现代数学基础

35 Toeplitz 系统预处理方法

■ 金小庆 著
■ 庞宏奎 译

高等教育出版社
HIGHER EDUCATION PRESS

To Kathy, Cecilia and Jeffrey

前言

Toeplitz 系统出现在许多应用问题中, 比如, 数值求解微分方程、数值求解卷积型积分方程、平稳自回归时间序列、系统识别问题以及图像恢复问题等 [24, 28, 40, 47, 56, 61, 73], 不胜枚举. 在本书中, 我们对最近发展的求解 Toeplitz 系统的预处理技术做一个基本的介绍.

用循环矩阵作为预处理共轭梯度法 (简称 PCG 方法) 的预处理矩阵来求解 Toeplitz 系统始于 1986 年 [74, 81]. 此迭代法的一个主要结果是对于一大类 $n \times n$ 的 Toeplitz 系统 $T_n\mathbf{x} = \mathbf{b}$, 其求解复杂度只要 $\mathcal{O}(n \log n)$. 此后, 许多不同的预处理矩阵被人们构造、使用和分析. 由于受篇幅和时间的限制, 我们主要从理论的角度研究一些著名的预处理矩阵, 并给出其在求解常微分方程 (简称 ODE) 系统中的应用.

本书分为七章. 在第一章中, 我们对本书将要用到的一些数值线性代数的背景知识做一简要的叙述. 然后介绍两个重要的 Krylov 子空间方法: 共轭梯度法 (简称 CG 方法) 和广义极小残量法 (简称 GMRES 方法). 最后给出一些用循环矩阵作为 CG 方法的预处理矩阵来求解 Toeplitz 系统的基本知识.

在第二章中, 我们引入 Strang 循环预处理矩阵 $s(T_n)$, 它是由 Strang 于 1986 年提出来的首个循环预处理矩阵 [81]. 此预处理矩阵是对给定的 Toeplitz 矩阵 T_n, 通过保留其中间的对角线部分, 然后把它反射到两边而构成的循环矩阵. 本章对用 $s(T_n)$ 作为 PCG 方法的预处理矩阵的收敛性进行了分析. Strang 循环预处理矩阵在许多文献中已被用来有效地求解一些良态的 Hermite Toeplitz 系统.

在第三章中, 我们研究 T. Chan 最优预处理矩阵 [35]. 对任意的 Toeplitz 矩阵 T_n, T. Chan 循环预处理矩阵 $c_F(T_n)$, 也称为最优循环预处理矩阵, 定义为使得 $\|T_n - W_n\|_F$ 达到最小值的矩阵, 其中 $\|\cdot\|_F$ 是 Frobenius 范数, W_n 取遍所有的循环矩阵. 我们对用 $c_F(T_n)$ 作为 PCG 方法的预处理矩阵的收敛性进行了分析, 并从算子的角度研究了 T. Chan 预处理矩阵的性质以及它的稳定性.

第四章讨论超最优预处理矩阵 [87]. 对任意的 Toeplitz 矩阵 T_n, 超最优循环预处理矩阵 $t_F(T_n)$ 定义为使得 $\|I_n - W_n^{-1}T_n\|_F$ 达到最小值的矩阵, 这里 I_n 指 $n \times n$ 的单位矩阵, W_n 取遍所有的非奇异循环矩阵. 我们对用 $t_F(T_n)$ 作为 PCG 方法的预处理矩阵的收敛性进行了分析, 并讨论了超最优预处理后的矩阵和最优预处理后的矩阵的谱之间的关系. 最后, 我们给出了一些数值结果.

在第五章中, 我们考察了用带状 Toeplitz 预处理矩阵 [32] 和 $\{\omega\}$–循环预处理矩阵 [75] 来求解病态的 Toeplitz 系统.

第六章着重讨论用 PCG 方法来求解块系统 $T_{mn}\mathbf{x} = \mathbf{b}$, 其中 T_{mn} 是一个 $m \times m$ 的块 Toeplitz 矩阵, 并且每一块都是 $n \times n$ 的 Toeplitz 矩阵. 本章定义了一个保持 T_{mn} 的块结构的块预处理矩阵, 并给出了算法的计算量估计和收敛性分析. 数值结果表明, 对于求解一些良态的块 Toeplitz 系统, 此块预处理矩阵是一个不错的预处理矩阵.

