}$, $\boldsymbol{A} = (a_{ij}) \in \mathbb{C}^{m\times n}$. 若

$$\lim_{k\to\infty} a_{ij}^{(k)} = a_{ij}, \quad i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, n,$$

则称序列 $\{\boldsymbol{A}^{(k)}\}$ 收敛于 $\boldsymbol{A}$, 记为 $\lim\limits_{k\to\infty} \boldsymbol{A}^{(k)} = \boldsymbol{A}$.

利用矩阵范数的等价性, 可得以下定理.

定理 1.18 $\lim\limits_{k\to\infty} \boldsymbol{A}^{(k)} = \boldsymbol{A}$ 当且仅当 $\lim\limits_{k\to\infty} \|\boldsymbol{A}^{(k)} - \boldsymbol{A}\| = 0$, 其中 $\|\cdot\|$ 是 $\mathbb{C}^{m\times n}$ 中任意的矩阵范数.

Back
Close

关于矩阵范数, 还有下述定理.

定理 1.19 若 $\|\boldsymbol{A}\| < 1$, 则矩阵 $\boldsymbol{I}-\boldsymbol{A}$ 非奇异, 且满足

$$\|(\boldsymbol{I}-\boldsymbol{A})^{-1}\| \leqslant \frac{1}{1-\|\boldsymbol{A}\|}.$$

证明 用反证法. 设 $\det(\boldsymbol{I}-\boldsymbol{A})=0$, 则方程 $(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$ 有非零解, 即存在 $\boldsymbol{x}_0 \in \mathbb{C}^n$, $\boldsymbol{x}_0 \neq \boldsymbol{0}$, 使得 $(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}_0=\boldsymbol{0}$. 故

$$\|\boldsymbol{A}\| = \max_{\|\boldsymbol{x}\|\neq 0} \frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \geqslant \frac{\|\boldsymbol{A}\boldsymbol{x}_0\|}{\|\boldsymbol{x}_0\|} = 1.$$

这与 $\|\boldsymbol{A}\|<1$ 矛盾.

进一步, 由于 $(\boldsymbol{I}-\boldsymbol{A})(\boldsymbol{I}-\boldsymbol{A})^{-1}=\boldsymbol{I}$, 则 $(\boldsymbol{I}-\boldsymbol{A})^{-1}=\boldsymbol{I}+\boldsymbol{A}(\boldsymbol{I}-\boldsymbol{A})^{-1}$, 从而

$$\|(\boldsymbol{I}-\boldsymbol{A})^{-1}\| \leqslant \|\boldsymbol{I}\| + \|\boldsymbol{A}\| \cdot \|(\boldsymbol{I}-\boldsymbol{A})^{-1}\|.$$

将上式整理即得要证的结论. 证毕. $\square$

定理 1.20 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, 则

(1) $\lim\limits_{k\to\infty} \boldsymbol{A}^k = \boldsymbol{O} \Longleftrightarrow \rho(\boldsymbol{A}) < 1.$

(2) $\lim\limits_{k\to\infty} \|\boldsymbol{A}^k\|^{\frac{1}{k}} = \rho(\boldsymbol{A}).$

证明 (1) 充分性. 设 $\rho(\boldsymbol{A}) < 1$, 则存在 $\varepsilon > 0$ 满足 $\rho(\boldsymbol{A}) + 2\varepsilon < 1$. 于是由定理 1.17 知存在算子范数 $\|\cdot\|$ 使得 $\|\boldsymbol{A}\| \leqslant \rho(\boldsymbol{A}) + \varepsilon$. 因此

$$\|\boldsymbol{A}^k\| \leqslant \|\boldsymbol{A}\|^k \leqslant \big[\rho(\boldsymbol{A}) + \varepsilon\big]^k < (1-\varepsilon)^k \to 0.$$

必要性. 设 $\lim\limits_{k\to\infty} \boldsymbol{A}^k = \boldsymbol{O}$, 并假定 $\lambda \in \lambda(\boldsymbol{A})$ 满足 $\rho(\boldsymbol{A}) = |\lambda|$. 由于 $\lambda \in \lambda(\boldsymbol{A})$ 蕴含 $\lambda^k \in \lambda(\boldsymbol{A}^k)$, 故由定理 1.17 知对一切的 $k = 1, 2, \cdots$ 成立

$$\big[\rho(\boldsymbol{A})\big]^k = |\lambda|^k = |\lambda^k| \leqslant \rho(\boldsymbol{A}^k) \leqslant \|\boldsymbol{A}^k\|,$$

故必有 $\rho(\boldsymbol{A}) < 1$.

(2) 由结论 (1) 的证明可知, 对任意的 $\varepsilon > 0$, 存在算子范数 $\|\cdot\|$, 满足

$$[\rho(\boldsymbol{A})]^k \leqslant \|\boldsymbol{A}^k\| \leqslant [\rho(\boldsymbol{A})+\varepsilon]^k,$$

即

$$\rho(\boldsymbol{A}) \leqslant \|\boldsymbol{A}^k\|^{\frac{1}{k}} \leqslant \rho(\boldsymbol{A})+\varepsilon,$$

由 $\varepsilon > 0$ 的任意性, 上式令 $k \to \infty$ 得

$$\lim_{k\to\infty} \|\boldsymbol{A}^k\|^{\frac{1}{k}} = \rho(\boldsymbol{A}).$$

再根据有限维空间范数的等价性, 对 $\mathbb{C}^{n\times n}$ 中的任意范数, 上述极限成立. 证毕. $\square$

利用定理 1.20 不难证明下面的重要结论.

定理 1.21 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, 则

(1) 矩阵级数 $\sum\limits_{k=0}^{\infty} \boldsymbol{A}^k$ 收敛的充分必要条件是 $\rho(\boldsymbol{A}) < 1$.

