计算机理论基础与应用丛书

矩阵计算与应用

胡茂林　编著

科 学 出 版 社

北　京

内 容 简 介

矩阵计算不仅是一门数学分支学科，也是众多理工科的重要的数学工具，计算机科学和工程的问题最终都变成关于矩阵的运算.

本书主要针对计算机科学、电子工程和计算数学等学科中的研究需求，以各种类型的线性方程组求解为主线进行阐述.内容侧重于分析各种矩阵分解及其应用，而不是矩阵的理论分析.介绍了各类算法在计算机上的实现方法，并讨论了各种算法的敏感性分析.在广度上和深度上较同类教材都有所加强.

本书适合相关领域广大研究生与高年级本科生阅读，也可作为这些领域中学者的参考书.

图书在版编目(CIP)数据

矩阵计算与应用/胡茂林编著.—北京：科学出版社，2008
(计算机理论基础与应用丛书)
ISBN 978-7-03-021226-9

Ⅰ.矩… Ⅱ.胡… Ⅲ.①矩阵-计算方法 ②矩阵-应用
Ⅳ.O241.6

中国版本图书馆 CIP 数据核字(2008)第 027556 号

责任编辑：姚庆爽 / 责任校对：陈玉凤
责任印制：刘士平 / 封面设计：王 浩

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

中国科学院印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2008 年 5 月第 一 版 开本：B5(720×1000)
2008 年 5 月第一次印刷 印张：23 3/4
印数：1—4 000 字数：464 000

定价：48.00 元

(如有印装质量问题，我社负责调换〈环伟〉)

前　　言

随着计算机硬件的发展和处理复杂算法能力的提高，近30年来，以人工智能为核心的相关学科群：计算机视觉、模式识别(含机器学习)、数字图像处理、数字信号处理和计算机图形学得到了迅速的发展.20世纪90年代，这些学科的发展逐步走向成熟，相关技术的融合和实际应用显著增长.而且，随着计算机应用深入到社会科学和生物学等学科，加之计算机网络的迅速扩展，数据的维数激增和数据量按指数增长，计算机所处理的数据发生了根本性的变化，这些都将进一步推动相关学科向纵深发展.

在这些学科研究的过程中，涉及数学知识的广度和深度都超出了人们的想象.在广度上，几乎所有数学科目都在这些学科的研究中出现过，而不像传统的学科，如物理主要应用微分几何、偏微分方程和群论；不仅如此，这些学科研究过程中所用的数学理论往往是当前数学界最新的研究成果，比如图像处理中所用的偏微分方程理论.这对没有受过严格数学训练的计算机学者提出了严峻挑战.

传统的计算机学科研究所用到的数学主要集中在离散数学、算法设计、数值计算和组合数学，这些19世纪的数学已经无法满足当前计算机科学发展的要求.为此，众多的计算机学者一方面呼吁数学工作者加入到计算机科学的研究中，同时也积极地将相关的数学理论引入到研究中.近年来，在国外著名大学的计算机系(或学院)都开设了介绍近代数学基础的课程，在这些学科近年出版的专著中，加入了相当篇幅的数学知识，另外在各类学术网站上，像著名的计算机视觉网站 http://homepages.inf.ed.ac.uk/rbf/CVonline/，开设有专门介绍相关数学知识的网页.但这些介绍仅是零散的，缺乏系统性.另外，在这些学科专著中介绍数学知识往往是菜单式的，很少给出证明，因此相关研究人员如果没有很强的数学背景，是难以理解其中的内容的，这为他们理解和设计算法带来了困难，使他们难以将理论研究的成果尽快地转化到实际应用中.

由于这些学科涉及众多数学知识，且分散在数学的不同学科中，所以要求计算机和工程专业的研究人员去学习和掌握这些知识是困难的.另外数学的理论研究与应用学科的研究对象不同，即使有学者愿意到数学学科去参考这些资料，也难以找到需要的内容.对于作者——一个有较强的数学背景(从本科到博士所学专业都是基础数学)的学者来说，这些问题也常常让他感到困惑.特别是要为计算机视觉和模式识别方向的研究生补充相关数学知识时，却苦于没有这方面的教材.鉴于此，作者最近几年以来，主要以计算机视觉和模式识别理论研究中所涉及的数学为

主线，将相关的数学知识收集起来，初步完成了《空间和变换》(已出版)、《矩阵计算与应用》、《优化算法》、《数据分析》等四本作为计算机科学中的数学基础读本.另外三本书《图论和组合优化》、《几何分析与计算》和《偏微分方程在计算机科学中的应用》(题目暂定)也将争取尽快完成.这部丛书着重介绍计算，而不是理论分析.这样有助于计算机等相关学科的研究人员和研究生能在最短的时间内掌握现代数学的一些知识和方法，为计算机科学的理论研究打下坚实的数学基础.

为了使丛书适合计算机及其相关学科的研究者阅读，所选内容基本是作者阅读计算机科学论文和专著中遇到过的.写作方式优先参考计算机学科的专著和论文，然后是工程类图书和论文，最后加以补充数学方面的材料使其完整.虽然丛书作为计算机科学中的数学基础读本，但并不表示书中的数学内容是简单的，只不过是把它们收集起来，方便相关人员参考而已.

最后，材料的选取是以作者个人对计算机学科的认识为主，加之涉及内容比较广泛，成书的时间较短，难免有疏漏和不适之处，希望读者给予批评和指正，以便再版时改进.

胡茂林
2007 年 12 月
于中国科学院上海微系统与信息技术研究所

《矩阵计算与应用》内容介绍

矩阵计算又称为数值线性代数.作为一门数学学科,它是众多理工学科重要的数学工具.矩阵理论既是经典数学的基础课程,是数学的一个重要且目前仍然非常活跃的领域,又是一门最有实用价值的数学理论,是计算机科学与工程计算的核心,已成为现代各科技领域处理大量有限维空间形式与数量关系强有力的工具,计算机科学和工程的问题最终都转化成矩阵的运算与求解.特别是计算机的广泛应用为矩阵论的应用开辟了广阔的前景.例如,系统工程、优化方法以及稳定性理论等,都与矩阵论有着密切的联系.目前,国内外已经出版了许多深受读者喜爱的有关矩阵理论的著作,但是大部分都以介绍理论为主,很少给出应用和算法实现上的说明,面向对象主要是数学专业.本卷只介绍了计算机科学应用中所涉及的矩阵理论,详细地讨论了具体的计算方法.与一般著作不同的是,本书更强调了矩阵的应用、算法、分解及其实际中非常重要的敏感性分析.

本书包含 17 章,分为五个部分.第一部分包含第 1~3 章,这是矩阵计算的基础,也是高等代数有关内容的复习和深入.第二部分包含第 4~7 章,介绍矩阵的各种分解,这是矩阵计算的核心.如果从线性方程组的求解角度看,这可以看做是系数矩阵从方阵求解开始,到最小二乘问题,到约束二次优化,到最后对任何类型的线性方程组(含过定和欠定)给出统一的描述.第三部分包含第 8~12 章,是关于矩阵的特征值和特征向量求解问题.第四部分包含第 12~14 章,是关于系数矩阵是大规模稀疏矩阵的迭代求解算法.第五部分包含第 15~17 章,是三个独立的内容,分别为矩阵函数、矩阵的点积和直积及非负矩阵.后面三个部分,相对比较独立,读者在阅读第一、二部分后,可以根据需要,选读后面的任意一部分.以下是这些章的内容介绍.

第 1 章介绍了本书所用到矩阵的记号和定义,详细研究了方阵的特征问题,尤其是求方阵的特征值所涉及的有关内容.基于这些讨论,给出了矩阵的实、复 Schur 分解.然后介绍了许多特殊的矩阵,简单描述了它们的性质.最后讨论了分块矩阵的计算及其应用,给出了矩阵逆和线性方程组的更新算法.

第 2 章详细地讨论了对称矩阵和正定矩阵的性质,相对一般矩阵,对称矩阵具有非常好的性质,实际应用时通常可以看做是对角矩阵.给出了广义特征值问题的定义,在此基础上给出了对称和正定矩阵同时对角化的方法.然后介绍了对称矩阵的变分原理,这是研究对称矩阵的特征值问题的好的载体,基于这个工具,讨论了对称矩阵扰动的特征值问题.再介绍了约束特征问题和广义特征问题的变分原理,

利用正交补空间将约束问题化成无约束问题,给出了约束特征问题与一般特征问题之间的关系.最后给出了广义特征问题的变分原理.

第 3 章介绍了向量和矩阵范数的各种定义,同时给出了与范数有关的一些定义,比如误差和收敛率.对矩阵范数,着重讨论了诱导范数,给出了计算机科学中常用的三种范数表达式.最后介绍了向量和矩阵范数在线性方程组求解和方阵逆中的敏感性分析的应用.

第 4 章首先给出了矩阵的三角分解和满秩分解,三角分解是求解阶数不太大的线性方程组的主要手段.虽然理论上这两个分解是基于初等变换,但在计算机科学中的应用却非常广泛,且具有计算代价低的优点.根据 Gauss 消去法给出矩阵可以三角分解的充分必要条件,介绍选主元法,并给出了实际的执行方法.然后介绍对称正定矩阵的 Cholesky 分解,这虽然可以看做三角分解的特例,但具有独特的性质,因此给出了 Cholesky 分解的更新算法.最后介绍矩阵的满秩分解.

第 5 章给出了矩阵的 QR 分解,即正交三角分解,它是矩阵计算中大多数算法的基础:包括最小二乘问题、特征值和奇异值问题的求解.这些迭代算法中的关键步骤就是 QR 分解.首先给出了两个正交变换——Givens 变换和 Householder 变换,讨论了它们的性质.然后给出了各种形式的 QR 分解算法,尤其是修正的 Gram-Schmidt,这是目前求矩阵的 QR 分解的通用算法,研究矩形矩阵的 QR 分解;最后介绍了矩阵的 QR 分解的更新算法和在最小二乘法中的应用.

第 6 章介绍了矩阵的奇异值分解,奇异值分解就是将矩阵分解成最简化形式的矩阵——对角矩阵与正交矩阵的乘积,因此在欧氏范数下,矩阵可以看做是对角矩阵.首先给出矩阵的奇异值分解的标准方法,然后解释分解的几何意义,讨论了其性质.在矩阵的奇异值分解下,给出一些矩阵范数的表达式.其次给出了如何利用奇异值分解来研究子空间之间的关系,基于矩阵奇异值分解来简化一些二次规划的问题.最后讨论了奇异值的极性和扰动理论.