在第七章中, 我们介绍用边值方法 (简称 BVM) 求解具有初值的 ODE 系统 [9, 29]. 此类方法需要求解一个或多个大型稀疏的非对称线性系统, GMRES 方法结合循环型预处理矩阵可用来求解这些线性系

统. 我们证明当用 A_{ν_1,ν_2} - 稳定的 BVM 求解一个 $m \times m$ 的 ODE 系统时, 预处理后的矩阵可以写成 $I + L$, 这里 I 是单位矩阵, L 是秩至多为 $2m(\nu_1 + \nu_2)$ 的矩阵. 因此, 当用 GMRES 方法求解这些预处理后的线性系统时, 最多需要 $2m(\nu_1 + \nu_2) + 1$ 步迭代即可收敛. 产生这一章数值结果的 MATLAB 程序见附录.

本书包含了近些年得到的关于 Toeplitz 迭代解法的一些重要的研究成果. 与 [24, 56, 73] 这些书不同, 它们面向的是数学专业的研究生或相关领域的专家, 而本书可为科学计算相关专业的高年级本科生所接受, 它要求读者只要具有基本的线性代数、微积分、数值分析和科学计算的知识即可. 另外, 本书也可作为对 Toeplitz 快速迭代解法感兴趣的科研和工程计算人员的参考书.

本书的成形, 得到了许多人的影响和帮助. 在此, 我对他们的引导、启发、讨论和鼓励表示衷心的感谢. 特别地, 我要感谢我的博士导师、香港中文大学数学系陈汉夫教授, 是他带我走进了 Toeplitz 快速迭代解法这个非常有趣的领域, 并感谢他一直以来对我的引导、鼓励及与我保持长期的友谊. 感谢加州大学洛杉矶分校数学系陈繁昌教授 (现为香港科技大学校长), 他富有启发性的想法、建议和意见让我受益匪浅. 感谢那些花费时间阅读本书的初稿并提出详细意见的同仁们, 他们是: 东南大学数学系陈建龙教授、浙江大学数学系韩丹夫教授、上海海事大学数学系徐兆亮教授以及我的博士研究生张应应. 同时, 感谢汕头大学数学系林福荣教授, 他为本书第七章提供了 MATLAB 程序. 当然, 特别要感谢的是在我的生活中有着重要地位的单位 —— 澳门大学, 感谢它为我提供了这么良好的一个学术氛围来写此书. 本书的写作主要是在晚上、周末和假期完成的. 最后, 我要感谢我的家人, 他们对我必要的鼓励、无尽的耐心和无私的爱是我能够完成此书的根本.

本书的写作得到了澳门大学基金 RG-UL/07-08S/Y1/JXQ/FST 和 UL020/08-Y2/MAT/JXQ01/FST 的资助.

—— 金小庆

译者的话: 本译本纠正了原书中的一些数学和打印错误, 并增加了一个参考文献 [82]. 本书在翻译过程中得到了上海海事大学数学系徐兆亮教授、华南师范大学数学科学学院黎稳教授和澳门大学数学系金小庆教授的大力支持和帮助, 他们非常仔细地阅读了本书的初稿, 并提出了许多宝贵的意见. 译者在这里向他们表示衷心的感谢!

本书的翻译得到了国家青年科学基金项目 11201192、江苏省科学基金项目 SBK201220691、江苏省高校科学基金项目 12KJB110004 和江苏高校优势学科建设工程资助项目的资助.

—— 庞宏奎

2013 年 2 月于江苏师范大学

目录

第一章　简介

“如果还有哪一个数学分支像微积分和微分方程一样基础的话, 那将是数值线性代数.”—— [84]

在本章中, 我们首先对本书将要用到的有关数值线性代数的背景知识做一简要叙述. 然后, 给出两个重要且被广泛使用的迭代方法: 共轭梯度法 (简称 CG 方法) 和广义极小残量法 (简称 GMRES 方法). 最后, 介绍一些有关 Toeplitz 迭代解法的基础知识.