(2) 若矩阵级数 $\sum\limits_{k=0}^{\infty} \boldsymbol{A}^k$ 收敛, 则

$$\sum_{k=0}^{\infty} \boldsymbol{A}^k = (\boldsymbol{I} - \boldsymbol{A})^{-1},$$

且对于 $\mathbb{C}^{n\times n}$ 中的任意矩阵范数都有

$$\left\| (\boldsymbol{I} - \boldsymbol{A})^{-1} - \sum_{k=0}^{m} \boldsymbol{A}^k \right\| \leqslant \frac{\|\boldsymbol{A}\|^{m+1}}{1 - \|\boldsymbol{A}\|}.$$

§1.4 矩阵的广义逆

矩阵的广义逆是研究最小二乘问题及矩阵方程的一个重要工具. 首先给出矩阵广义逆的定义.

定义 1.20 设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$, 若有 $\boldsymbol{X} \in \mathbb{C}^{n\times m}$ 满足 Penrose (彭罗斯) 方程, 即

(1) $\boldsymbol{AXA} = \boldsymbol{A}$.

(2) $\boldsymbol{XAX} = \boldsymbol{X}$.

(3) $(\boldsymbol{AX})^{\mathrm{H}} = \boldsymbol{AX}$.

(4) $(\boldsymbol{XA})^{\mathrm{H}} = \boldsymbol{XA}$.

则称 $\boldsymbol{X}$ 为 $\boldsymbol{A}$ 的 Moore–Penrose 逆, 记为 $\boldsymbol{A}^{\dagger}$.

注 1.4 若 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, 则 $\boldsymbol{X} = \boldsymbol{A}^{\dagger} \in \mathbb{R}^{n\times m}$ 定义为满足下列四个方程的 $\boldsymbol{X}$:

$$(1)\ \boldsymbol{AXA}=\boldsymbol{A}. \qquad (2)\ \boldsymbol{XAX}=\boldsymbol{X}.$$

$$(3)\ (\boldsymbol{AX})^{\mathrm{T}}=\boldsymbol{AX}. \qquad (4)\ (\boldsymbol{XA})^{\mathrm{T}}=\boldsymbol{XA}.$$

由定义 1.20 容易得到:

(1) $\boldsymbol{A}\in\mathbb{C}^{n\times n}$ 可逆时, $\boldsymbol{X}=\boldsymbol{A}^{-1}$ 满足 Penrose 方程, 故 $\boldsymbol{A}^{\dagger}=\boldsymbol{A}^{-1}$.

(2) $\boldsymbol{A}=\boldsymbol{O}_{m\times n}$ 时, $\boldsymbol{X}=\boldsymbol{O}_{n\times m}$ 满足 Penrose 方程, 故 $\boldsymbol{O}_{m\times n}^{\dagger}=\boldsymbol{O}_{n\times m}$.

(3) $\boldsymbol{A}=\boldsymbol{x}\in\mathbb{C}^{n\times 1}$ 时, $\boldsymbol{X}=\dfrac{1}{\|\boldsymbol{x}\|_2^2}\boldsymbol{x}^{\mathrm{H}}$ 满足 Penrose 方程, 故 $\boldsymbol{x}^{\dagger}=\dfrac{1}{\|\boldsymbol{x}\|_2^2}\boldsymbol{x}^{\mathrm{H}}$. 例如:

$$\begin{bmatrix}1\\1+2\mathrm{i}\end{bmatrix}^{\dagger}=\frac{1}{6}(1,1-2\mathrm{i}).$$

定理 1.22 设 $\boldsymbol{F} \in \mathbb{C}_r^{m\times r}$, $\boldsymbol{G} \in \mathbb{C}_r^{r\times n}\,(r \geqslant 1)$, 则有

(1) $\boldsymbol{F}^{\dagger} = (\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F})^{-1}\boldsymbol{F}^{\mathrm{H}}$, $\quad \boldsymbol{F}^{\dagger}\boldsymbol{F} = \boldsymbol{I}_r$.

(2) $\boldsymbol{G}^{\dagger} = \boldsymbol{G}^{\mathrm{H}}(\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}})^{-1}$, $\quad \boldsymbol{G}\boldsymbol{G}^{\dagger} = \boldsymbol{I}_r$.

证明 先证结论 (1). 对于 $\boldsymbol{F}$, 记 $\boldsymbol{X} = (\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F})^{-1}\boldsymbol{F}^{\mathrm{H}}$, 则有

$$\boldsymbol{F}\boldsymbol{X}\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F})^{-1}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F} = \boldsymbol{F},$$

$$\boldsymbol{X}\boldsymbol{F}\boldsymbol{X} = (\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F})^{-1}\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F}\boldsymbol{X} = \boldsymbol{X},$$

$$(\boldsymbol{F}\boldsymbol{X})^{\mathrm{H}} = \boldsymbol{X}^{\mathrm{H}}\boldsymbol{F}^{\mathrm{H}} = \boldsymbol{F}(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F})^{-1}\boldsymbol{F}^{\mathrm{H}} = \boldsymbol{F}\boldsymbol{X},$$

$$(\boldsymbol{X}\boldsymbol{F})^{\mathrm{H}} = \boldsymbol{I}_r^{\mathrm{H}} = \boldsymbol{I}_r = \boldsymbol{X}\boldsymbol{F}.$$

所以 $\boldsymbol{F}^{\dagger} = (\boldsymbol{F}^{\mathrm{H}}\boldsymbol{F})^{-1}\boldsymbol{F}^{\mathrm{H}}$, 并且 $\boldsymbol{F}^{\dagger}\boldsymbol{F} = \boldsymbol{I}_r$.

同理, 可证结论 (2). 证毕. $\square$

定理 1.23 设 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$, 则 $\boldsymbol{A}^\dagger$ 存在且唯一.