第 7 章介绍了广义逆和伪逆,它们可以给出线性方程组求解的统一理论.首先从单边逆开始引入广义逆的概念,给出了基于广义逆的线性方程组的通解表达式.然后给出了伪逆的定义,讨论了伪逆的性质和不同的计算方法.基于伪逆给出了欠定和过定线性方程组具有特殊性质解的表达式,将伪逆的定义推广到有线性约束的情形.最后给出了伪逆的扰动理论,分析了最小二乘问题的敏感性.

第 8 章首先介绍了求特征向量的幂法,逆迭代和移位逆迭代,详细分析了 Rayleigh 商迭代,这些算法虽然比较简单,但它们是后来介绍的高级算法的重要成分和思想.然后简单介绍求矩阵特征值和特征向量最常用的 QR 算法,目前这个算法还在发展中.最后给出了基于移位的 QR 算法加速收敛方法,讨论了不同情况下的移位策略.

第 9 章介绍了 QR 算法具体实现问题.首先介绍了常用 QR 算法的执行方

法——隐 QR 算法，即不显示地进行移位；为了在实域中来解实矩阵的特征值，对实矩阵讨论了双步 QR 算法. 其次介绍了运用 QR 算法计算特征向量的方法、矩阵奇异值分解的有效计算方法，最后介绍了线性子空间的迭代，这是在高层次上讨论矩阵特征值和特征向量的计算方法. 并且介绍了子空间的迭代，说明了同时迭代和 QR 算法之间的关系.

第 10 章是关于特征值的估计和敏感性分析. 首先对矩阵的特征值所在复平面上的位置进行估计，介绍了著名的 Gershgorin 定理和它在对角占优矩阵中的应用. 然后给出了特征值整体和单个的敏感性分析. 最后讨论了特征向量的敏感性.

第 11 章介绍了对称矩阵的特征计算方法. 针对矩阵的对称性，给出了有些独特且有效的算法. 首先给出求对称矩阵特征值经典的 Jacobi 算法，介绍了这个算法的各种推广. 然后介绍三对角矩阵的特征求解算法，因为一般对称矩阵都可以很容易化简到三对角矩阵，这些是目前求对称矩阵常用的方法；对三对角矩阵给出了特征值不同的求解方法. 最后给出用逆迭代计算对称矩阵的特征向量的各种方法.

第 12 章介绍了线性方程组的迭代求解方法. 首先给出了线性方程组求解的一些经典迭代方法——Jacobi 方法、Gauss-Seidel 以及超松弛迭代. 其次对这些迭代方法进行了收敛性分析，给出了基于迭代矩阵的范数的判断收敛的充分条件. 最后给出了迭代收敛的例子，主要讨论了对角占优矩阵超松弛迭代的最佳收敛参数，也简单地给出了关于正定矩阵的结论.

第 13 章对系数矩阵是对称正定矩阵的线性方程组介绍了最速下降方法，这是优化问题基于下降方法的重要求解方法和推广其他算法的基础；同时介绍了预优化的重要思想，通过应用预优方法，下降方法的收敛率将显著提高. 然后描述了计算上非常有效的共轭梯度法. 共轭梯度方法只是 Krylov 子空间方法这一大类中的一个. 本章以对非正定和非对称问题进行简单的 Krylov 子空间方法的讨论作为结束.

第 14 章给出了大规模稀疏矩阵的线性方程组求解和特征问题的计算. 首先讨论了如何有效地对大规模稀疏矩阵进行 LU 分解，介绍缩减矩阵带宽的 CM 算法和将矩阵分解为块矩阵的结构分解矩阵. 然后介绍经典的 Arnoldi 算法及其矩阵表达式；对对称矩阵给出了 Lanczos 计算方法. 最后介绍特殊的隐重新开始的 Arnoldi 算法.

第 15 章是关于矩阵函数的. 首先定义了矩阵序列收敛性，说明收敛矩阵序列具有一般收敛数列所具有的性质. 给出了幂级数的定义. 其次利用函数的 Taylor 级数展开，引入了矩阵函数. 最后给出了矩阵的微积分及其应用.

第 16 章介绍了矩阵之间的特殊运算——Hadamard 积和直积（或称 Kronecker 积）. 它们不仅在实际中经常出现，而且在矩阵的理论研究和计算方法中都有十分重要的应用. 特别地，运用矩阵的直积，能够将线性矩阵方程转化为线性代数

方程组进行讨论或计算.

第 17 章介绍了非负矩阵的某些性质,包括著名的关于正矩阵的 Perron 定理和不可约非负矩阵的 Frobenius 定理;介绍了素矩阵,它对应于正矩阵的结论,是非负矩阵的中心理论.在实际应用中,矩阵的逆是非负矩阵也非常重要,称为单调矩阵,对此也进行了简单介绍.

目　　录

第 1章 矩阵的基本知识

本章和第 2、3 章主要讨论有关矩阵的一些基础知识,它们是高等代数中有关矩阵知识的总结和提高.在 1.1 节,首先给出本书有关矩阵的记号和定义,对矩阵知识进行了简单总结,讨论了矩阵计算的复杂度问题,然后详细地研究了方阵特征值的问题,给出了基于已知特征向量和特征值的矩阵简化方法.在 1.2 节研究了一些特殊矩阵及其性质,最后讨论了分块矩阵的计算及其应用.本章是本书以后各章的思想和知识基础.

1.1 基本概念

在高等代数中,我们已经学习过矩阵的一些基本概念和性质,为了以后行文方便和避免不必要的重复,首先对本书所用的一些记号和术语做简要说明.

基本定义和符号 本书的记号采用国际标准记号.大写字母 R 表示实数,R^n 表示实 n 维向量空间,所有 $m\times n$实矩阵的集合用 $R^{m\times n}$ 表示;与实元素对应的复元素集合分别用 C,C^n 和 $C^{m\times n}$ 表示 .记号 $\mathrm{diag}(a_1,\cdots,a_n)$ 表示对角元素为 $a_1,\cdots,a_n$ 的 n阶对角阵,当 a_i 被方阵 $\boldsymbol{A}_i$ 代替后,表示块对角阵,见 1.2 节.矩阵 $\boldsymbol{A}$ 的**秩**表示 $\boldsymbol{A}$ 的列(或行)向量中最大线性无关个数,记为 $\mathrm{rank}\boldsymbol{A}$. $R_r^{m\times n}$ 表示 $R^{m\times n}$ 中所有秩为 r 的矩阵集合.

对矩阵 $\boldsymbol{A}\in R^{m\times n}$,$\boldsymbol{A}^{\mathrm{T}}$ 表示矩阵 $\boldsymbol{A}$ 的转置.对复矩阵 $\boldsymbol{A}\in C^{m\times n}$,$\overline{\boldsymbol{A}}$ 和 $\boldsymbol{A}^{\mathrm{H}}$ 分别表示它的共轭和共轭转置矩阵.本书主要讨论实矩阵,对于复矩阵的情形,只要将转置换成共轭转置,通常就可以进行类似推导.

当 $m=n$时,矩阵 $\boldsymbol{A}\in R^{n\times n}$称为 $\boldsymbol{n}$ **阶方阵**,方阵比长方阵具有更多的代数性质. $\boldsymbol{I}_n$ 表示 n 阶单位方阵,它的第 k列记为 $\boldsymbol{e}_k^{(n)}$,即

$$\boldsymbol{I}_n=\begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix},\quad \boldsymbol{e}_k^{(n)}=(0,\cdots,0,1,0,\cdots,0)^{\mathrm{T}}\in R^n$$

当维数在上下文中是清楚的,就略去指标 n,简记为 $\boldsymbol{I}$ 和 $\boldsymbol{e}_k$.

n阶方阵 $\boldsymbol{A}$ 的**行列式**和**迹**分别记为 $\det\boldsymbol{A}$ 和 $\mathrm{tr}\boldsymbol{A}$.如果 $\det\boldsymbol{A}\neq 0$,则 $\boldsymbol{A}$ 是**可逆**的,它的逆矩阵用 $\boldsymbol{A}^{-1}$ 表示,即 $\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{-1}=\boldsymbol{I}$.可逆矩阵也称为**非奇异**的,不可

逆矩阵也称为**奇异**的.

若 $\boldsymbol{A}\in C^{n\times n}$满足 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{I}$,则称 $\boldsymbol{A}$ 是**酉矩阵**,n 阶酉矩阵的全体记作 $U(n)$;如果 $\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$,则称 $\boldsymbol{A}$ 是 **Hermite 矩阵**;如果 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$,则称 $\boldsymbol{A}$ 是**正规矩阵**.显然,酉矩阵和 Hermite 矩阵都是正规矩阵.实酉矩阵称作**实正交矩阵**,n 阶实正交矩阵的集合记为 $O(n)$;实 Hermite 矩阵称作**实对称矩阵**.如果 $\boldsymbol{A}=-\boldsymbol{A}^{\mathrm{H}}$,则称 $\boldsymbol{A}$ 为**反 Hermite 矩阵**,对应的实矩阵就是**反对称矩阵**.对于一个 Hermite 矩阵 $\boldsymbol{A}$,如果对任意的非零向量 $\boldsymbol{x}\in C^n$,都有 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}>0$,则称 $\boldsymbol{A}$ 是**正定**的;如果对任意向量 $\boldsymbol{x}\in C^n$,都有 $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}\geqslant 0$,则称 $\boldsymbol{A}$ 是**半正定**的.

如果向量 $\boldsymbol{a}_k\in R^m$ 为 $\boldsymbol{A}\in R^{m\times n}$的第 k 列,则 $\boldsymbol{A}$ 可记为

$$\boldsymbol{A}=[\boldsymbol{a}_1,\cdots,\boldsymbol{a}_n]$$

同样,如果 $\boldsymbol{r}_k\in R^n$ 表示 $\boldsymbol{A}$ 的第 k 行的 n 维列向量,则 $\boldsymbol{A}$ 可记为

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{r}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{r}_m^{\mathrm{T}}\end{bmatrix}$$

下面的内容在文献(胡茂林,2007)中介绍过,现在简单回顾一下.设 $L\subset R^n$ 是一个子空间.$L^{\perp}$ 记为 L 的**正交补空间**,即 $L^{\perp}=\{\boldsymbol{x}\in R^n : \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=0,\forall\,\boldsymbol{y}\in L\}$.给定 m 个向量 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_m\in R^n$,$\boldsymbol{a}_1,\cdots,\boldsymbol{a}_m$ 所有线性组合构成的集合称为 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_m$ 的**扩张子空间**,记为 $\mathrm{span}\{\boldsymbol{a}_1,\cdots,\boldsymbol{a}_m\}$.