§1.1　数值线性代数的背景知识

本节回顾一下数值线性代数中的一些重要概念, 这里介绍的内容将在后面的章节中用到.

§1.1.1　基本的符号、记号和定义

在本书中, 我们使用如下的符号、记号和定义.

— 令 $\mathbb{R}$ 表示实数的全体; $\mathbb{C}$ 为复数的全体; $\mathbf{i} \equiv \sqrt{-1}$; 符号 $\mathbb{R}^n$ 表示 n 维实向量的全体; $\mathbb{C}^n$ 表示 n 维复向量的全体. 注意, 本书提到的向量, 若无特殊说明, 均指列向量. 我们用 $\mathbb{R}^{m\times n}$ 表示所有 $m\times n$ 的实矩阵构成的线性向量空间; 用 $\mathbb{C}^{m\times n}$ 表示所有 $m\times n$ 的复矩阵构成的线性向量空间. 符号 $\dim(S)$ 表示线性向量空间 S 的维数.

— 大写字母, 如 A, B, C, Δ, Λ 等, 表示矩阵; 黑体小写字母, 如 $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ 等, 表示向量.

— 记号 $(A)_{ij} = a_{ij}$ 表示矩阵 $A = (a_{ij})$ 的 (i,j) 位置的元素. 对任意的 $n\times n$ 矩阵, 指标 i 和 j 一般取从 1 到 n, 但有时候为了方便起见也取从 0 到 $n-1$. 我们用 A^T 和 A^* 分别表示矩阵 A 的转置矩阵和共轭转置矩阵; 用 $\mathrm{rank}(A)$ 表示 A 的秩; 用 $\mathrm{tr}(A) \equiv \sum_{i=1}^{n} a_{ii}$ 表示 A 的迹; 用

$$\mathrm{diag}(a_{11},\ldots,a_{nn}) = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix}$$

表示一个 $n\times n$ 的对角矩阵; 用 I_n 表示 $n\times n$ 的单位矩阵, 在不至于引起混淆的情况下, 我们简记它为 I; 用符号 $\mathbf{e}_j$ 表示单位矩阵 I 的第 j 列.

— 矩阵 $A\in\mathbb{R}^{n\times n}$ 称为对称矩阵, 若 $A^T = A$. 一个对称矩阵 A 是正定的 (半正定的), 如果对任意的非零向量 $\mathbf{x}\in\mathbb{R}^n$, 都有 $\mathbf{x}^T A\mathbf{x} > 0$ $(\geqslant 0)$. 矩阵 $A\in\mathbb{C}^{n\times n}$ 称为 Hermite 矩阵, 若 $A^* = A$. 一个 Hermite 矩阵 A 是正定的 (半正定的), 如果对任意的非零向量 $\mathbf{x}\in\mathbb{C}^n$, 都有 $\mathbf{x}^* A\mathbf{x} > 0$ $(\geqslant 0)$.

— 设 $A\in\mathbb{C}^{n\times n}$. 如果存在非零向量 $\mathbf{x}\in\mathbb{C}^n$ 和标量 $\lambda\in\mathbb{C}$ 使得 $A\mathbf{x} = \lambda\mathbf{x}$, 则称 λ 是 A 的特征值, $\mathbf{x}$ 是 A 的相应于 λ 的特征

向量. 众所周知, Hermite 矩阵的特征值都是实数, 记 $\lambda_{\min}(A)$ 和 $\lambda_{\max}(A)$ 分别为 Hermite 矩阵 A 的最小特征值和最大特征值. 我们用 $\rho(A) \equiv \max|\lambda_i(A)|$ 表示 A 的谱半径, 这里的 $\lambda_i(A)$ 取遍 A 的谱. 所谓 A 的谱是指 A 的所有特征值构成的集合.

— 我们用 A_{mn} 表示任意一个 $m \times m$ 的分块矩阵并且每一块为 $n \times n$ 的矩阵; 用 $(A_{mn})_{i,j;k,l}$ 表示 A_{mn} 的 (k,l) 块的 (i,j) 位置的元素.

— 我们用 $\|\cdot\|$ 表示一个向量或矩阵的范数; 用符号 $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$ 表示 p – 范数, 这里 p 的取值分别为 1, 2, ∞.