证明 存在性. 当 $\boldsymbol{A}=\boldsymbol{O}$ 时, $\boldsymbol{A}^\dagger=\boldsymbol{O}$; 当 $\boldsymbol{A}\neq\boldsymbol{O}$ 时, $\operatorname{rank}(\boldsymbol{A}) = r \geqslant 1$, 从而 $\boldsymbol{A}$ 有满秩分解 $\boldsymbol{A}=\boldsymbol{F}\boldsymbol{G}$ ($\boldsymbol{F}\in\mathbb{C}_r^{m\times r}$, $\boldsymbol{G}\in\mathbb{C}_r^{r\times n}$). 记 $\boldsymbol{X}=\boldsymbol{G}^\dagger\boldsymbol{F}^\dagger$, 则有

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{X}\boldsymbol{A} &= \boldsymbol{F}\boldsymbol{G}\boldsymbol{G}^\dagger\boldsymbol{F}^\dagger\boldsymbol{F}\boldsymbol{G} = \boldsymbol{F}(\boldsymbol{G}\boldsymbol{G}^\dagger)(\boldsymbol{F}^\dagger\boldsymbol{F})\boldsymbol{G} = \boldsymbol{F}\boldsymbol{G} = \boldsymbol{A},\\
\boldsymbol{X}\boldsymbol{A}\boldsymbol{X} &= \boldsymbol{G}^\dagger\boldsymbol{F}^\dagger\boldsymbol{F}\boldsymbol{G}\boldsymbol{G}^\dagger\boldsymbol{F}^\dagger = \boldsymbol{G}^\dagger(\boldsymbol{F}^\dagger\boldsymbol{F})(\boldsymbol{G}\boldsymbol{G}^\dagger)\boldsymbol{F}^\dagger = \boldsymbol{G}^\dagger\boldsymbol{F}^\dagger = \boldsymbol{X},\\
(\boldsymbol{A}\boldsymbol{X})^{\mathrm{H}} &= (\boldsymbol{F}\boldsymbol{G}\boldsymbol{G}^\dagger\boldsymbol{F}^\dagger)^{\mathrm{H}} = (\boldsymbol{F}\boldsymbol{F}^\dagger)^{\mathrm{H}} = \boldsymbol{F}\boldsymbol{F}^\dagger = \boldsymbol{F}\boldsymbol{G}\boldsymbol{G}^\dagger\boldsymbol{F}^\dagger = \boldsymbol{A}\boldsymbol{X},\\
(\boldsymbol{X}\boldsymbol{A})^{\mathrm{H}} &= (\boldsymbol{G}^\dagger\boldsymbol{F}^\dagger\boldsymbol{F}\boldsymbol{G})^{\mathrm{H}} = (\boldsymbol{G}^\dagger\boldsymbol{G})^{\mathrm{H}} = \boldsymbol{G}^\dagger\boldsymbol{G} = \boldsymbol{G}^\dagger\boldsymbol{F}^\dagger\boldsymbol{F}\boldsymbol{G} = \boldsymbol{X}\boldsymbol{A},
\end{aligned}$$

由定义可得

$$\boldsymbol{A}^\dagger = \boldsymbol{G}^\dagger\boldsymbol{F}^\dagger = \boldsymbol{G}^{\mathrm{H}}(\boldsymbol{F}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{G}^{\mathrm{H}})^{-1}\boldsymbol{F}^{\mathrm{H}}. \tag{1.24}$$

唯一性. 对于 $\boldsymbol{A}$, 如果 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 都满足 Penrose 方程, 则有

$$
\begin{aligned}
\boldsymbol{X} &= \boldsymbol{XAX} = \boldsymbol{X}\cdot\boldsymbol{AYA}\cdot\boldsymbol{X} = \boldsymbol{X}\cdot(\boldsymbol{AY})(\boldsymbol{AX})\\
&= \boldsymbol{X}\cdot(\boldsymbol{AY})^{\mathrm{H}}(\boldsymbol{AX})^{\mathrm{H}} = \boldsymbol{X}\cdot(\boldsymbol{AXAY})^{\mathrm{H}} = \boldsymbol{X}\cdot(\boldsymbol{AY})^{\mathrm{H}}\\
&= \boldsymbol{XAY} = \boldsymbol{X}\cdot\boldsymbol{AYA}\cdot\boldsymbol{Y} = (\boldsymbol{XA})(\boldsymbol{YA})\cdot\boldsymbol{Y}\\
&= (\boldsymbol{XA})^{\mathrm{H}}(\boldsymbol{YA})^{\mathrm{H}}\cdot\boldsymbol{Y} = (\boldsymbol{YAXA})^{\mathrm{H}}\cdot\boldsymbol{Y}\\
&= (\boldsymbol{YA})^{\mathrm{H}}\cdot\boldsymbol{Y} = \boldsymbol{YAY} = \boldsymbol{Y},
\end{aligned}
$$

所以 $\boldsymbol{A}^{\dagger}$ 唯一. 证毕. $\square$

定理 1.23 的存在性证明实际上给出了使用矩阵满秩分解的方法来计算非零矩阵广义逆. 下面给出一个例子.

例 1.3 已知

$$\boldsymbol{A}=\begin{bmatrix}1&2&1&2\\0&1&0&1\\1&0&1&0\\2&1&2&1\end{bmatrix},$$

求 $\boldsymbol{A}$ 的满秩分解和 $\boldsymbol{A}^{\dagger}$.