给定矩阵 $\boldsymbol{A}\in R^{m\times n}$,与 $\boldsymbol{A}$ 有关的两个重要的子空间是 $\boldsymbol{A}$ 的**值域**(或**像空间**),定义为

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

根据扩张子空间的定义,$R(\boldsymbol{A})=\mathrm{span}\{\boldsymbol{a}_1,\cdots,\boldsymbol{a}_n\}$,即 $\boldsymbol{A}$ 的值域是 $\boldsymbol{A}$ 的列的扩张子空间;$\boldsymbol{A}$ 的**零空间**(或**核**),定义为

$$\ker\boldsymbol{A}=\{\boldsymbol{x}\in R^n : \boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}\}$$

特别地,$\ker\boldsymbol{A}^{\mathrm{T}}$ 表示 $\boldsymbol{A}$ 的左零空间.这两个子空间的维数与 $\boldsymbol{A}$ 的秩有如下关系:

$$\dim\ker(\boldsymbol{A})=n-\mathrm{rank}(\boldsymbol{A}),\quad \dim\ker(\boldsymbol{A}^{\mathrm{T}})=m-\mathrm{rank}(\boldsymbol{A}) \tag{1.1}$$

具体请参见文献(胡茂林,2007).

算法复杂度　计算机科学和工程中问题的求解最终都归结为矩阵运算.在本书中,将介绍各种与矩阵有关的算法.算法一般分成直接法和迭代法两类.**直接法**是指在没有误差的情况下,可在有限步得到所计算问题精确解的算法,比如,将在第 3 章介绍的求解线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 的 Gauss 消去法和求解最小二乘问题等方

法都是直接方法.相对地,另一类算法是**迭代法**,即采取逐次逼近的方法来逼近问题的精确解,而在任意有限步都不能得到精确解的算法.迭代法产生的是一序列收敛到问题真正解的逼近解,算法的每一步或迭代都产生一个新的逼近.理论上逼近序列是无限的,但在实际中,算法不能永远进行下去.一旦逼近解足够好,或可以被接受时,算法就停止,见第 3 章.达到这一目标的迭代步骤通常预先是不知道的,虽然有时能估计出.

对给定的算法来说,算法的快慢是衡量其性能优劣的一个重要标志.对于直接法,其运算量的大小通常可作为其快慢的一个主要标志.算法复杂度分析就是计算或估计算法的运算量,它是一个算法的所有运算次数的总和,即所有四则运算$+$,$-$,$\times$,$\div$的总和.运算量是基于 flop(floating octal point,浮点八进制)来计量的,一个 flop 就是一个浮点运算.例如,计算两个长度为 n 的向量内积需要 n 次乘法和 $n-1$ 次加法,因此需要 $2n-1$ 个 flops.对于 $\boldsymbol{A}\in R^{m\times n}$,$\boldsymbol{x}\in R^{n}$,$\boldsymbol{b}\in R^{m}$,计算 $\boldsymbol{Ax}+\boldsymbol{b}$需要 $2mn$个 flops,同样,矩阵 $\boldsymbol{A}$ 秩一更新 $\boldsymbol{A}+\boldsymbol{xy}^{\mathrm{T}}$,其中 $\boldsymbol{x}\in R^{m}$,$\boldsymbol{y}\in R^{n}$ 也需要 $2mn$个 flops,因此,对 $\boldsymbol{A}\in R^{m\times p}$,$\boldsymbol{B}\in R^{p\times n}$ 和 $\boldsymbol{C}\in R^{m\times n}$,计算 $\boldsymbol{AB}+\boldsymbol{C}$ 需要 $2mnp$ 个 flops.

flops 数的计算可以通过对算法最基本循环的运算数相加得到,例如,对矩阵乘法和加法,基本的算法是 $a_{ik}b_{kj}+c_{ij}$,它有 2 个 flops.通过简单的循环计数得到此计算需要进行 mnp 次,所以一般矩阵相乘和加法的计算需要 $2mnp$ flops.

现在考虑两个上三角矩阵乘积的运算量.矩阵 $\boldsymbol{A}$ 称为上三角的,如果 $\boldsymbol{A}$ 的元素 a_{ij} 在 $i>j$ 时为零(见 1.2 节).因为对任意上三角矩阵 $\boldsymbol{A},\boldsymbol{B}\in R^{n\times n}$,乘积 $\boldsymbol{C}=\boldsymbol{AB}$ 也是上三角的,即对任何 $k<i$ 或 $j<k$,都有 $a_{ik}b_{kj}=0$,因此

$$c_{ij}=\sum_{k=i}^{j}a_{ik}b_{kj}$$

所以,计算 c_{ij}($i\leqslant j$)需要 $2(j-i+1)$个 flops.两个矩阵乘积的运算量是

$$\sum_{i=1}^{n}\sum_{j=1}^{n}2(j-i+1)=\sum_{i=1}^{n}\sum_{j=1}^{n-i+1}2j\approx\sum_{i=1}^{n}\frac{2(n-i+1)^2}{2}=\sum_{i=1}^{n}i^2\approx\frac{n^3}{3}$$

因此三角矩阵乘积的计算量仅是一般稠密矩阵的六分之一.

对较大的 n,低阶项对 flops 数的影响较小,因此通常被舍去.例如,通过精确的浮点运算数的计算,上述算法需要 $3n^3+n^2+2n$个 flops.当 $n=10^3$ 时,精确的浮点数 3001000200 与近似数 3000000000 相对误差非常小.因此精确的 flops 数与近似估计 $3n^3$ 相比,多出的两项没有实质的意义.

应该指出,虽然运算量在一定程度上反映了算法的快慢程度,但它不是确定算法快慢的唯一因素.现代计算机的运算速度远远高于数据传输速度,因此,一个算法实际运行的效率在很大程度上依赖于数据传输量的快慢.基于 flops 数来衡量

算法的效率是粗糙的,因此讨论计算的效率不能过分地依赖于 flops 数的比较.例如,不能断定三角矩阵相乘要比满矩阵快 6 倍.flops 数只是一种简洁、但不一定完全准确的评价方式,它仅是影响效率的多个因素之一.

对于迭代法,除了对每步所需的运算量进行分析外,还需要对其收敛速度进行分析,将在第 3 章介绍向量范数后,讨论迭代的收敛速度问题.

特征值和特征向量　特征值和特征向量的求解是矩阵计算的基本问题之一,这里复习一下矩阵特征值和特征向量的一些基本性质,相关的求解算法将在第8～12 章介绍.

设 $\boldsymbol{A}\in C^{n\times n}$,如果对于 $\lambda\in C$,存在非零向量 $\boldsymbol{x}\in C^n$ 使得

$$\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}\tag{1.2}$$

则称 λ 是 $\boldsymbol{A}$ 的**特征值**,$\boldsymbol{x}$ 是 $\boldsymbol{A}$ 属于 λ 的**特征向量**,偶对$(\lambda,\boldsymbol{v})$称为 $\boldsymbol{A}$ 的**特征对**.求解一个方阵的特征值和特征向量称为**特征问题**.式(1.2)表示,$\boldsymbol{A}$ 对 $\boldsymbol{x}$ 的作用等价于用标量乘以 $\boldsymbol{x}$.如果 $\boldsymbol{x}$ 是 $\boldsymbol{A}$ 对应 λ 的特征向量,则 $\boldsymbol{x}$ 的非零数乘也是 $\boldsymbol{A}$ 对应 λ 的特征向量,即特征向量是与尺度无关的,通常取特征向量为单位向量,即 $\hat{\boldsymbol{x}}=\boldsymbol{x}/\|\boldsymbol{x}\|$.

$\boldsymbol{A}$ 的全体特征值组成的集合称为 $\boldsymbol{A}$ 的谱,记作 $\lambda(\boldsymbol{A})$.由式(1.1),$\lambda\in\lambda(\boldsymbol{A})$的充要条件是

$$\det(\lambda\boldsymbol{I}-\boldsymbol{A})=0$$

这是关于 λ 的 n 次多项式,称 $p(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})$为 $\boldsymbol{A}$ 的**特征多项式**.因为特征多项式的根一般是复数,所以研究矩阵的特征问题一般是在复域中进行的.

根据方阵的行列式和它的转置的行列式是相同的,则 $\det(\lambda\boldsymbol{I}-\boldsymbol{A})=\det(\lambda\boldsymbol{I}-\boldsymbol{A}^{\mathrm{T}})$,因此 $\lambda(\boldsymbol{A})=\lambda(\boldsymbol{A}^{\mathrm{T}})$.于是,对任意的 $\lambda\in\lambda(\boldsymbol{A})$,一定存在非零向量 $\boldsymbol{y}\in C^n$,使得 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}=\lambda\boldsymbol{y}$,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**;相对应地,属于 λ 的特征向量亦称作 $\boldsymbol{A}$ 属于 λ 的**右特征向量**.需要注意的是虽然矩阵和它的转置的特征值是相同的,但是通常左、右特征向量是不相等的,读者可以举例说明.

几何重数与代数重数　方阵 $\boldsymbol{A}$ 的代数重数就是它的特征多项式根的重数,即如果 $\boldsymbol{A}$ 有 r 个互不相同的特征值 $\lambda_1,\lambda_2,\cdots,\lambda_r$,且

$$p(\lambda)=\prod_{i=1}^{r}(\lambda-\lambda_i)^{n(\lambda_i)},\quad \lambda_i\neq\lambda_j(i\neq j),\quad \sum_{i=1}^{r}n(\lambda_i)=n$$

正整数 $n(\lambda_i)$称为 λ_i 的**代数重数**.

对应 λ 的特征向量组成的子空间称为 $\boldsymbol{A}$ 的**特征空间**,即 $\lambda_i\boldsymbol{I}-\boldsymbol{A}$ 的零空间,记为 L_λ,它是 $\boldsymbol{A}$ 的不变子空间,即 $\boldsymbol{A}L_\lambda\subseteq L_\lambda$.由式(1.1),$\lambda_i$ 的特征空间的维数是

$$m(\lambda_i)=n-\operatorname{rank}(\lambda_i\boldsymbol{I}-\boldsymbol{A})$$

称为 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.显然有

$$1 \leqslant m(\lambda_i) \leqslant n(\lambda_i) \leqslant n$$

即几何重数总不大于代数重数.例如,矩阵 $\boldsymbol{A}=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$ 的特征值 1 的代数重数是 2,而几何重数是 1,因此几何重数和代数重数是两个不同的概念.习惯上,将**代数重数**简称为**重数**.在今后的叙述中,说 λ 是 $\boldsymbol{A}$ 的 k 重特征值,是指 λ 的代数重数是 k.通常将代数重数为 1 的特征值称作**单特征值**.

代数重数等于几何重数的特征值称为**半单特征值**,否则称为**亏损特征值**.如果 $\boldsymbol{A}$ 的所有特征值都是半单的,则称 $\boldsymbol{A}$ 为**半单矩阵**(或**非亏损**的),有一个或多个亏损特征值的矩阵称为**亏损矩阵**.所有对角矩阵都是非亏损的,对于这样的矩阵,特征值 λ 的代数重数和几何重数均等于 λ 在对角线上出现的个数.