— 符号 $\mathbf{C}_{2\pi}$ 表示以 2π 为周期的连续实值函数 $f(x)$ 的全体; $\mathbf{C}_{2\pi}^{+}$ 表示在 $\mathbf{C}_{2\pi}$ 中所有满足条件 $f(x) \geqslant 0$ 的函数构成的子集, 这里要求 $f(x)$ 不恒为零; 符号 $\mathbf{C}_{2\pi\times 2\pi}$ 表示以 2π 为周期的 (在每个方向上) 连续实值函数 $f(x,y)$ 的全体.

§1.1.2 Hermite 矩阵谱的性质

Hermite 矩阵的谱有很多优美而又重要的性质. 这里我们给出一些经典的结果 [50, 61, 84, 93], 它们将在后面的讨论中用到.

定理 1.1 (谱定理) 矩阵 $A \in \mathbb{C}^{n\times n}$ 是 Hermite 矩阵, 当且仅当存在一个酉矩阵 $Q \in \mathbb{C}^{n\times n}$ 和一个对角矩阵 $\Lambda \in \mathbb{R}^{n\times n}$, 使得 $A = Q\Lambda Q^*$.

我们称一个矩阵 $M \in \mathbb{C}^{n\times n}$ 为酉矩阵, 如果 $M^{-1} = M^*$.

定理 1.2 (Cauchy 交错定理) 设 $A \in \mathbb{C}^{n\times n}$ 是 Hermite 矩阵, 其特征值按递增顺序排列

$$\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n.$$

如果

$$\mu_1 \leqslant \mu_2 \leqslant \cdots \leqslant \mu_m$$

是 A 的一个阶数为 m 的主子矩阵的特征值, 则有

$$\lambda_k \leqslant \mu_k \leqslant \lambda_{k+n-m},$$

其中 $k=1,2,\ldots,m$.

定理 1.3 (Weyl 定理) 设 $A, E\in\mathbb{C}^{n\times n}$ 是 Hermite 矩阵, 又设诸特征值 $\lambda_i(A)$, $\lambda_i(E)$ 以及 $\lambda_i(A+E)$ 均按递增顺序排列. 则对每个 $k=1,2,\ldots,n$, 有

$$\lambda_k(A)+\lambda_1(E)\leqslant\lambda_k(A+E)\leqslant\lambda_k(A)+\lambda_n(E).$$

定理 1.4 (Courant-Fischer 极大极小定理) 设 $A\in\mathbb{C}^{n\times n}$ 是 Hermite 矩阵, 其特征值按递增顺序排列 $\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$. 则对每个 $k=1,2,\ldots,n$, 有

$$\begin{aligned}\lambda_k &= \min_{\substack{S\subseteq\mathbb{C}^n\\ \dim(S)=k}}\ \max_{\mathbf{0}\neq\mathbf{x}\in S}\frac{\mathbf{x}^*A\mathbf{x}}{\mathbf{x}^*\mathbf{x}} = \min_{\substack{S\subseteq\mathbb{C}^n\\ \dim(S)=k}}\ \max_{\substack{\mathbf{x}\in S\\ \|\mathbf{x}\|_2=1}}\mathbf{x}^*A\mathbf{x}\\ &= \max_{\substack{S\subseteq\mathbb{C}^n\\ \dim(S)=n-k+1}}\ \min_{\mathbf{0}\neq\mathbf{x}\in S}\frac{\mathbf{x}^*A\mathbf{x}}{\mathbf{x}^*\mathbf{x}} = \max_{\substack{S\subseteq\mathbb{C}^n\\ \dim(S)=n-k+1}}\ \min_{\substack{\mathbf{x}\in S\\ \|\mathbf{x}\|_2=1}}\mathbf{x}^*A\mathbf{x},\end{aligned}$$