解 $\quad \boldsymbol{A} \xrightarrow{\text{初等行变换}} \begin{bmatrix}1&0&1&0\\0&1&0&1\\0&0&0&0\\0&0&0&0\end{bmatrix}: \quad c_1=1, c_2=2,$

$$\boldsymbol{A}=\boldsymbol{F}\boldsymbol{G}: \quad \boldsymbol{F}=\begin{bmatrix}1&2\\0&1\\1&0\\2&1\end{bmatrix}, \quad \boldsymbol{G}=\begin{bmatrix}1&0&1&0\\0&1&0&1\end{bmatrix}.$$

因此, 有

$$\boldsymbol{F}^{\dagger}=(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F})^{-1}\boldsymbol{F}^{\mathrm{T}}=\begin{bmatrix}6&4\\4&6\end{bmatrix}^{-1}\boldsymbol{F}^{\mathrm{T}}=\frac{1}{10}\begin{bmatrix}-1&-2&3&4\\4&3&-2&-1\end{bmatrix},$$

$$\boldsymbol{G}^{\dagger}=\boldsymbol{G}^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}})^{-1}=\boldsymbol{G}^{\mathrm{T}}\begin{bmatrix}2&0\\0&2\end{bmatrix}^{-1}=\frac{1}{2}\begin{bmatrix}1&0\\0&1\\1&0\\0&1\end{bmatrix},$$

$$\boldsymbol{A}^{\dagger}=\boldsymbol{G}^{\dagger}\boldsymbol{F}^{\dagger}=\frac{1}{20}\begin{bmatrix}-1&-2&3&4\\4&3&-2&-1\\-1&-2&3&4\\4&3&-2&-1\end{bmatrix}.$$

下面的定理说明使用矩阵奇异值分解的方法来计算广义逆更为便捷.

定理 1.24 设 $\boldsymbol{A} \in \mathbb{C}_r^{m \times n}\,(r \geqslant 1)$ 的奇异值分解为

$$\boldsymbol{A} = \boldsymbol{U}\begin{bmatrix} \boldsymbol{\Sigma}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}\boldsymbol{V}^{\mathrm{H}},$$

式中: $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 及 $\boldsymbol{\Sigma}_r$ 的意义同式 (1.5), 则

$$\boldsymbol{A}^{\dagger} = \boldsymbol{V}\begin{bmatrix} \boldsymbol{\Sigma}_r^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}\boldsymbol{U}^{\mathrm{H}}. \tag{1.25}$$

证明 根据 Moore–Penrose 广义逆的定义, 直接验证即可证明.

注 1.5 需要指出, 可逆矩阵的逆 $\boldsymbol{A}^{-1}$ 具有的性质, 对于一般矩阵的广义逆 $\boldsymbol{A}^{\dagger}$ 不一定具有, 例如:

(1) $(\boldsymbol{AB})^\dagger \neq \boldsymbol{B}^\dagger \boldsymbol{A}^\dagger$：取 $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$，$\boldsymbol{B} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$，则

$$\boldsymbol{AB} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad (\boldsymbol{AB})^\dagger = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix},$$

$$\boldsymbol{A}^\dagger = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{B}^\dagger = \boldsymbol{B}^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{B}^\dagger \boldsymbol{A}^\dagger = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

(2) $A^{\dagger}A \neq AA^{\dagger}$： A 不是方阵时，这是明显的. A 是方阵但不可逆时，取

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad A^{\dagger} = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix},$$

则

$$A^{\dagger}A = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad AA^{\dagger} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

§1.5 几种特殊的矩阵类型

1. 不可约矩阵和对角占优矩阵

定义 1.21 设 $\boldsymbol{A}$ 是一个 $n\,(n \geqslant 2)$ 阶矩阵, 如果存在 n 阶置换矩阵 $\boldsymbol{P}$, 使得

$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{O} & \boldsymbol{A}_{22} \end{bmatrix},$$

式中: $\boldsymbol{A}_{11}$ 和 $\boldsymbol{A}_{22}$ 分别为 r 阶和 $n-r$ 阶方阵 $(1 \leqslant r \leqslant n-1)$. 则称 $\boldsymbol{A}$ 为可约矩阵. 如果不存在这样的置换矩阵, 则称 $\boldsymbol{A}$ 为不可约矩阵.

可约矩阵的一个等价定义如下.

定义 1.21′ 设 $\boldsymbol{A} = (a_{ij}) \in \mathbb{C}^{n\times n}\,(n \geqslant 2)$. 又设 $\mathcal{I} = \{1, 2, \cdots, n\}$. 若存在 $\mathcal{I}$ 的一个非空真子集 $\mathcal{K}$, 使得 $a_{ij} = 0\,(i \notin \mathcal{K}, j \in \mathcal{K})$, 则称 $\boldsymbol{A}$ 为可约矩阵. 否则, 称 $\boldsymbol{A}$ 为不可约矩阵.

Back
Close

用定义 $1.21'$ 判定一个矩阵是否可约更为方便.

例 1.4 设 $\boldsymbol{A} = (a_{ij})$ 为三对角矩阵, 即满足 $a_{ij} = 0\,(|i-j| > 1)$. 证明: 若

$$a_{i+1,i} \neq 0, \quad a_{i,i+1} \neq 0, \quad i = 1, 2, \cdots, n-1,$$

则 $\boldsymbol{A}$ 是不可约矩阵.

证明 用反证法. 若 $\boldsymbol{A}$ 可约, 则由定义 $1.21'$ 知, 存在 $\mathcal{I}$ 的非空真子集 $\mathcal{K}$, 使得 $a_{ij} = 0\,(i \notin \mathcal{K}, j \in \mathcal{K})$. 由于 $\mathcal{K}$ 非空, 则 $\mathcal{K}$ 至少有一个数, 设为 r. 由 $a_{r,r+1} \neq 0$ 知, $r+1 \in \mathcal{K}$. 类似可得 $r+2 \in \mathcal{K}, \cdots, n \in \mathcal{K}$. 另外, 又由 $a_{r-1,r} \neq 0$ 知, $r-1 \in \mathcal{K}$. 类似可得 $r-2 \in \mathcal{K}, \cdots, 1 \in \mathcal{K}$. 即有 $\mathcal{K} = \mathcal{I}$, 这与 $\mathcal{K}$ 是 $\mathcal{I}$ 的真子集矛盾. 故 $\boldsymbol{A}$ 为不可约矩阵.

