


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
www.sciencep.com

0241.6/32

2008

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.一北京: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>

源海印刷有限责任公司印刷

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<长虹>)

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,迭代法和移位迭代法,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

前言

《矩阵计算与应用》内容介绍

第 1 章 矩阵的基本知识	(1)
1.1 基本概念	(1)
1.2 特殊矩阵及其性质	(13)
1.3 分块矩阵	(18)
习题 1	(24)
第 2 章 对称矩阵的特征问题	(26)
2.1 特征值问题	(26)
2.2 对称矩阵的变分原理	(31)
2.3 约束特征问题和广义特征问题的变分原理	(37)
习题 2	(43)
第 3 章 向量和矩阵的范数及其应用	(44)
3.1 向量范数	(44)
3.2 矩阵范数	(49)
3.3 范数的应用	(62)
习题 3	(70)
第 4 章 三角分解和满秩分解	(72)
4.1 Gauss 消去法与矩阵的三角分解	(72)
4.2 对称正定矩阵的 Cholesky 分解	(83)
4.3 矩阵的满秩分解	(88)
习题 4	(91)
第 5 章 矩阵的 QR 分解	(92)
5.1 Givens 变换和 Householder 变换	(92)
5.2 矩阵的 QR 分解	(99)
5.3 QR 分解的更新和应用	(108)
习题 5	(116)
第 6 章 奇异值分解	(118)
6.1 奇异值分解	(118)
6.2 奇异值分解的应用	(130)

6.3 奇异值的极性和扰动理论	(138)
习题 6	(141)
第 7 章 广义逆和伪逆	(143)
7.1 矩阵的广义逆	(143)
7.2 矩阵的伪逆	(152)
7.3 伪逆的扰动理论	(163)
习题 7	(169)
第 8 章 特征值与特征向量的求解算法	(171)
8.1 幂法及其推广	(171)
8.2 QR 算法	(180)
8.3 QR 算法的收敛加速方法	(187)
习题 8	(192)
第 9 章 QR 算法执行	(194)
9.1 QR 算法的执行	(194)
9.2 基于 QR 算法特征向量的计算	(201)
9.3 矩阵奇异值分解的计算	(203)
9.4 子空间迭代和同时迭代	(209)
习题 9	(217)
第 10 章 特征值的估计和敏感性分析	(218)
10.1 特征值的估计	(218)
10.2 特征值的敏感性分析	(223)
10.3 特征向量的敏感性分析	(227)
习题 10	(230)
第 11 章 对称矩阵的特征计算方法	(232)
11.1 Jacobi 算法	(232)
11.2 三对角矩阵的特征值求解算法	(234)
11.3 特征向量的逆迭代算法	(241)
习题 11	(246)
第 12 章 线性方程组的迭代求解方法	(248)
12.1 经典迭代法	(248)
12.2 迭代的收敛分析	(253)
12.3 迭代收敛的例子	(259)
习题 12	(264)
第 13 章 共轭梯度法	(266)
13.1 最速下降法	(266)

13.2	共轭梯度法	(273)
13.3	共轭梯度法的收敛分析	(279)
	习题 13	(286)
第 14 章	大规模稀疏矩阵的方程求解和特征问题	(287)
14.1	稀疏线性方程组的求解	(287)
14.2	Arnoldi 算法	(293)
14.3	隐重新开始的 Arnoldi 算法	(299)
	习题 14	(306)
第 15 章	矩阵函数	(307)
15.1	矩阵序列	(307)
15.2	矩阵函数	(313)
15.3	矩阵函数的微积分及其应用	(321)
	习题 15	(326)
第 16 章	Hadamard 积和 Kronecker 积	(328)
16.1	矩阵的 Hadamard 积	(328)
16.2	直积的概念	(331)
16.3	线性矩阵方程的可解性	(336)
	习题 16	(340)
第 17 章	非负矩阵	(341)
17.1	非负矩阵的基本概念	(341)
17.2	正矩阵和非负矩阵	(345)
17.3	不可约非负矩阵和素矩阵	(351)
	习题 17	(358)
	参考文献	(359)
	索引	(362)
	致谢	(366)

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 $\text{diag}(a_1, \dots, a_n)$ 表示对角元素为 a_1, \dots, a_n 的 n 阶对角阵,当 a_i 被方阵 A_i 代替后,表示块对角阵,见 1.2 节.矩阵 A 的秩表示 A 的列(或行)向量中最大线性无关个数,记为 $\text{rank}A$. $R_r^{m \times n}$ 表示 $R^{m \times n}$ 中所有秩为 r 的矩阵集合.

对矩阵 $A \in R^{m \times n}$, A^T 表示矩阵 A 的转置.对复矩阵 $A \in C^{m \times n}$, \bar{A} 和 A^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 $A \in R^{n \times n}$ 称为 n 阶方阵,方阵比方阵具有更多的代数性质. I_n 表示 n 阶单位方阵,它的第 k 列记为 $e_k^{(n)}$,即

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \quad e_k^{(n)} = (0, \dots, 0, 1, 0, \dots, 0)^T \in R^n$$

当维数在上下文中是清楚的,就略去指标 n ,简记为 I 和 e_k .

n 阶方阵 A 的行列式和迹分别记为 $\det A$ 和 $\text{tr}A$.如果 $\det A \neq 0$,则 A 是可逆的,它的逆矩阵用 A^{-1} 表示,即 $A^{-1}A = AA^{-1} = I$.可逆矩阵也称为非奇异的,不可

逆矩阵也称为奇异的.