其中 S 表示 $\mathbb{C}^n$ 的任一子空间, $\|\mathbf{x}\|_2=\left(\sum\limits_{i=1}^{n}|x_i|^2\right)^{1/2}$, 这里的

$$\mathbf{x}=(x_1,x_2,\ldots,x_n)^T\in\mathbb{C}^n.$$

特别地,

$$\begin{aligned}\lambda_1&=\min_{\mathbf{x}\neq\mathbf{0}}\frac{\mathbf{x}^*A\mathbf{x}}{\mathbf{x}^*\mathbf{x}}=\min_{\|\mathbf{x}\|_2=1}\mathbf{x}^*A\mathbf{x},\\ \lambda_n&=\max_{\mathbf{x}\neq\mathbf{0}}\frac{\mathbf{x}^*A\mathbf{x}}{\mathbf{x}^*\mathbf{x}}=\max_{\|\mathbf{x}\|_2=1}\mathbf{x}^*A\mathbf{x}.\end{aligned}$$

现在, 我们介绍矩阵的奇异值. 众所周知 [50], 对任意的矩阵 $A\in\mathbb{C}^{m\times n}$, 不妨假设 $m\geqslant n$, 总存在矩阵 $U\in\mathbb{C}^{m\times n}$ 满足 $U^*U=I$ 和矩阵 $V\in\mathbb{C}^{n\times n}$ 满足 $V^*V=I$, 使得

$$A=U\Delta V^*,$$

其中 $\Delta=\operatorname{diag}(\sigma_1,\sigma_2,\ldots,\sigma_n)$, $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_n\geqslant 0$. 标量 σ_k $(k=1,2,\ldots,n)$ 称为 A 的奇异值. 注意到, 矩阵 A^*A 的特征值为 σ_k^2, $k=1,2,\ldots,n$.

§1.1.3 范数和条件数

任意一个标量的大小都可以很容易地用它的模来度量, 这个模是一个非负的量, 例如 $c \in \mathbb{C}$ 可以用它的模 $|c|$ 来度量其大小. 类似地, 对于 $\mathbb{C}^n$ 中的向量或 $\mathbb{C}^{n\times n}$ 中的矩阵, 我们也可以用一个称作范数的非负测度来度量其大小. 通常, 范数用来度量矩阵计算中的误差. 下面, 我们首先讨论向量的范数, 然后再介绍矩阵的范数.

定义 1.5 设 $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T \in \mathbb{C}^n$, 一个定义在 $\mathbb{C}^n$ 上的向量范数是一个函数, 它对每个 $\mathbf{x} \in \mathbb{C}^n$ 都定义了一个实数 $\|\mathbf{x}\|$, 称为 $\mathbf{x}$ 的范数, 使得对所有的 $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ 和 $\alpha \in \mathbb{C}$ 都满足下列三条性质:

(a) $\|\mathbf{x}\| \geqslant 0 \quad (\|\mathbf{x}\| = 0$ 当且仅当 $\mathbf{x} = \mathbf{0})$;

(b) $\|\alpha\mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\|$;

(c) $\|\mathbf{x}+\mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|$.

一类有用的向量范数是 p – 范数, 定义为

$$\|\mathbf{x}\|_p \equiv \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p},$$

其中 $p \geqslant 1$. 在实际中最常用的 p – 范数是:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|, \quad \|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}, \quad \|\mathbf{x}\|_\infty = \max_{1\leqslant i\leqslant n} |x_i|.$$

$\mathbb{C}^n$ 上的向量范数有一个非常重要的性质, 即所有定义在 $\mathbb{C}^n$ 上的向量范数都是等价的. 也就是说, 如果 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是 $\mathbb{C}^n$ 上的任意两个向量范数, 则存在两个正的常数 c_1 和 c_2, 使得对所有的 $\mathbf{x} \in \mathbb{C}^n$ 都有 [50]

$$c_1\|\mathbf{x}\|_\alpha \leqslant \|\mathbf{x}\|_\beta \leqslant c_2\|\mathbf{x}\|_\alpha.$$

容易证明

$$\|\mathbf{x}\|_2 \leqslant \|\mathbf{x}\|_1 \leqslant \sqrt{n}\|\mathbf{x}\|_2, \qquad \|\mathbf{x}\|_\infty \leqslant \|\mathbf{x}\|_2 \leqslant \sqrt{n}\|\mathbf{x}\|_\infty.$$