Back
Close

定义 1.22 设 $\boldsymbol{A}$ 是一个 n 阶矩阵.

(1) 若 $\boldsymbol{A}$ 满足

$$|a_{ii}| \geqslant \sum_{j \neq i} |a_{ij}|, \ i = 1, 2, \cdots, n,$$

且其中至少有个严格不等式成立, 则称 $\boldsymbol{A}$ 为 (行) 弱对角占优.

(2) 若 $\boldsymbol{A}$ 满足

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|, \ i = 1, 2, \cdots, n,$$

则称 $\boldsymbol{A}$ 为 (行) 严格对角占优.

类似可定义 (列) 弱对角占优和 (列) 严格对角占优.

以后往往不严格区分按行 (严格) 对角占优或按列 (严格) 对角占优而模糊地统称为 (严格) 对角占优. 弱对角占优且不可约的矩阵简称为不可约对角占优.

Back
Close

引理 1.1 若 n 阶矩阵 $\boldsymbol{A}$ 满足下列条件之一:

(1) 严格对角占优.

(2) 不可约对角占优.

则 $\boldsymbol{A}$ 非奇异, 即其行列式 $\det(\boldsymbol{A}) \neq 0$.

证明 首先考虑 $\boldsymbol{A}$ 为严格对角占优的情形. 用反证法. 假设 $\boldsymbol{A}$ 是奇异的, 则存在非零向量 $\boldsymbol{x}$ 使得 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$. 不妨设 $|x_i| = \|\boldsymbol{x}\|_\infty = 1$. 则有

$$|a_{ii}| = |a_{ii}x_i| = \left|\sum_{j\neq i} a_{ij}x_j\right| \leqslant \sum_{j\neq i} |a_{ij}|,$$

这与 $\boldsymbol{A}$ 严格对角占优矛盾.

再考虑 $\boldsymbol{A}$ 为不可约对角占优时结论成立. 仍用反证法. 设非零向量 $\boldsymbol{x}$ 满足 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, 并定义

$$\mathcal{K} = \{k : |x_k| < 1\}.$$

Back
Close

易知 $\mathcal{K}$ 非空. 否则 $\boldsymbol{x}$ 的所有分量的模均为 1, 则对于所有的 $i \in \mathcal{K}$, 均有

$$|a_{ii}| \leqslant \sum_{j \neq i} |a_{ij}|,$$

这与 $\boldsymbol{A}$ 对角占优矛盾. 此外, 因 $\boldsymbol{A}$ 不可约, 必存在 i, k 使得

$$a_{ik} \neq 0,\ i \notin \mathcal{K},\ k \in \mathcal{K}.$$

于是 $|a_{ik}x_k| < |a_{ik}|$, 且

$$\begin{aligned} |a_{ii}| &\leqslant \sum_{j \notin \mathcal{K}, j \neq i} |a_{ij}| \cdot |x_j| + \sum_{j \in \mathcal{K}} |a_{ij}| \cdot |x_j| \\ &< \sum_{j \notin \mathcal{K}, j \neq i} |a_{ij}| + \sum_{j \in \mathcal{K}} |a_{ij}| = \sum_{j \neq i} |a_{ij}|, \end{aligned}$$

这又与 $\boldsymbol{A}$ 对角占优矛盾. 证毕. $\square$

2. 非负矩阵和 M 矩阵

非负矩阵在迭代法中起着非常重要的作用. 一个非负矩阵就是其元素全为非负数的矩阵. 对两个 $m \times n$ 阶矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$, 如果其所有相应的元素满足 $a_{ij} \geqslant b_{ij}$, 则称 $\boldsymbol{A} \geqslant \boldsymbol{B}$, 这样 "$\geqslant$" 就定义了矩阵集合上的一个偏序关系. 利用这种偏序关系, 非负矩阵 $\boldsymbol{A}$ 可以表示为 $\boldsymbol{A} \geqslant \boldsymbol{O}$. $\boldsymbol{A} \geqslant \boldsymbol{B}$ 也可表示为 $\boldsymbol{B} \leqslant \boldsymbol{A}$, 即 "$\leqslant$" 也定义了矩阵集合上的一个偏序关系.

性质 1.9 关于非负矩阵, 下列性质成立:

(1) 若 $\boldsymbol{A} \geqslant \boldsymbol{O}$, $\boldsymbol{B} \geqslant \boldsymbol{O}$, 则 $\boldsymbol{AB} \geqslant \boldsymbol{O}$, $\boldsymbol{A}+\boldsymbol{B} \geqslant \boldsymbol{O}$ 且 $\boldsymbol{A}^k \geqslant \boldsymbol{O}$.

(2) 若 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ 为非负矩阵且 $\boldsymbol{A} \leqslant \boldsymbol{B}$, 则 $\boldsymbol{AC} \leqslant \boldsymbol{BC}$, $\boldsymbol{CA} \leqslant \boldsymbol{CB}$.

Back
Close

(3) 若 $\boldsymbol{O} \leqslant \boldsymbol{A} \leqslant \boldsymbol{B}$, 则 $\boldsymbol{A}^{\mathrm{T}} \leqslant \boldsymbol{B}^{\mathrm{T}}$ 且 $\boldsymbol{A}^k \leqslant \boldsymbol{B}^k$ 对任意的 k 均成立.

(4) 若 $\boldsymbol{O} \leqslant \boldsymbol{A} \leqslant \boldsymbol{B}$, 则 $\|\boldsymbol{A}\|_1 \leqslant \|\boldsymbol{B}\|_1$, 且类似地 $\|\boldsymbol{A}\|_\infty \leqslant \|\boldsymbol{B}\|_\infty$.