相似矩阵　下面介绍矩阵的相似变换,相似变换保持矩阵的谱,因此是一个非常重要的变换,矩阵的特征值求解算法的思想就是用相似变换将矩阵变换到特征值容易求解的形式.

两个矩阵 $\boldsymbol{A},\boldsymbol{B}\in C^{n\times n}$ 称为**相似**的,如果存在 n 阶非奇异方阵 $\boldsymbol{X}$,使得 $\boldsymbol{A}=\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}$.与对角矩阵相似的矩阵称为**可对角化矩阵**.

容易验证,如果 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相似,则 $\det(\lambda\boldsymbol{I}-\boldsymbol{A})=\det(\lambda\boldsymbol{I}-\boldsymbol{B})$.因此,相似矩阵有相同的特征多项式,从而有相同的特征值和代数重数.事实上,由于 $\boldsymbol{X}$ 是非奇异的,可以得到,相似矩阵也有相同的几何重数.

基于相似矩阵就能证明特征值的代数重数不小于几何重数.假设 $\boldsymbol{A}$ 的特征值 λ 的几何重数为 k,取特征空间 L_λ 的 k 个标准正交基 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k$.由 $\boldsymbol{A}\boldsymbol{x}_i=\lambda\boldsymbol{x}_i$,得

$$\boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}_j=\begin{cases}\lambda, & i=j\\ 0, & i\neq j\end{cases}$$

再补充 $n-k$ 个向量,使它们构成 C^n 的一组标准正交基.因此,它们按列组成一个酉矩阵 $\boldsymbol{V}$,则由上式得

$$\boldsymbol{B}=\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V}=\begin{bmatrix}\lambda\boldsymbol{I} & \boldsymbol{C}\\ \boldsymbol{0} & \boldsymbol{D}\end{bmatrix}$$

其中 $\boldsymbol{I}\in C^{k\times k}$是单位矩阵,$\boldsymbol{C}\in C^{k\times(n-k)}$,$\boldsymbol{D}\in C^{(n-k)\times(n-k)}$.根据行列式的定义,有

$$\det(z\boldsymbol{I}-\boldsymbol{B})=\det(z\boldsymbol{I}-\lambda\boldsymbol{I})\det(z\boldsymbol{I}-\boldsymbol{D})=(z-\lambda)^k\det(z\boldsymbol{I}-\boldsymbol{D})$$

因此,$\boldsymbol{B}$ 的特征值 λ 的代数重数至少为 k.由于相似变换保持重数不变,因此 $\boldsymbol{A}$ 的特征值 λ 的代数重数至少为 k,即矩阵特征值的代数重数大于几何重数.

特征值求解和多项式求解等价性　上面说明了求矩阵 $\boldsymbol{A}$ 的特征值等价于求

解多项式$\det(\lambda \boldsymbol{I}-\boldsymbol{A})=0$的根.因此,如果能有方法求任意多项式的根,则理论上就能求出任意矩阵的特征值.下面的定理1表明反过来也是正确的,即如果能有方法求任意矩阵的特征值,则理论上能求任意阶多项式的根.

给定n次多项式$q(\lambda)=b_n\lambda^n+b_{n-1}\lambda^{n-1}+\cdots+b_0$,其中$b_n\neq0$.如果$q(\lambda)$的所有的系数除以$b_n$,则得到新的方程$p(\lambda)=\lambda^n+a_{n-1}\lambda^{n-1}+\cdots+a_0\,(a_k=b_k/b_n)$,它与原方程有相同的根.因为$p(\lambda)$的首项系数$a_n=1$,所以称为**首一多项式**.因为多项式$q(\lambda)=0$总是能被等价的$p(\lambda)=0$代替,所以只要考虑首一多项式就可以了.给定首一多项式p,定义它的$n\times n$**友矩阵**为

$$\boldsymbol{A}=\begin{bmatrix} -a_{n-1} & -a_{n-2} & -a_{n-3} & \cdots & -a_0 \\ 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \end{bmatrix} \tag{1.3}$$

其中次对角上元素全为1,第一行元素依次由多项式p的后n项系数的负值组成,其余元素都是零.

定理1 设$p(\lambda)=\lambda^n+a_{n-1}\lambda^{n-1}+\cdots+a_0$是次数为$n$的首一多项式,且它的友矩阵$\boldsymbol{A}$为式(1.3),则$\boldsymbol{A}$的特征多项式是$p(\lambda)$,即方程$p(\lambda)=0$的根是$\boldsymbol{A}$的特征值.

迭代的必要性 定理1给出了求矩阵的特征值问题和多项式求解问题的等价性.多项式求解是一个古老的问题,引起了许多著名科学家的兴趣.特别地,在1824年,Abel证明了对$n>4$的n次多项式,没有类似于二次多项式的求根公式.Abel理论表明,没有直接的方法求解矩阵的特征值.因为如果存在有限步的方法将表明存在任意多项式方程的求解公式,即使很复杂.因此,由定理1,对$n>4$的$n\times n$矩阵的特征值没有解的表达式,必须用迭代的方法求解.

矩阵多项式 因为方阵的任何正整数幂和线性组合都是同阶的矩阵,所以多项式

$$q(t)=a_kt^k+a_{k-1}t^{k-1}+\cdots+a_1t+a_0$$

在矩阵$\boldsymbol{A}\in R^{n\times n}$处取值是有意义的.定义**矩阵多项式**为

$$q(\boldsymbol{A})\equiv a_k\boldsymbol{A}^k+a_{k-1}\boldsymbol{A}^{k-1}+\cdots+a_1\boldsymbol{A}+a_0\boldsymbol{I}$$

矩阵多项式$q(\boldsymbol{A})$与$\boldsymbol{A}$的特征向量是相同的,而其特征值与$\boldsymbol{A}$的特征值有如下的简单关系.

定理2 给定多项式$q(t)$,如果λ是$\boldsymbol{A}\in R^{n\times n}$的特征值,$\boldsymbol{x}$是对应的特征向量,那么$q(\lambda)$是矩阵$q(\boldsymbol{A})$的特征值,且$\boldsymbol{x}$是对应$q(\lambda)$的特征向量.

证明　首先,反复应用特征值-特征向量方程(1.2),得

$$\boldsymbol{A}^j\boldsymbol{x} = \boldsymbol{A}^{j-1}(\boldsymbol{A}\boldsymbol{x}) = \lambda\boldsymbol{A}^{j-1}\boldsymbol{x} = \cdots = \lambda^j\boldsymbol{x}$$

于是

$$\begin{aligned} q(\boldsymbol{A})\boldsymbol{x} &= a_k\boldsymbol{A}^k\boldsymbol{x} + a_{k-1}\boldsymbol{A}^{k-1}\boldsymbol{x} + \cdots + a_1\boldsymbol{A}\boldsymbol{x} + a_0\boldsymbol{x} \\ &= a_k\lambda^k\boldsymbol{x} + \cdots + a_1\lambda\boldsymbol{x} + a_0\boldsymbol{x} \\ &= (a_k\lambda^k + \cdots + a_1\lambda + a_0)\boldsymbol{x} = q(\lambda)\boldsymbol{x} \end{aligned}$$

因此,$q(\boldsymbol{A})$的特征值是 $q(\lambda)$,而特征向量是相同的. □

矩阵的对角化　根据代数基本定理,特征多项式 $p(\lambda)$最多有 n 个根,它们分别是 $\boldsymbol{A}\in R^{n\times n}$的特征值.

如果 $\boldsymbol{A}$ 有线性独立的特征向量$\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$,它们构成了 R^n 的基称为**完备特征向量系**.在这个基下,矩阵 $\boldsymbol{A}$ 是可对角化的(胡茂林,2007).此时有下面的定理.

定理 3　n阶矩阵 $\boldsymbol{A}$ 与对角矩阵相似,当且仅当 $\boldsymbol{A}$ 有 n 个线性无关的特征向量.

证明　如果 $\boldsymbol{A}$ 与对角矩阵相似,即存在非奇异矩阵 $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]$,使得

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

两边左乘以 $\boldsymbol{X}$,得

$$\boldsymbol{A}[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n] = \boldsymbol{X}\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n) = [\lambda_1\boldsymbol{x}_1,\lambda_2\boldsymbol{x}_2,\cdots,\lambda_n\boldsymbol{x}_n]$$

即

$$\boldsymbol{A}\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i,\quad i = 1,2,\cdots,n$$

这就表明 $\boldsymbol{X}$ 的列向量是 $\boldsymbol{A}$ 的特征向量.又因为 $\boldsymbol{X}$ 是非奇异的,所以 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 是 $\boldsymbol{A}$ 的 n 个线性无关的特征向量.

反之,如果 $\boldsymbol{A}$ 有 n 个线性无关的特征向量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$,则

$$\boldsymbol{A}\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i,\quad i = 1,2,\cdots,n$$

记 $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]$,显然 $\boldsymbol{X}$ 是非奇异的.又因为

$$\begin{aligned} \boldsymbol{A}\boldsymbol{X} &= [\boldsymbol{A}\boldsymbol{x}_1,\boldsymbol{A}\boldsymbol{x}_2,\cdots,\boldsymbol{A}\boldsymbol{x}_n] = [\lambda_1\boldsymbol{x}_1,\lambda_2\boldsymbol{x}_2,\cdots,\lambda_n\boldsymbol{x}_n] \\ &= [\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n) = \boldsymbol{X}\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n) \end{aligned}$$

所以

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

故 $\boldsymbol{A}$ 与对角矩阵 $\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$相似.　□

需要指出,在 $\boldsymbol{X}$ 的构造中,应该保证 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 与 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 的排列顺序保持一致,否则 $\boldsymbol{X}$ 就不是原来的形式.矩阵 $\boldsymbol{A}$ 有 n 个特征向量,即映射 $\boldsymbol{A}$ 有 n 个不变子空间,根据不变子空间的理论,$\boldsymbol{A}$ 一定可对角化(胡茂林,2007),因此这个定理是这个结论的不同叙述.