例如,

$$\|\mathbf{x}\|_\infty = \left(\max_{1\leqslant i\leqslant n}|x_i|^2\right)^{1/2} \leqslant \|\mathbf{x}\|_2 \equiv \left(\sum_{i=1}^n |x_i|^2\right)^{1/2} \leqslant \left(n\max_{1\leqslant i\leqslant n}|x_i|^2\right)^{1/2} = \sqrt{n}\|\mathbf{x}\|_\infty.$$

接下来我们介绍矩阵范数. 令矩阵 $A=(a_{ij})\in\mathbb{C}^{n\times n}$ 的列分块形式为

$$A = (\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n),$$

行分块形式为

$$A = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_n \end{pmatrix}.$$

定义 1.6 设 $A=(a_{ij})\in\mathbb{C}^{n\times n}$, 矩阵范数是一个函数, 它对每个 $A\in\mathbb{C}^{n\times n}$ 都定义了一个实数 $\|A\|$, 称为 A 的范数, 使得对所有的 $A,B\in\mathbb{C}^{n\times n}$ 和 $\alpha\in\mathbb{C}$ 都满足下列四条性质:

(a) $\|A\|\geqslant 0 \quad (\|A\|=0$ 当且仅当 $A=\mathbf{0})$;

(b) $\|\alpha A\| = |\alpha|\cdot\|A\|$;

(c) $\|A+B\|\leqslant\|A\|+\|B\|$;

(d) $\|AB\|\leqslant\|A\|\cdot\|B\|$.

对于每个向量范数, 我们都可以用一个自然的方法 "诱导" 出一个矩阵范数. 给定一个向量范数 $\|\cdot\|_v$, 由其诱导出来的矩阵范数 $\|\cdot\|_M$ 定义为

$$\|A\|_M \equiv \max_{\mathbf{x}\neq\mathbf{0}}\frac{\|A\mathbf{x}\|_v}{\|\mathbf{x}\|_v} = \max_{\|\mathbf{x}\|_v=1}\|A\mathbf{x}\|_v.$$

最常用的矩阵范数是由向量 p – 范数诱导出来的矩阵 p – 范数, 其中 $p=1,2,\infty$. 易证 [44, 61]:

$$\|A\|_1 \equiv \max_{\|\mathbf{x}\|_1=1}\|A\mathbf{x}\|_1 = \max_{1\leqslant j\leqslant n}\sum_{i=1}^n|a_{ij}| = \max_{1\leqslant j\leqslant n}\|\mathbf{c}_j\|_1,$$

$$\|A\|_\infty \equiv \max_{\|\mathbf{x}\|_\infty=1} \|A\mathbf{x}\|_\infty = \max_{1\leqslant i\leqslant n} \sum_{j=1}^{n} |a_{ij}| = \max_{1\leqslant i\leqslant n} \|\mathbf{r}_i\|_1$$

和

$$\|A\|_2 \equiv \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \sqrt{\lambda_{\max}(A^*A)} = \sigma_{\max}(A), \tag{1.1}$$

其中 $\lambda_{\max}(A^*A)$ 表示 A^*A 的最大特征值, $\sigma_{\max}(A)$ 表示 A 的最大奇异值.

作为例子, 我们下面给出 (1.1) 的证明. 注意到

$$\begin{aligned}\|A\|_2 &= \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \max_{\|\mathbf{x}\|_2=1} [(A\mathbf{x})^*(A\mathbf{x})]^{1/2} \\ &= \max_{\|\mathbf{x}\|_2=1} [\mathbf{x}^*(A^*A)\mathbf{x}]^{1/2}.\end{aligned}$$

由于 A^*A 是 Hermite 半正定的, 可假设它的特征值按如下顺序排列:

$$0 \leqslant \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n.$$

根据定理 1.1, 令 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{C}^n$ 是分别对应于特征值 $\lambda_1, \ldots, \lambda_n$ 的标准正交特征向量. 则对于任意的向量 $\mathbf{x} \in \mathbb{C}^n$, $\|\mathbf{x}\|_2 = 1$, 我们有

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i, \qquad \sum_{i=1}^{n} |\alpha_i|^2 = 1.$$

因此

$$\mathbf{x}^*A^*A\mathbf{x} = \sum_{i=1}^{n} \lambda_i |\alpha_i|^2 \leqslant \lambda_n.$$