定理 1.25 (Perron–Frobenius 定理) 设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, 则有下列结论成立:

(1) 若 $\boldsymbol{A} \geqslant \boldsymbol{O}$, 则 $\boldsymbol{A}$ 有一个非负特征值等于 $\rho(\boldsymbol{A})$; 对 $\rho(\boldsymbol{A}) > 0$, 相应的特征向量 $\boldsymbol{x} \geqslant \boldsymbol{0}$, 且当 $\boldsymbol{A}$ 的任意元素增加时, $\rho(\boldsymbol{A})$ 不减少.

(2) 若 $\boldsymbol{A} \geqslant \boldsymbol{O}$ 不可约, 则 $\boldsymbol{A}$ 有一个正特征值等于 $\rho(\boldsymbol{A})$; 对 $\rho(\boldsymbol{A}) > 0$, 其对应的特征向量 $\boldsymbol{x} > \boldsymbol{0}$; 当 $\boldsymbol{A}$ 的任意元素增加时, $\rho(\boldsymbol{A})$ 增加; 且 $\rho(\boldsymbol{A})$ 是 $\boldsymbol{A}$ 的一个简单特征值.

性质 1.10 设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, 则有下列性质成立:

(1) 若 $\boldsymbol{O} \leqslant \boldsymbol{A} \leqslant \boldsymbol{B}$, 则 $\rho(\boldsymbol{A}) \leqslant \rho(\boldsymbol{B})$.

(2) 若 $|\boldsymbol{A}| \leqslant \boldsymbol{B}$ ($|\boldsymbol{A}|$ 表示 $\boldsymbol{A}$ 的元素取绝对值后得到的矩阵), 则 $\rho(\boldsymbol{A}) \leqslant \rho(\boldsymbol{B})$.

(3) 设 $\boldsymbol{A} \geqslant \boldsymbol{O}$, 则 $\rho(\boldsymbol{A}) < 1$ 当且仅当 $\boldsymbol{I}-\boldsymbol{A}$ 非奇异且 $(\boldsymbol{I}-\boldsymbol{A})^{-1} \geqslant \boldsymbol{O}$.

(4) 设 $\boldsymbol{A} \geqslant \boldsymbol{O}$, $\boldsymbol{x} \geqslant \boldsymbol{0}$ 为非零向量, $\alpha > 0$. 若 $\boldsymbol{A}\boldsymbol{x} > (\geqslant)\alpha\boldsymbol{x}$, 则 $\rho(\boldsymbol{A}) > (\geqslant)\alpha$; 若 $\boldsymbol{A}\boldsymbol{x} < \alpha\boldsymbol{x}$, 则 $\rho(\boldsymbol{A}) < \alpha$.

定义 1.23 (1) 设 $\boldsymbol{A}$ 是一个 n 阶矩阵. 若 $\boldsymbol{A}$ 的主对角元是正的, 即 $a_{ii} > 0$, $i = 1, 2, \cdots, n$, 而其非主对角元是非正的, 即 $a_{ij} \leqslant 0$, $i \neq j$, $i, j = 1, 2, \cdots, n$, 则称 $\boldsymbol{A}$ 为 Z 矩阵. 若 Z 矩阵 $\boldsymbol{A}$ 为非奇异的, 且 $\boldsymbol{A}^{-1} \geqslant \boldsymbol{O}$, 则称 $\boldsymbol{A}$ 为 M 矩阵.

(2) 记矩阵 $\langle \boldsymbol{A} \rangle = (\alpha_{ij})$, 其中主对角元 $\alpha_{ii} = |a_{ii}|$, 而非主对

角元 $\alpha_{ij} = -|a_{ij}|$, 则称 $\langle \boldsymbol{A} \rangle$ 为 $\boldsymbol{A}$ 的比较矩阵. 若 $\langle \boldsymbol{A} \rangle$ 为非奇异的 M 矩阵, 则称 $\boldsymbol{A}$ 为 H 矩阵.

关于 M 矩阵和 H 矩阵, 有下面一些常用的性质.

性质 1.11 设 $\boldsymbol{A}$ 是 Z 矩阵, 则 $\boldsymbol{A}$ 是 M 矩阵当且仅当 $\rho(\boldsymbol{B}) < 1$, 其中 $\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{D}^{-1}\boldsymbol{A}$ 且 $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{A})$.

性质 1.12 设 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, 则有下列性质成立:

(1) 严格对角占优或不可约对角占优矩阵是 H 矩阵.

(2) 严格对角占优或不可约对角占优矩阵的 Z 矩阵是 M 矩阵.

(3) 若 $\boldsymbol{A}$ 为 H 矩阵, 则 $|\boldsymbol{A}|^{-1} \leqslant \langle \boldsymbol{A} \rangle^{-1}$.

(4) 若 $\boldsymbol{A} = \boldsymbol{D} - \boldsymbol{B}$ 为 H 矩阵, $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{A})$, 则 $|\boldsymbol{D}|$ 非奇异, 且 $\rho(|\boldsymbol{D}|^{-1}|\boldsymbol{B}|) < 1$.

性质 1.13 设 $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$ 是两个非奇异的 M 矩阵, 且 $\boldsymbol{A} \leqslant \boldsymbol{B}$, 则

(1) $\boldsymbol{A}^{-1} \geqslant \boldsymbol{B}^{-1}$.

(2) $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的每个主子矩阵也是 M 矩阵.

(3) 满足 $\boldsymbol{A} \leqslant \boldsymbol{C} \leqslant \boldsymbol{B}$ 的 $\boldsymbol{C}$ 都是 M 矩阵. 特别地, 若 $\boldsymbol{A} \leqslant \boldsymbol{C} \leqslant \mathrm{diag}(\boldsymbol{A})$, 则 $\boldsymbol{C}$ 是 M 矩阵.

性质 1.14 若 $\boldsymbol{A}$ 是 M 矩阵, $\boldsymbol{B}$ 是 Z 矩阵, 且 $A \leqslant B$, 则 $\boldsymbol{B}$ 也是 M 矩阵.