现在说明,如果 $\boldsymbol{A}$ 有 n 个不同的特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$,则对应的特征向量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 是线性独立的.这可以通过反证法证明.如果 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 是线性相关的,不妨假设其最大线性无关向量个数是 k,且前 k 个向量是线性无关的(如有必要,可以重排编序).因此存在 $\alpha_1,\cdots,\alpha_k\in C$,使得

$$\boldsymbol{x}_{k+1}=\alpha_1\boldsymbol{x}_1+\cdots+\alpha_r\boldsymbol{x}_k$$

用 $\boldsymbol{A}$ 对上式作用,得

$$\lambda_{k+1}\boldsymbol{x}_{k+1}=\boldsymbol{A}\boldsymbol{x}_{k+1}=\boldsymbol{A}(\alpha_1\boldsymbol{x}_1+\cdots+\alpha_k\boldsymbol{x}_k)=\alpha_1\lambda_1\boldsymbol{x}_1+\cdots+\alpha_k\lambda_k\boldsymbol{x}_k$$

将 $\boldsymbol{x}_{k+1}$ 的表达式代入到上式,得

$$\alpha_1(\lambda_1-\lambda_{k+1})\boldsymbol{x}_1+\cdots+\alpha_k(\lambda_k-\lambda_{k+1})\boldsymbol{x}_k=\boldsymbol{0}$$

而 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k$ 线性无关,且 $\lambda_i-\lambda_{k+1}\neq 0$, $i=1,\cdots,k$,因此 $\alpha_1=\cdots=\alpha_k=0$.这与 $\boldsymbol{x}_{k+1}$ 线性依赖于 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k$ 矛盾,因此假设是错误的,$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 线性无关.

因此,由定理 3, 如果 n 阶矩阵有 n 个不同的特征值,则它与对角矩阵相似.

特征问题简化　在特征值求解过程中,通常是把一个给定的问题分解为若干小的问题来逐一解决,即将矩阵分块为较小的矩阵,对小矩阵求解特征问题,得到对大的矩阵特征求解.下面是相关的简化理论.

定理 4　如果 $A\in C^{n\times n}$有如下分块形式:

$$\begin{array}{cl}\boldsymbol{A}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{0} & \boldsymbol{A}_{22}\end{bmatrix} & \begin{matrix}k\\ n-k\end{matrix}\\ \quad\ \begin{matrix}k & \quad n-k\end{matrix} & \end{array}$$

则 $\lambda(\boldsymbol{A})=\lambda(\boldsymbol{A}_{11})\cup\lambda(\boldsymbol{A}_{22})$.

证明　若 $\lambda,\boldsymbol{x}$ 是 $\boldsymbol{A}$ 的特征对,则

$$\boldsymbol{A}\boldsymbol{x}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{0} & \boldsymbol{A}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}_{11}\boldsymbol{x}_1+\boldsymbol{A}_{22}\boldsymbol{x}_2\\ \boldsymbol{A}_{22}\boldsymbol{x}_2\end{bmatrix}=\lambda\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}$$

其中 $\boldsymbol{x}_1\in C^k$ 和 $\boldsymbol{x}_2\in C^{n-k}$是 $\boldsymbol{x}$ 的分划.如果 $\boldsymbol{x}_2\neq\boldsymbol{0}$,那么 $\boldsymbol{A}_{22}\boldsymbol{x}_2=\lambda\boldsymbol{x}_2$,因此 λ 是 $\boldsymbol{A}_{22}$ 的特征值,即 $\lambda\in\lambda(\boldsymbol{A}_{22})$.如果 $\boldsymbol{x}_2=\boldsymbol{0}$,则 $\boldsymbol{A}_{11}\boldsymbol{x}_1=\lambda\boldsymbol{x}_1$,此时 λ 是 $\boldsymbol{A}_{11}$ 的特征值,即 $\lambda\in\lambda(\boldsymbol{A}_{11})$.由此可得

$$\lambda(\boldsymbol{A})\subseteq\lambda(\boldsymbol{A}_{11})\cup\lambda(\boldsymbol{A}_{22})$$

又因为 $\lambda(\boldsymbol{A})$和 $\lambda(\boldsymbol{A}_{11})\cup\lambda(\boldsymbol{A}_{22})$的个数都是 n,所以,上面的集合包含关系应该是等价的,得到结论. □

注意这个结论只是关于特征值的,对特征向量没有类似的结果,特征向量的求解必须整体地考虑;但是如果知道特征值,求矩阵的特征向量并不是特别困难的,见第 9 章.

我们知道相似变换可以将矩阵简化,且简化后的矩阵与原矩阵的特征值相同.将这个事实与定理 4 结合起来,有下面的结论.

定理 5　如果矩阵 $\boldsymbol{A}\in C^{n\times n}$,$\boldsymbol{B}\in C^{k\times k}$和 $\boldsymbol{X}\in C^{n\times k}$,其中 rank $\boldsymbol{X}=k$,满足

$$\boldsymbol{AX}=\boldsymbol{XB} \tag{1.4}$$

则存在 Hermite 矩阵 $\boldsymbol{Q}\in C^{n\times n}$,使得

$$\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{AQ}=\boldsymbol{T}=\begin{array}{cc} \begin{bmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} \\ \boldsymbol{0} & \boldsymbol{T}_{22} \end{bmatrix} & \begin{array}{l} k \\ n-k \end{array} \\ \begin{array}{cc} k & n-k \end{array} & \end{array} \tag{1.5}$$

且 $\lambda(\boldsymbol{T}_{11})=\lambda(\boldsymbol{A})\cap\lambda(\boldsymbol{B})$.

证明　记 $\boldsymbol{X}=[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k]$,由 rank $\boldsymbol{X}=k$,$\boldsymbol{AX}=\boldsymbol{XB}$ 说明 $\boldsymbol{AX}$ 的每一列可以用 $\boldsymbol{X}$ 的列向量的线性组合表示,即子空间 span$\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k\}$是 $\boldsymbol{A}$ 的不变子空间.对向量 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k$ 进行 Gram-Schmidt 正交化,得

$$\begin{aligned} \boldsymbol{x}_1 &= r_{11}\boldsymbol{q}_1 \\ \boldsymbol{x}_2 &= r_{21}\boldsymbol{q}_1+r_{22}\boldsymbol{q}_2 \\ &\vdots \\ \boldsymbol{x}_k &= r_{k1}\boldsymbol{q}_1+r_{k2}\boldsymbol{q}_2+\cdots+r_{kk}\boldsymbol{q}_k \end{aligned}$$

因此

$$\boldsymbol{X}=[\boldsymbol{q}_1,\vdots,\boldsymbol{q}_k]\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ & r_{22} & \cdots & r_{2k} \\ & & \ddots & \vdots \\ & & & r_{kk} \end{bmatrix}$$

记

$$\boldsymbol{R}_1=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ & r_{22} & \cdots & r_{2k} \\ & & \ddots & \vdots \\ & & & r_{kk} \end{bmatrix}$$

由 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k$ 是线性无关的,得 $\boldsymbol{R}_1\in C^{k\times k}$是非奇异的.将向量 $\boldsymbol{q}_1,\cdots,\boldsymbol{q}_k$ 扩展到 C^n

的标准正交基 $\boldsymbol{q}_1,\cdots,\boldsymbol{q}_k,\boldsymbol{q}_{k+1},\cdots,\boldsymbol{q}_n$,且记 $\boldsymbol{Q}=[\boldsymbol{q}_1,\cdots,\boldsymbol{q}_n]$,则

$$\boldsymbol{X}=\boldsymbol{Q}\boldsymbol{R} \tag{1.6}$$

其中 $\boldsymbol{R}=\begin{bmatrix}\boldsymbol{R}_1\\ \boldsymbol{0}\end{bmatrix}$.上式称为 $\boldsymbol{X}$ 的 QR 分解(见第 5 章).将式(1.6)代入到式(1.4)中,得

$$\begin{bmatrix}\boldsymbol{T}_{11} & \boldsymbol{T}_{12}\\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{R}_1\\ \boldsymbol{0}\end{bmatrix}=\begin{bmatrix}\boldsymbol{R}_1\\ \boldsymbol{0}\end{bmatrix}\boldsymbol{B}$$

其中

$$\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{Q}=\boldsymbol{T}=\begin{bmatrix}\boldsymbol{T}_{11} & \boldsymbol{T}_{12}\\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22}\end{bmatrix}$$

因此得到两个方程组 $\boldsymbol{T}_{21}\boldsymbol{R}_1=\boldsymbol{0}$ 和 $\boldsymbol{T}_{11}\boldsymbol{R}_1=\boldsymbol{R}_1\boldsymbol{B}$.因为 $\boldsymbol{R}_1$ 是非奇异的,由前一个方程得 $\boldsymbol{T}_{21}=\boldsymbol{0}$,后一个方程表示 $\boldsymbol{T}_{11}$ 与 $\boldsymbol{B}$ 是相似矩阵,因此 $\lambda(\boldsymbol{T}_{11})=\lambda(\boldsymbol{B})$.由定理 4,$\lambda(\boldsymbol{A})=\lambda(\boldsymbol{T}_{11})\cup\lambda(\boldsymbol{T}_{22})$得到结论. □

式(1.5)实际上表示由不变子空间可以将矩阵简化.

Schur 分解 下面在复空间中,介绍将给定的向量变换成与坐标向量 $\boldsymbol{e}_1$ 平行的线性变换.对于任意的 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}\in C^n$,$\boldsymbol{x}\neq\boldsymbol{0}$,记

$$p=\begin{cases}\|\boldsymbol{x}\|_2, & x_1=0\\ -\mathrm{e}^{\mathrm{i}\arg x_1}\|\boldsymbol{x}\|_2, & x_1\neq 0\end{cases}$$

其中 $\|\boldsymbol{x}\|_2=\sqrt{x_1^2+\cdots+x_n^2}$是复平面上的欧氏范数.定义

$$\boldsymbol{H}(\omega)=\boldsymbol{I}-2\omega\omega^{\mathrm{H}} \tag{1.7}$$

其中

$$\omega=(\boldsymbol{x}-p\boldsymbol{e}_1)/\|\boldsymbol{x}-p\boldsymbol{e}_1\|_2$$

则可以直接验证有

$$\boldsymbol{H}(\omega)\boldsymbol{x}=p\boldsymbol{e}_1$$

容易证明,式(1.7)所定义的矩阵 $\boldsymbol{H}(\omega)$既是 Hermite 矩阵又是酉矩阵,即

$$\boldsymbol{H}(\omega)^{\mathrm{H}}=\boldsymbol{H}(\omega)=\boldsymbol{H}(\omega)^{-1}$$

在实空间中,$\boldsymbol{H}(\omega)$对应的是 Housholder 矩阵,其几何意义见第 4 章.利用变换 $\boldsymbol{H}(\omega)$和数学归纳法可以证明下面的定理.

定理 6(Schur 分解) 给定 $\boldsymbol{A}\in C^{n\times n}$,则存在酉矩阵 $\boldsymbol{U}$,使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\boldsymbol{T} \tag{1.8}$$

其中 $\boldsymbol{T}$ 是上三角矩阵，且适当选取 $\boldsymbol{U}$，可使 $\boldsymbol{T}$ 的对角元素按任意指定的顺序排列.

证明　用数学归纳法对矩阵的阶数 n 进行归纳证明.

当 $n=1$ 时，定理显然成立.下面考虑 $n>1$ 的情形，并假设定理在 $n-1$ 时成立.