另一方面, 取 $\mathbf{x} = \mathbf{v}_n$, 我们可得

$$\mathbf{x}^*A^*A\mathbf{x} = \mathbf{v}_n^*A^*A\mathbf{v}_n = \mathbf{v}_n^*\lambda_n\mathbf{v}_n = \lambda_n.$$

所以

$$\|A\|_2 = \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \sqrt{\lambda_n} = \sqrt{\lambda_{\max}(A^*A)} = \sigma_{\max}(A).$$

另外一个有用的矩阵范数称为 Frobenius 范数, 其定义为

$$\begin{aligned}\|A\|_F &\equiv \Big(\sum_{j=1}^{n}\sum_{i=1}^{n} |a_{ij}|^2\Big)^{1/2} = \Big(\sum_{j=1}^{n} \|\mathbf{c}_j\|_2^2\Big)^{1/2} = \Big(\sum_{i=1}^{n} \|\mathbf{r}_i\|_2^2\Big)^{1/2} \\ &= [\mathrm{tr}(A^*A)]^{1/2} = \Big[\sum_{k=1}^{n} \lambda_k(A^*A)\Big]^{1/2},\end{aligned}$$

其中 $\mathrm{tr}(A^*A)$ 表示矩阵 A^*A 的迹. 范数 $\|\cdot\|_2$ 和 $\|\cdot\|_F$ 有一个非常重要的性质, 即酉不变性.

定理 1.7 设 $A \in \mathbb{C}^{n\times n}$, 对任意的酉矩阵 Q 和 Z, 我们有

$$\|A\|_2 = \|QAZ\|_2, \qquad \|A\|_F = \|QAZ\|_F.$$

证明 设 $\mathbf{y} = Q\mathbf{x}$, $\mathbf{x} \in \mathbb{C}^n$. 对于向量范数 $\|\cdot\|_2$,

$$\|\mathbf{y}\|_2^2 = \mathbf{y}^*\mathbf{y} = (Q\mathbf{x})^*(Q\mathbf{x}) = \mathbf{x}^*Q^*Q\mathbf{x} = \mathbf{x}^*\mathbf{x} = \|\mathbf{x}\|_2^2.$$

而对于矩阵范数 $\|\cdot\|_2$, 根据 (1.1) 可得

$$\begin{aligned}\|QAZ\|_2^2 &= \lambda_{\max}[(QAZ)^*(QAZ)] = \lambda_{\max}(Z^*A^*AZ)\\ &= \lambda_{\max}(A^*A) = \|A\|_2^2.\end{aligned}$$

对于范数 $\|\cdot\|_F$, 令

$$AZ = W = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n).$$

由向量范数 $\|\cdot\|_2$ 的酉不变性, 可推出

$$\begin{aligned}\|QAZ\|_F^2 = \|QW\|_F^2 &= \sum_{j=1}^{n}\|Q\mathbf{w}_j\|_2^2 = \sum_{j=1}^{n}\|\mathbf{w}_j\|_2^2\\ &= \|W\|_F^2 = \|AZ\|_F^2.\end{aligned}$$

接下来我们只需要证明 $\|AZ\|_F^2 = \|A\|_F^2$. 根据 A 的行分块形式, 再次利用向量范数 $\|\cdot\|_2$ 的酉不变性, 即得

$$\begin{aligned}\|AZ\|_F^2 &= \sum_{i=1}^{n}\|\mathbf{r}_iZ\|_2^2 = \sum_{i=1}^{n}\|Z^*\mathbf{r}_i^*\|_2^2\\ &= \sum_{i=1}^{n}\|\mathbf{r}_i^*\|_2^2 = \|A\|_F^2.\end{aligned}$$

□

此外, 我们指出谱半径是所有矩阵范数的一个下界.

定理 1.8 设 $A \in \mathbb{C}^{n\times n}$, 则对任意的矩阵范数 $\|\cdot\|$, 有 $\rho(A) \leqslant \|A\|$.

证明 设 $\mathbf{x} \in \mathbb{C}^n$ 满足

$$\mathbf{x} \neq \mathbf{0}, \qquad A\mathbf{x} = \lambda\mathbf{x}, \qquad |\lambda| = \rho(A).$$