若 $A \in C^{n \times n}$ 满足 $A^H A = I$, 则称 A 是酉矩阵, n 阶酉矩阵的全体记作 $U(n)$; 如果 $A^H = A$, 则称 A 是 Hermite 矩阵; 如果 $A^H A = A A^H$, 则称 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 $A = -A^H$, 则称 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 A , 如果对任意的非零向量 $x \in C^n$, 都有 $x^H A x > 0$, 则称 A 是正定的; 如果对任意向量 $x \in C^n$, 都有 $x^H A x \geq 0$, 则称 A 是半正定的.

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

$$A = [a_1, \dots, a_n]$$

同样, 如果 $r_k \in R^n$ 表示 A 的第 k 行的 n 维列向量, 则 A 可记为

$$A = \begin{bmatrix} r_1^T \\ \vdots \\ r_m^T \end{bmatrix}$$

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 $L^\perp = \{x \in R^n : x^T y = 0, \forall y \in L\}$. 给定 m 个向量 $a_1, \dots, a_m \in R^n$, a_1, \dots, a_m 所有线性组合构成的集合称为 a_1, \dots, a_m 的扩张子空间, 记为 $\text{span}\{a_1, \dots, a_m\}$.

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

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

根据扩张子空间的定义, $R(A) = \text{span}\{a_1, \dots, a_n\}$, 即 A 的值域是 A 的列的扩张子空间; A 的零空间(或核), 定义为

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

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

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

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 $Ax = b$ 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 $A \in R^{m \times n}$, $x \in R^n$, $b \in R^m$, 计算 $Ax + b$ 需要 $2mn$ 个 flops, 同样, 矩阵 A 秩一更新 $A + xy^T$, 其中 $x \in R^m$, $y \in R^n$ 也需要 $2mn$ 个 flops, 因此, 对 $A \in R^{m \times p}$, $B \in R^{p \times n}$ 和 $C \in R^{m \times n}$, 计算 $AB + C$ 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 $a_k b_{kj} + c_{ij}$, 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 A 称为上三角的, 如果 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 $A, B \in R^{n \times n}$, 乘积 $C = AB$ 也是上三角的, 即对任何 $k < i$ 或 $j < k$, 都有 $a_k b_{kj} = 0$, 因此

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

所以, 计算 c_{ij} ($i \leq j$) 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 $A \in C^{n \times n}$, 如果对于 $\lambda \in C$, 存在非零向量 $x \in C^n$ 使得

$$Ax = \lambda x \quad (1.2)$$

则称 λ 是 A 的特征值, x 是 A 属于 λ 的特征向量, 偶对 (λ, v) 称为 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, A 对 x 的作用等价于用标量乘以 x . 如果 x 是 A 对应 λ 的特征向量, 则 x 的非零数乘也是 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 $\hat{x} = x / \|x\|$.

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

$$\det(\lambda I - A) = 0$$

这是关于 λ 的 n 次多项式, 称 $p(\lambda) = \det(\lambda I - A)$ 为 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 $\det(\lambda I - A) = \det(\lambda I - A^T)$, 因此 $\lambda(A) = \lambda(A^T)$. 于是, 对任意的 $\lambda \in \lambda(A)$, 一定存在非零向量 $y \in C^n$, 使得 $A^T y = \lambda y$, 即 $y^T A = \lambda y^T$, 称 y 为 A 属于 λ 的左特征向量; 相应地, 属于 λ 的特征向量亦称作 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 A 的代数重数就是它的特征多项式根的重数, 即如果 A 有 r 个互不相同的特征值 $\lambda_1, \lambda_2, \dots, \lambda_r$, 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 A 的特征空间, 即 $\lambda_i I - A$ 的零空间, 记为 L_λ , 它是 A 的不变子空间, 即 $AL_\lambda \subseteq L_\lambda$. 由式(1.1), λ_i 的特征空间的维数是

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

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

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

即几何重数总不大于代数重数. 例如, 矩阵 $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ 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 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 A 的所有特征值都是半单的, 则称 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 $A, B \in C^{n \times n}$ 称为相似的, 如果存在 n 阶非奇异方阵 X , 使得 $A = X^{-1}BX$. 与对角矩阵相似的矩阵称为可对角化矩阵.

容易验证, 如果 A 与 B 相似, 则 $\det(\lambda I - A) = \det(\lambda I - B)$. 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 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 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 x_1, \dots, x_k . 由 $Ax_i = \lambda x_i$, 得

$$x_i^H A x_j = \begin{cases} \lambda, & i = j \\ 0, & i \neq j \end{cases}$$

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 V , 则由上式得

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

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

$$\det(zI - B) = \det(zI - \lambda I) \det(zI - D) = (z - \lambda)^k \det(zI - D)$$

因此, 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 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 A 的特征值等价于求