如果 $\lambda\in\lambda(\boldsymbol{A})$ 是排在左上角的特征值，且 $\boldsymbol{x}\in C^n$ 是对应 λ 的特征向量，即 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$.构造 Hermite 酉矩阵 $\boldsymbol{H}$ 和非零常数 p，使得 $\boldsymbol{H}\boldsymbol{x}=p\boldsymbol{e}_1$.由 $\boldsymbol{H}=\boldsymbol{H}^{-1}$ 得

$$\boldsymbol{H}\boldsymbol{e}_1=\frac{1}{p}\boldsymbol{x}$$

因此

$$\boldsymbol{H}\boldsymbol{A}\boldsymbol{H}\boldsymbol{e}_1=\boldsymbol{H}\boldsymbol{A}\left(\frac{1}{p}\boldsymbol{x}\right)=\lambda\boldsymbol{e}_1$$

这表明 $\boldsymbol{HAH}$ 有如下形式：

$$\boldsymbol{HAH}=\boldsymbol{HAHI}=\begin{bmatrix}\lambda & * \\ \boldsymbol{0} & \boldsymbol{A}_1\end{bmatrix}$$

其中 $\boldsymbol{A}_1$ 是 $n-1$ 阶方阵.由定理 4，$\lambda(\boldsymbol{A}_1)=\lambda(\boldsymbol{A})-\{\lambda\}$，再由归纳假设知，存在 $\boldsymbol{U}_1\in U(n-1)$ 使得

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

其中 $\boldsymbol{T}_1$ 是上三角矩阵，而且 $\boldsymbol{T}_1$ 的对角元素可以按指定的次序排列.记

$$\boldsymbol{U}=\boldsymbol{H}\begin{bmatrix}1 & \boldsymbol{0} \\ \boldsymbol{0}^{\mathrm{T}} & \boldsymbol{U}_1\end{bmatrix}$$

则 $\boldsymbol{U}$ 是酉矩阵，且 $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}$ 是上三角矩阵.　□

定理 6 的证明过程给出了求一个矩阵特征值的思想：可以依次求一个阶数较低的矩阵，来得到原矩阵的其余特征值.具体地说，在已知矩阵 $\boldsymbol{A}$ 的一个特征值和对应的特征向量时，如果需要求其他的特征值和特征向量，可以对低一阶的矩阵 $\boldsymbol{A}_1$ 进行处理，这个方法称为**收缩**.

分解式(1.8)称作 $\boldsymbol{A}$ 的 **Schur 分解**，其右端的 $\boldsymbol{T}$ 称作 $\boldsymbol{A}$ 的 **Schur 标准型**，可记为

$$\boldsymbol{T}=\boldsymbol{D}+\boldsymbol{M},\quad \boldsymbol{D}=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$$

其中 $\boldsymbol{M}$ 为严格上三角矩阵，$\lambda(\boldsymbol{A})=\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$.

如果 $\boldsymbol{A}$ 是 Hermite 矩阵，即 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}$，则 $(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U})^{\mathrm{H}}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}^{\mathrm{H}\mathrm{H}}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}$，即

$U^{\mathrm{H}}AU$ 也是 Hermite 矩阵,而 Hermite 上三角矩阵一定是对角矩阵,因此 T 是对角矩阵.又因为上三角正规矩阵也是对角矩阵,所以从定理 6 可立即得到下面的结论.

推论 1 设 $A\in C^{n\times n}$,则:

(1) A 是正规矩阵当且仅当存在酉矩阵 U,使得 $U^{\mathrm{H}}AU=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$.

(2) A 是 Hermite 矩阵当且仅当存在酉矩阵 U 和实对角矩阵 D,使得 $U^{\mathrm{H}}AU=D$.

实 Schur 分解 在实际问题中,大部分特征问题的求解都是关于实矩阵的,因此给出实的 Schur 分解是有意义的.所谓实 Schur 分解就是分解式(1.8)中的矩阵元素都是实的.因为实矩阵的特征值可能是复的,所以不会有“特征值显现”的三角阵,必须降低要求.

定理 7(实 Schur 分解) 若 $A\in R^{n\times n}$,则存在正交矩阵 $Q\in R^{n\times n}$,使得

$$Q^{\mathrm{T}}AQ=\begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1k} \\ 0 & R_{22} & \cdots & R_{2k} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & R_{kk} \end{bmatrix} \tag{1.9}$$

其中每个 $R_{ii}(i=1,\cdots,k)$是实数或二阶实矩阵.

证明 因为 A 的特征多项式 $\det(\lambda I-A)$的系数都是实数,所以复特征值一定是共轭出现的.记 $\lambda(A)$中复共轭对的个数为 l.对 l 进行归纳证明.当 $l=0$ 时,此时特征值都是实的,且特征向量也是实向量,这就是分解式(1.8),其中元素都是实的.现在假设 $l\geqslant 1$,如果 $\lambda=\alpha+\beta\mathrm{i}$,且 $\beta\neq 0$,那么一定存在线性无关向量 $y,z\in R^n$,使得

$$A(y+\mathrm{i}z)=(\alpha+\mathrm{i}\beta)(y+\mathrm{i}z)$$

否则 $\beta=0$,即

$$A[y,z]=[y,z]\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$$

因此 y 和 z 张成 A 的一个二维不变子空间.由定理 5,存在正交矩阵 $U\in R^{n\times n}$,使得

$$U^{\mathrm{T}}AU=\begin{array}{cc} \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} & \begin{array}{c} 2 \\ n-2 \end{array} \\ \begin{array}{cc} 2 & \quad n-2 \end{array} & \end{array}$$

其中 $\lambda(T_{11})=\{\lambda,\bar{\lambda}\}$.由归纳法,存在正交矩阵 $\tilde{U}\in R^{(n-2)\times(n-2)}$ 使得 $\tilde{U}^{\mathrm{T}}T_{22}\tilde{U}$ 具有所需的结构.记 $Q=U\begin{bmatrix} I_{2\times 2} & 0 \\ 0 & \tilde{U} \end{bmatrix}$,则分解式(1.9)成立. □

实 Schur 分解式(1.9)的右边称为拟上三角阵,这是一个简单的拟上三角阵(见 1.2 节).这个定理表明任一实阵可以正交相似于一个拟上三角阵,显然复特征值的实部和虚部都可以由 2×2 的对角块阵得到.

1.2　特殊矩阵及其性质

在计算机科学和工程研究中,经常会遇到一些特殊形式的矩阵,其中矩阵元素之间存在着一定关系.了解这些矩阵的内部结构,有助于灵活地使用这些矩阵,简化一些问题的表达和求解.本节将介绍本书所用到的一些特殊矩阵.

对角矩阵　如果矩阵 $\boldsymbol{D}=[d_{ij}]\in R^{n\times n}$,当 $i\neq j$ 时,有 $d_{ij}=0$,就称 $\boldsymbol{D}$ 为**对角矩阵**,即对角矩阵的主对角线以外的元素全为零,主对角线上的元素也可以为零,如果全部是零,则为零矩阵.通常记对角矩阵为 $\boldsymbol{D}=\mathrm{diag}(d_{11},\cdots,d_{nn})$或 $\boldsymbol{D}=\mathrm{diag}\boldsymbol{d}$,其中 $\boldsymbol{d}$ 是由 $\boldsymbol{D}$ 的对角元素组成的 n 维向量.对一般的矩阵 $\boldsymbol{X}\in C^{n\times n}$,$\mathrm{diag}\boldsymbol{X}$ 表示取 $\boldsymbol{X}$ 的对角元素组成的对角矩阵.如果一个对角矩阵的所有对角元都是正(非负)实数,就称它为**正(非负)对角矩阵**.注意,术语正对角矩阵指的是,矩阵是对角的,且对角元素是正的,而不是指所有对角元素是正的一般矩阵.正对角矩阵的一个例子是单位矩阵.如果对角矩阵 $\boldsymbol{D}$ 的各对角元素都是相等的,就称 $\boldsymbol{D}$ 为**纯量矩阵**,即存在数 $\alpha\in C$,有 $\boldsymbol{D}=\alpha\boldsymbol{I}$.矩阵左乘或右乘纯量矩阵等价于数乘矩阵.

对角矩阵的行列式正好是对角元素的乘积,即 $\det\boldsymbol{D}=\prod\limits_{i=1}^{n}d_{ii}$.因而,一个对角矩阵是非奇异的,当且仅当所有对角元素非零.对角矩阵 $\boldsymbol{D}$ 左乘 $\boldsymbol{A}\in R^{n\times n}$,即 $\boldsymbol{DA}$,就是用 $\boldsymbol{D}$ 的各对角元素乘 $\boldsymbol{A}$ 的各行,即 $\boldsymbol{A}$ 的第 i 行乘 d_{ii},$i=1,2,\cdots,n$.而 $\boldsymbol{D}$ 右乘 $\boldsymbol{A}$,即 $\boldsymbol{AD}$,就是用 $\boldsymbol{D}$ 的各对角元素乘 $\boldsymbol{A}$ 的各列.特别地,两个对角矩阵的乘积是可交换的.

分块对角矩阵　具有形式

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{A}_{11} & & & \boldsymbol{0}\\ & \boldsymbol{A}_{22} & & \\ & & \ddots & \\ \boldsymbol{0} & & & \boldsymbol{A}_{kk}\end{bmatrix}$$

的矩阵 $\boldsymbol{A}\in R^{n\times n}$称为**分块对角矩阵**,其中 $\boldsymbol{A}_{ii}\in R^{n_i\times n_i}$,$i=1,2,\cdots,k$ 且 $\sum\limits_{i=1}^{k}n_i=n$.由文献(胡茂林,2007)知,这样的矩阵对应的映射有 k 个不变子空间,因此,这个矩阵一般用 $\boldsymbol{A}=\boldsymbol{A}_{11}\oplus\boldsymbol{A}_{22}\oplus\cdots\oplus\boldsymbol{A}_{kk}$ 来表示,简记作$\oplus\sum\limits_{i=1}^{k}\boldsymbol{A}_{ii}$,称矩阵 $\boldsymbol{A}$ 是 $\boldsymbol{A}_{11},\cdots,\boldsymbol{A}_{kk}$

的**直和**.从乘法角度来考虑分块矩阵,分块对角矩阵的许多性质是对角矩阵的推广,例如,$\det(\oplus\sum_{i=1}^{k}\boldsymbol{A}_{ii})=\prod_{i=1}^{k}\det(\boldsymbol{A}_{ii})$,因而,$\boldsymbol{A}=\oplus\sum_{i=1}^{k}\boldsymbol{A}_{ii}$是非奇异的,当且仅当每个$\boldsymbol{A}_{ii}$,$i=1,\cdots,k$是非奇异的.另外,给定矩阵$\boldsymbol{A}=\oplus\sum_{i=1}^{k}\boldsymbol{A}_{ii}$与$\boldsymbol{B}=\oplus\sum_{i=1}^{k}\boldsymbol{B}_{ii}$,其中$\boldsymbol{A}_{ii}$,$\boldsymbol{B}_{ii}$都是同阶的,则$\boldsymbol{A}+\boldsymbol{B}=\oplus\sum_{i=1}^{k}(\boldsymbol{A}_{ii}+\boldsymbol{B}_{ii})$,$\boldsymbol{AB}$可交换当且仅当$\boldsymbol{A}_{ii}$和$\boldsymbol{B}_{ii}$的乘积可交换.