性质 1.15 (1) 若分块矩阵 $\boldsymbol{A} = \left(\boldsymbol{A}_{ij}\right)_{i,j=1}^{m}$ 是 M 矩阵, 且其主对角块 $\boldsymbol{A}_{ii}$ 为方阵, 则 $\boldsymbol{A}_{ii}$ 也是 M 矩阵.

(2) 若 $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}$ 是 M 矩阵, 则 Schur 补 $\boldsymbol{S} = \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$ 也是 M 矩阵.

3. 正定矩阵

关于正定矩阵, 在实矩阵的情形, 前面已作了介绍. 下面考虑复矩阵的情形.

定义 1.24 (1) 设矩阵 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, 若对任意非零向量 $\boldsymbol{x} \in \mathbb{C}^n$ 都有 $\mathrm{Re}(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}) > 0$, 则称 $\boldsymbol{A}$ 为正定的. 若 $\boldsymbol{A}$ 还是 Hermite 的, 则称为 Hermite 正定的, 记为 HPD.

(2) 设矩阵 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, 若对任意非零向量 $\boldsymbol{x} \in \mathbb{R}^n$ 都有 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} > 0$, 则称 $\boldsymbol{A}$ 为正实的.

任意 (实或复) 方阵 $\boldsymbol{A}$ 均可分裂为

$$\boldsymbol{A} = \boldsymbol{H} + \mathrm{i}\boldsymbol{S}, \tag{1.26}$$

式中:

$$\boldsymbol{H} = \frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^{\mathrm{H}}), \quad \boldsymbol{S} = \frac{1}{2\mathrm{i}}(\boldsymbol{A} - \boldsymbol{A}^{\mathrm{H}}).$$

注意到 $\boldsymbol{H}$ 和 $\boldsymbol{S}$ 都是 Hermite 的, 而 $\mathrm{i}\boldsymbol{S}$ 是反 Hermite 的. 矩阵 $\boldsymbol{H}$ 和 $\mathrm{i}\boldsymbol{S}$ 分别称为 $\boldsymbol{A}$ 的 Hermite 部分和反 Hermite 部分. 当 $\boldsymbol{A}$ 是实矩阵且 $\boldsymbol{u}$ 是实向量时, 有 $(\boldsymbol{A}\boldsymbol{u},\boldsymbol{u})=(\boldsymbol{H}\boldsymbol{u},\boldsymbol{u})$. 分裂 (1.26) 称为矩阵 $\boldsymbol{A}$ 的一个 Hermite–反 Hermite 分裂, 简称 HS 分裂.

正定矩阵具有下面一些常用的性质.

性质 1.16 关于矩阵 $\boldsymbol{A}$, 下列性质成立:

(1) 设 $\boldsymbol{A}$ 为实正定矩阵, 则 $\boldsymbol{A}$ 是非奇异的, 且对任意的实向量 $\boldsymbol{x}$, 存在标量 $\alpha>0$ 满足 $(\boldsymbol{A}\boldsymbol{x},\boldsymbol{x})\geqslant\alpha\|\boldsymbol{x}\|^2$.

(2) 正定矩阵 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$ 的所有特征特均有正的实部.

(3) 设 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$ 为正定矩阵, $\boldsymbol{P}\in\mathbb{C}^{n\times n}$ 为 n 阶非奇异矩阵, 则 $\boldsymbol{P}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{P}$ 亦为正定矩阵.

(4) 正定矩阵 $\boldsymbol{A}\in\mathbb{C}^{n\times n}$ 的所有主子阵均为正定矩阵. 特别地, $\boldsymbol{A}$ 的所有对角元素均有正的实部.

(5) 设 $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, 则 $\boldsymbol{A}$ 是正定矩阵当且仅当 $\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}}$ 为 Hermite 正定的.

定理 1.26 对于矩阵 $\boldsymbol{A}$ 的 HS 分裂 (1.26), 其特征值 λ_i 满足

$$\lambda_{\min}(\boldsymbol{H}) \leqslant \mathrm{Re}(\lambda_i) \leqslant \lambda_{\max}(\boldsymbol{H}), \quad \lambda_{\min}(\boldsymbol{S}) \leqslant \mathrm{Im}(\lambda_i) \leqslant \lambda_{\max}(\boldsymbol{S}).$$

当 $\boldsymbol{B}$ 为对称正定矩阵时, $\mathbb{C}^n \times \mathbb{C}^n$ 到 $\mathbb{C}$ 上的映射 $\boldsymbol{x},\boldsymbol{y} \to (\boldsymbol{x},\boldsymbol{y})_{\boldsymbol{B}} \equiv (\boldsymbol{Bx},\boldsymbol{y})$ 是 $\mathbb{C}^n$ 上的一个内积, 称此内积为能量内积. 而称其诱导范数为能量范数. 有时, 可以找到一个 HPD 矩阵 $\boldsymbol{B}$, 使得一个给定的在欧几里得内积下的非 Hermite 矩阵 $\boldsymbol{A}$ 在能量内积下成为 Hermite 的, 即:

$$(\boldsymbol{Ax},\boldsymbol{y})_{\boldsymbol{B}} = (\boldsymbol{x},\boldsymbol{Ay})_{\boldsymbol{B}}, \ \ \forall\, \boldsymbol{x},\boldsymbol{y}.$$

最简单的例子是 $\boldsymbol{A}=\boldsymbol{B}^{-1}\boldsymbol{C}$ 和 $\boldsymbol{A}=\boldsymbol{CB}$, 其中 $\boldsymbol{C}$ 是 Hermite 的且 $\boldsymbol{B}$ 是 Hermite 正定的.