三角矩阵　如果当$i>j$时,$r_{ij}=0$,称矩阵$\boldsymbol{R}=[r_{ij}]\in C^{n\times n}$为**上三角矩阵**,如果当$i\geqslant j$时,$r_{ij}=0$,就称$\boldsymbol{R}$是**严格上三角矩阵**,类似地,$\boldsymbol{L}$称为**下三角**(或**严格下三角**)**矩阵**,是指它的转置是上三角(或严格上三角)矩阵.如果实三角矩阵$\boldsymbol{R}$的主对角元素全为正数,则称为**正线上三角矩阵**.与对角矩阵类似,三角矩阵的行列式是它的各对角元素的乘积,因此$\boldsymbol{R}$是非奇异的当且仅当主对角元素全不为零.上三角矩阵的乘积是上三角矩阵,上三角矩阵的逆是上三角矩阵.三角矩阵的秩至少是(可能大于)主对角线上非零元的个数.以上这些性质对下三角矩阵也成立.

分块三角矩阵　矩阵$\boldsymbol{A}\in C^{n\times n}$称为**分块上三角矩阵**,如果它具有形状

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{A}_{11} & & & * \\ & \boldsymbol{A}_{22} & & \\ & & \ddots & \\ \boldsymbol{0} & & & \boldsymbol{A}_{kk}\end{bmatrix}$$

其中$\boldsymbol{A}_{ii}\in C^{n_i\times n_i}$,$i=1,\cdots,k$,$\sum_{i=1}^{k}n_i=n$,而“$*$”表示可以为任意的元素,分块上三角矩阵称为**严格分块上三角矩阵**,如果它的所有对角子块都是零方阵.类似地,可以定义**分块下三角矩阵**、**严格分块下三角矩阵**.分块三角矩阵的行列式是各对角子块的行列式的乘积.分块三角矩阵的秩至少是(可能大于)诸对角子块秩的和.

三角方程组　经典线性方程组的求解通常是将给定的方程组转化为具有同解的两个三角方程组,即 Gauss 消元法(见第 4 章).下面讨论三角方程组的求解算法.

前代法　考虑如下非奇异的3×3下三角方程组:

$$\begin{bmatrix}l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33}\end{bmatrix}\begin{bmatrix}x_1 \\ x_2 \\ x_3\end{bmatrix}=\begin{bmatrix}b_1 \\ b_2 \\ b_3\end{bmatrix}$$

因为矩阵是非奇异的,对角元素$l_{ii}\neq 0$,$i=1,2,3$,所以未知数x_i,$i=1,2,3$可依次确定为

$$x_1 = b_1 / l_{11}$$
$$x_2 = (b_2 - l_{21} x_1) / l_{22}$$
$$x_3 = (b_3 - l_{31} x_1 - l_{32} x_2) / l_{33}$$

这就是方程组的**前代算法**,此算法可以立即推广到非奇异的 $n \times n$ 三角方程组 $\boldsymbol{Lx} = \boldsymbol{b}$.这个算法的一般形式是

$$x_1 = \frac{b_1}{l_{11}}$$
$$x_i = \frac{1}{l_{ii}}\left(b_i - \sum_{j=1}^{i-1} l_{ij} x_j \right), \quad i = 2, \cdots, n$$

算法所需要的$+,-,\times,\div$运算次数是 $\sum_{i=1}^{n}(2i-1) = 2\times\frac{n(n+1)}{2} - n = n^2$,因此该算法计算量是 n^2 flops.

回代法　类似地,给定 $n(n \geqslant 2)$阶非奇异的下三角矩阵 $\boldsymbol{R}$,从下到上解上三角方程组 $\boldsymbol{Rx} = \boldsymbol{b}$ 的算法叫**回代法**.其计算公式为

$$x_n = \frac{b_n}{r_{nn}}$$
$$x_i = \frac{1}{r_{ii}}\left(b_i - \sum_{j=i+1}^{n} r_{ij} x_j \right), \quad i = n-1, \cdots, 1$$

此算法计算量也是 n^2 flops.

从上述讨论可以看出,对三角方程组,解有非常简单的表达式,因此对一般的线性方程组,一般都是将它们转化为三角方程组来求解,转化的方法有三角(或 LU)分解和 QR 分解,将分别在第 4、5 章介绍.

Hessenberg 矩阵　在矩阵的特征值求解中,Hessenberg 矩阵起着重要的作用.矩阵 $\boldsymbol{A} = [a_{ij}] \in C^{n\times n}$称为**上 Hessenberg 矩阵**,如果对任何 $i > j+1$,都有 $a_{ij} = 0$,即

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & & \cdots & a_{2,n-1} & a_{2n} \\ 0 & a_{32} & & \cdots & a_{3,n-1} & a_{3n} \\ 0 & 0 & & & & \\ \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & a_{n,n-1} & a_{n,n} \end{bmatrix}$$

如果 $\boldsymbol{A}^{\mathrm{T}}$ 是上 Hessenberg 矩阵,就称 $\boldsymbol{A}$ 是**下 Hessenberg 矩阵**.如果 $\boldsymbol{A}$ 的所有次对角元素非零,则称为**不可约**的(见第 10 章,不可约矩阵部分).

在实际问题中,主要研究不可约的上 Hessenberg 矩阵,因为当某个次主对角元素为零时(对超大规模矩阵,这是常有的),上 Hessenberg 矩阵能被看做是两个不可约上 Hessenberg 矩阵,所以可以分别处理,这样就可以简化问题,即它能被分割成两个小问题.比如求某个上 Hessenberg 矩阵 $\boldsymbol{A}$ 的特征值问题,如果 $a_{i,i+1}=0$,即

$$\boldsymbol{A}=\begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{0} & \boldsymbol{B}_{22} \end{bmatrix}$$

其中 $\boldsymbol{B}_{11}\in C^{j\times j}$,$\boldsymbol{B}_{22}\in C^{k\times k}$,$j+k=n$.由定理 4,可以分别求 $\boldsymbol{B}_{11}$ 和 $\boldsymbol{B}_{22}$ 的特征值来得到 $\boldsymbol{A}$ 的特征值.如果分割点在矩阵的中间,就能节省许多计算量.因为如果对 Hessenberg 矩阵 $\boldsymbol{A}$ 求特征值的计算量是 $O(n^2)$,将 n 分割一半,求 $\boldsymbol{B}_{11}$ 和 $\boldsymbol{B}_{22}$ 特征值的计算量就分别减少 1/4,整个计算量减少 1/2.

三对角矩阵 如果矩阵 $\boldsymbol{A}=[a_{ij}]\in R^{n\times n}$既是上 Hessenberg 矩阵,又是下 Hessenberg 矩阵,就称为**三对角矩阵**,即当$|i-j|>1$时,$a_{ij}=0$,三对角矩阵的形式是

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} & & & \boldsymbol{0} \\ a_{21} & a_{22} & a_{23} & & \\ & a_{32} & a_{33} & & \\ & & & \ddots & \\ & & \ddots & \ddots & a_{n-1,n} \\ \boldsymbol{0} & & & a_{n,n-1} & a_{n,n} \end{bmatrix}$$

三对角矩阵的行列式可以用递推公式计算:记 $\boldsymbol{A}_k$ 为 k 阶三对角矩阵,对 $k=2,\cdots,n-1$,有

$$\det \boldsymbol{A}_{(k+1)\times(k+1)}=a_{k+1,k+1}\det\boldsymbol{A}_{k\times k}-a_{k+1,k}a_{k,k+1}\det\boldsymbol{A}_{(k-1)\times(k-1)}.$$

正交矩阵 如果实方阵 $\boldsymbol{U}$ 的转置等于它的逆矩阵,即 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}$,其中 $\boldsymbol{I}$ 是恒等矩阵,则称 $\boldsymbol{U}$ 为正交矩阵.它表示 $\boldsymbol{U}$ 的列向量的长度都是 1,且任意两个不同的列都是正交的.用 Kronecker 记号表示为 $\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_j=\delta_{ij}$.由条件 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}$,可以推出 $\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}=\boldsymbol{I}$.所以 $\boldsymbol{U}$ 的行向量长度也都是 1,并且两两是正交的.对矩阵等式 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}$ 两边取行列式,由 $\det\boldsymbol{U}=\det\boldsymbol{U}^{\mathrm{T}}$,得到$(\det\boldsymbol{U})^2=1$.因此,对正交矩阵 $\boldsymbol{U}$,有 $\det\boldsymbol{U}=\pm1$.

如果 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 是正交的,则$(\boldsymbol{UV})^{\mathrm{T}}(\boldsymbol{UV})=\boldsymbol{V}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{UV}=\boldsymbol{I}$,即正交矩阵的乘积仍然是正交矩阵,且正交矩阵的逆也是正交矩阵,因此所有 n 阶正交矩阵形成一个群,记为 $O(n)$.此外,行列式为正的 n 阶正交矩阵也形成一个子群,称为 n 维**旋转群**,记为 $SO(n)$.

正交矩阵的保范性质 向量 $\boldsymbol{x}$ 的长度定义为 $\|\boldsymbol{x}\|_2=(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})^{1/2}$,也称为欧氏范数(胡茂林,2007).正交矩阵的一个重要性质是向量乘以正交矩阵后,欧氏范数保

持不变.这可通过下面的简单计算进行验证

$$\|\boldsymbol{Ux}\|^2=(\boldsymbol{Ux})^{\mathrm{T}}(\boldsymbol{Ux})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{Ux}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=\|\boldsymbol{x}\|^2$$

因此正交矩阵对应的变换是**保距**的,即保持变换前后两点之间的距离.由上面的证明过程可以看出,正交变换也是保持内积的,因此正交变换也保持向量之间的夹角.

半正交矩阵　下面介绍具有正交矩阵部分性质的长方阵.

矩阵 $\boldsymbol{Q}\in R^{m\times n}$, $m\geqslant n$ 称为**半正交矩阵**,如果它的列向量在 R^m 中是正交的,即 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\boldsymbol{I}_n$.因为只有方阵才有逆矩阵,因此 $\boldsymbol{Q}^{-1}\neq\boldsymbol{Q}^{\mathrm{T}}$,且 $\boldsymbol{QQ}^{\mathrm{T}}$ 也不是单位矩阵.

对半正交矩阵 $\boldsymbol{Q}\in R^{m\times n}$, $m\geqslant n$,记它的列向量为 $\boldsymbol{q}_1,\boldsymbol{q}_2,\cdots,\boldsymbol{q}_n$,则可以简单地证明下面的三个结论:

(1) 如果 $\boldsymbol{x}\in R^m$ 与 $\boldsymbol{q}_1,\boldsymbol{q}_2,\cdots,\boldsymbol{q}_n$ 正交,则 $\boldsymbol{QQ}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{0}$.

(2) 对 $i=1,\cdots,n$, $\boldsymbol{QQ}^{\mathrm{T}}\boldsymbol{q}_i=\boldsymbol{q}_i$.因此在 R^m 的子空间 $\mathrm{span}\{\boldsymbol{q}_1,\cdots,\boldsymbol{q}_n\}$上,$\boldsymbol{QQ}^{\mathrm{T}}$ 类似于单位矩阵.

(3) $(\boldsymbol{QQ}^{\mathrm{T}})^2=\boldsymbol{QQ}^{\mathrm{T}}$,因此 $\boldsymbol{QQ}^{\mathrm{T}}$ 是一个投影矩阵.又因为$(\boldsymbol{QQ}^{\mathrm{T}})^{\mathrm{T}}=\boldsymbol{QQ}^{\mathrm{T}}$,因此是正交投影矩阵.

半正交矩阵也是保持距离的,即如果 $\boldsymbol{Q}\in R^{m\times n}$ 是半正交矩阵,则对所有的 $\boldsymbol{x}$, $\boldsymbol{y}\in R^n$,有

$$\langle\boldsymbol{Qx},\boldsymbol{Qy}\rangle=\langle\boldsymbol{x},\boldsymbol{y}\rangle \text{和} \|\boldsymbol{Qx}\|=\|\boldsymbol{x}\|$$

置换矩阵　如果矩阵 $\boldsymbol{P}\in R^{n\times n}$ 的每一行和每一列恰好有一个元素等于 1,而其余的元素都是 0,就称 $\boldsymbol{P}$ 是一个**置换矩阵**.置换矩阵实际上就是对 n 阶单位矩阵的 n 个列向量进行重新排列而组成的矩阵.一个矩阵左(或右)乘以置换矩阵就是将被乘矩阵的行(或列)作相应的置换.例如,取置换矩阵

$$\boldsymbol{P}=\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}\in R^{3\times 3}$$

则

$$\boldsymbol{P}\begin{pmatrix}1\\2\\3\end{pmatrix}=\begin{pmatrix}2\\1\\3\end{pmatrix}$$

$\boldsymbol{P}$ 将向量$(1,2,3)^{\mathrm{T}}$ 的行分量进行互换,即它把第一个元素换到第二个位置,而把第二个元素换到第一个位置,第三项保持不动.一般,矩阵 $\boldsymbol{A}\in R^{m\times n}$ 左乘置换矩阵 $\boldsymbol{P}\in R^{m\times m}$ 就是互换 $\boldsymbol{A}$ 的行,右乘置换矩阵 $\boldsymbol{P}\in R^{n\times n}$ 就是互换 $\boldsymbol{A}$ 的列.

根据行列式展开公式,置换矩阵的行列式是 ± 1,因而置换矩阵一定是非奇异

的.虽然置换矩阵关于矩阵乘法是不可交换的,但两个置换矩阵的乘积还是置换矩阵.因为单位矩阵是一个置换矩阵,且对每个置换矩阵 $\boldsymbol{P}$,有 $\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{P}^{-1}$.所以,置换矩阵集合构成 $R^{n\times n}$中的正交变换群 $O(n)$的一个子群,该子群有 $n!$ 个元素.

如果置换矩阵 $\boldsymbol{P}\in R^{n\times n}$置换矩阵的行,则 $\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{P}^{-1}\in R^{n\times n}$就是置换相同标号的列,所以变换 $\boldsymbol{A}\rightarrow\boldsymbol{PAP}^{\mathrm{T}}$ 置换 $\boldsymbol{A}\in R^{n\times n}$相同标号的行和列,这个变换相当于重排这两行和列的各元素.

方阵 $\boldsymbol{A}\in R^{n\times n}$称为**本性三角矩阵**,如果存在某个置换矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{PAP}^{\mathrm{T}}$ 是三角矩阵.本性三角矩阵与三角矩阵之间有许多共同的性质.

1.3　分块矩阵

在矩阵计算中,采用分块处理可以简化许多核心算法,目前在高性能计算中,“分块算法”越来越重要.分块算法实质上是指大量运用矩阵乘矩阵的算法.在许多计算环境下,这类算法比基于低层线性代数的算法更为有效.通过合理地利用矩阵的结构,可以优化储存空间和减少求解线性方程组的计算量.因此熟悉和掌握矩阵分块计算是很重要的.

块矩阵记号　通常的矩阵按行和列分划可以看做是矩阵分块的特殊情形.在一般情况下,对 $n\times n$ 矩阵 $\boldsymbol{A}$ 的行和列的指标集进行分划,并用 $q\times r$ 个子矩阵把 $\boldsymbol{A}$ 记成

$$\boldsymbol{A}=\begin{array}{cc}\begin{bmatrix}\boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1r}\\ \vdots & & \vdots\\ \boldsymbol{A}_{q1} & \cdots & \boldsymbol{A}_{qr}\end{bmatrix} & \begin{array}{c}m_1\\ \vdots\\ m_q\end{array}\\ \begin{array}{ccc}n_1 & \cdots & n_r\end{array} & \end{array}\tag{1.10}$$

其中 $m_1+\cdots+m_q=n_1+\cdots+n_r=n$,$\boldsymbol{A}_{\alpha\beta}\in R^{m_\alpha\times n_\beta}$,表示在 α行 β列的**块矩阵**,也称为**子矩阵**,称 $\boldsymbol{A}=[\boldsymbol{A}_{\alpha\beta}]$是一个 $q\times r$ 的**分块矩阵**.

分块矩阵的运算　只要满足对应的阶数条件,分块矩阵的运算就和普通矩阵的运算完全一致.例如,若 $n\times n$ 矩阵 $\boldsymbol{B}$ 也有与式(1.10)相同的分划,即

$$\boldsymbol{B}=\begin{bmatrix}\boldsymbol{B}_{11} & \cdots & \boldsymbol{B}_{1r}\\ \vdots & & \vdots\\ \boldsymbol{B}_{q1} & \cdots & \boldsymbol{B}_{qr}\end{bmatrix}$$

则矩阵 $\boldsymbol{C}=\boldsymbol{A}+\boldsymbol{B}$ 也可看做是 $q\times r$ 分块矩阵

$$\boldsymbol{C}=\begin{bmatrix}\boldsymbol{C}_{11} & \cdots & \boldsymbol{C}_{1r}\\ \vdots & & \vdots\\ \boldsymbol{C}_{q1} & \cdots & \boldsymbol{C}_{qr}\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}_{11}+\boldsymbol{B}_{11} & \cdots & \boldsymbol{A}_{1r}+\boldsymbol{B}_{1r}\\ \vdots & & \vdots\\ \boldsymbol{A}_{q1}+\boldsymbol{B}_{q1} & \cdots & \boldsymbol{A}_{qr}+\boldsymbol{B}_{qr}\end{bmatrix}$$

类似地，可以定义分块矩阵的乘法．设 $\boldsymbol{A},\boldsymbol{B}\in R^{n\times n}$ 的分划为

$$\boldsymbol{A}=\begin{array}{c}\left[\begin{array}{ccc}\boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1s}\\ \vdots & & \vdots\\ \boldsymbol{A}_{q1} & \cdots & \boldsymbol{A}_{qs}\end{array}\right]\begin{array}{c}m_1\\ \vdots\\ m_q\end{array}\\ \begin{array}{ccc}p_1 & \cdots & p_s\end{array}\end{array}\quad \text{和}\quad \boldsymbol{B}=\begin{array}{c}\left[\begin{array}{ccc}\boldsymbol{B}_{11} & \cdots & \boldsymbol{B}_{1r}\\ \vdots & & \vdots\\ \boldsymbol{B}_{s1} & \cdots & \boldsymbol{B}_{sr}\end{array}\right]\begin{array}{c}p_1\\ \vdots\\ p_s\end{array}\\ \begin{array}{ccc}n_1 & \cdots & n_r\end{array}\end{array}$$

则 $\boldsymbol{C}=\boldsymbol{AB}$ 可以按如下方式分块计算：

$$\boldsymbol{C}=\begin{array}{c}\left[\begin{array}{ccc}\boldsymbol{C}_{11} & \cdots & \boldsymbol{C}_{1r}\\ \vdots & & \vdots\\ \boldsymbol{C}_{q1} & \cdots & \boldsymbol{C}_{qr}\end{array}\right]\begin{array}{c}m_1\\ \vdots\\ m_q\end{array}\\ \begin{array}{ccc}n_1 & \cdots & n_r\end{array}\end{array}$$

其中 $\boldsymbol{C}_{\alpha\beta}=\sum\limits_{r=1}^{s}\boldsymbol{A}_{\alpha r}\boldsymbol{B}_{r\beta}$，$\alpha=1,\cdots,q$，$\beta=1,\cdots,r$．

分块矩阵的应用　在对高阶方阵进行计算时分块计算无疑会带来便利，并能极大地减少计算的工作量．下面为简单计，只限于考虑在纵向及横向各分划成两块的矩阵运算，反复使用所得结果，很容易得到在纵向及横向分划成三块或更多块情况下的计算公式．

给定 $\boldsymbol{A}\in R^{n\times n}$，将 $\boldsymbol{A}$ 分划成

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}\tag{1.11}$$

其中 $\boldsymbol{A}_{11}$ 是 n_1 阶矩阵，$\boldsymbol{A}_{22}$ 是 n_2 阶矩阵，$n_1+n_2=n$．

如果 $\boldsymbol{A}_{11}$ 可逆，构造拟下三角矩阵

$$\begin{bmatrix}\boldsymbol{I}_{n_1} & \boldsymbol{0}\\ -\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{I}_{n_2}\end{bmatrix}$$

并用它左乘 $\boldsymbol{A}$ 得

$$\begin{bmatrix}\boldsymbol{I}_{n_1} & \boldsymbol{0}\\ -\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{I}_{n_2}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{0} & \boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\end{bmatrix}$$

这相当于对 $\boldsymbol{A}$ 进行 n_1 个倍加的初等变换．由上式得到分块矩阵式(1.11)的拟三角分解(见第 4 章)

$$\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}=\begin{bmatrix}\boldsymbol{I}_{n_1} & \boldsymbol{0}\\ \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{I}_{n_2}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{0} & \boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\end{bmatrix}$$