§1.6 模型问题: Poisson 问题

考虑二维规则区域上的 Poisson 问题

$$\begin{cases} -\left(\dfrac{\partial^2 u}{\partial x^2}+\dfrac{\partial^2 u}{\partial y^2}\right)+c\dfrac{\partial u}{\partial x}+d\dfrac{\partial u}{\partial y}+eu=f(x,y), \quad (x,y)\in\Omega, \\ u(x,y)|_{\partial\Omega}=0, \end{cases} \tag{1.27}$$

式中: $\Omega=(0,1)\times(0,1)$; f 为给定的函数; c,d,e 为非负常数.

设 x,y 方向均取 $n+1$ 个等距网格, 则内网格点共有 n^2 个, 步长为 $h=\dfrac{1}{n+1}$, 如图 1.1 所示.

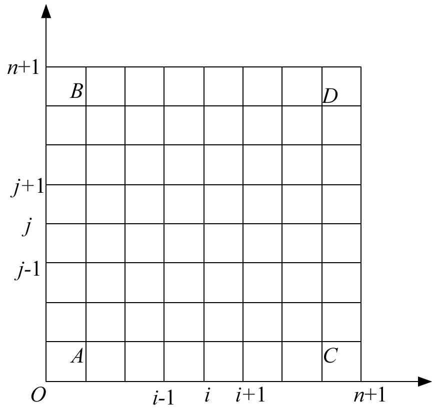

图 1.1 区域离散和节点标号

将

$$u_{xx}(x_i,y_j) \approx \frac{u_{i+1,j}-2u_{ij}+u_{i-1,j}}{h^2}, \quad u_{yy}(x_i,y_j) \approx \frac{u_{i,j+1}-2u_{ij}+u_{i,j-1}}{h^2},$$

$$u_x(x_i,y_j) \approx \frac{u_{i+1,j}-u_{i-1,j}}{2h}, \quad u_y(x_i,y_j) \approx \frac{u_{i,j+1}-u_{i,j-1}}{2h},$$

$$u(x_i,y_j) \approx u_{ij}, \quad f(x_i,y_j) \approx f_{ij}.$$

代入式 (1.27) 的第 1 式并整理, 得

$$(4+h^2e)u_{ij} - (u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}) + \frac{hc}{2}(u_{i+1,j} - u_{i-1,j})$$

$$+\frac{hd}{2}(u_{i,j+1} - u_{i,j-1}) = h^2 f_{ij}, \quad i,j = 1,2,\cdots,n.$$

该方程组包含了 n^2 个未知量, 它的矩阵形式为

$$\boldsymbol{Au} = \boldsymbol{f}, \tag{1.28}$$

式中: $\boldsymbol{A} = \boldsymbol{A}_1 + \boldsymbol{A}_2 + \boldsymbol{A}_3$ 为 n^2 阶的块三对角矩阵, 其具体形式为

$$
\boldsymbol{A}_1 = \begin{bmatrix} \boldsymbol{B} & -\boldsymbol{I} & & & \\ -\boldsymbol{I} & \boldsymbol{B} & -\boldsymbol{I} & & \\ & \ddots & \ddots & \ddots & \\ & & -\boldsymbol{I} & \boldsymbol{B} & -\boldsymbol{I} \\ & & & -\boldsymbol{I} & \boldsymbol{B} \end{bmatrix}, \tag{1.29}
$$

$$
\boldsymbol{A}_2 = \frac{hc}{2} \begin{bmatrix} \boldsymbol{O} & \boldsymbol{I} & & & \\ -\boldsymbol{I} & \boldsymbol{O} & \boldsymbol{I} & & \\ & \ddots & \ddots & \ddots & \\ & & -\boldsymbol{I} & \boldsymbol{O} & \boldsymbol{I} \\ & & & -\boldsymbol{I} & \boldsymbol{O} \end{bmatrix}, \tag{1.30}
$$

$$\boldsymbol{A}_3 = \frac{hd}{2}\begin{bmatrix} \boldsymbol{O} & -\boldsymbol{I} & & & \\ \boldsymbol{I} & \boldsymbol{O} & -\boldsymbol{I} & & \\ & \ddots & \ddots & \ddots & \\ & & \boldsymbol{I} & \boldsymbol{O} & -\boldsymbol{I} \\ & & & \boldsymbol{I} & \boldsymbol{O} \end{bmatrix}, \tag{1.31}$$

这里

$$\boldsymbol{B} = \begin{bmatrix} 4+h^2e & -1 & & & \\ -1 & 4+h^2e & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4+h^2e & -1 \\ & & & -1 & 4+h^2e \end{bmatrix} \tag{1.32}$$

Back
Close

为对称三对角矩阵, $\boldsymbol{I}$ 为 n 阶单位矩阵. n^2 维列向量 $\boldsymbol{u}$ 定义为

$$\boldsymbol{u} = (u_{11}, u_{12}, \cdots, u_{1n}, u_{21}, u_{22}, \cdots, u_{2n}, \cdots, u_{n1}, u_{n2}, \cdots, u_{nn})^{\mathrm{T}},$$

网格中的每个内部节点都对应一个分量, u_{ij} 表示 $u(ih, jh)$ 的近似, 这些分量按沿 y 方向变化最快的方式排列. 图 1.1 中节点 A, B, C, D 的标号分别为 $(1,1), (1,n), (n,1), (n,n)$, 节点 (i,j) 对应的标号为 $i*n+j$, 它的上下左右的四个相邻节点为

$$i*n+(j+1), \quad i*n+(j-1), \quad (i-1)*n+j, \quad (i+1)*n+j.$$

右端向量 $\boldsymbol{f}$ 为

$$\boldsymbol{f} = h^2(f_{11}, f_{12}, \cdots, f_{1n}, f_{21}, f_{22}, \cdots, f_{2n}, \cdots, f_{n1}, f_{n2}, \cdots, f_{nn})^{\mathrm{T}}.$$

以后常常用此模型做数值实验, 故称它为模型问题.