

工科研究生矩阵论课程的实践应用研究

郭 杰, 郭淑妹, 韩松辉

信息工程大学基础部, 河南 郑州
Email: guojie10142008@163.com

收稿日期: 2020年12月27日; 录用日期: 2021年1月28日; 发布日期: 2021年2月4日

摘 要

本文凝练了作者在工科研究生矩阵论教学中关于应用研究的一些实际做法, 首先从方阵的特征值与特征向量的定义出发, 结合专业背景探索了特征值与特征向量在信息检索、机械振动以及图像压缩等方面的应用, 加深学生对特征值与特征向量涵义的理解, 培养学生的创新实践意识, 提升学生学以致用能力。

关键词

矩阵论, 实践思维, 特征值与特征向量, 图像压缩, 机械振动

Practical Application of Matrix Theory for Engineering Postgraduate Students

Jie Guo, Shumei Guo, Songhui Han

Foundation Department of Information Engineering University, Zhengzhou Henan
Email: guojie10142008@163.com

Received: Dec. 27th, 2020; accepted: Jan. 28th, 2021; published: Feb. 4th, 2021

Abstract

This paper summarized some of the author's practical approaches to applied research in the teaching of matrix theory. Starting from the definition of eigenvalue & eigenvector of square matrix and combined with professional background, this paper explores the application of eigenvalue & eigenvector in information retrieval, mechanical vibration and image compression etc. This design deepens students' understanding of the meaning of eigenvalue & eigenvectors, cultivates students' awareness of innovation practice, and enhances students' ability to apply what they have learned.

Keywords

Matrix Theory, Practical Thinking, Eigenvalue and Eigenvector, Image Compression, Mechanical Vibration

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

矩阵论[1][2]是一个重要的数学分支,属于代数学范畴,这门课程在培养学生的抽象思维能力、数学建模能力以及科学计算能力等方面发挥着举足轻重的作用,而这些能力的培养对于提升学生的数学素养至关重要。目前矩阵论的思想方法已经渗透到理工科、经济管理以及军事学等各个领域,尤其是上世纪五六十年代以来,随着计算机科学技术的发展,矩阵论所提供的科学计算方法在科学技术、工程应用以及数学本身都有着广泛而深刻的应用。基于此,所有工科专业,比如网络工程、信息工程、测绘工程以及密码工程等各个专业都把矩阵论课程作为研究生的学位课和必修课开设。矩阵论课程的内容丰富,但是高度抽象,主讲内容包括线性空间与线性变换、矩阵的范数理论、矩阵分析、矩阵分解、矩阵的特征值估计以及矩阵的广义逆等六个章节,通常博士研究生入学考试范围也是这些板块的内容。而随着研究生学制的改革,通常矩阵论课程只有40学时,因此,如何在这么短的学时中讲授矩阵论课程的核心思想和方法,同时又要顾及到学生创新实践能力的培养,是摆在任课教师面前亟需解决的一个问题。

培养学生的创新实践能力是研究生期间的重要目标,一方面通过课堂知识的消化吸收,继而利用所学内容解决实际问题,也即是工程中的问题进行抽象建模,理论分析,然后科学计算,最终完成工程作业,这是一个学以致用过程。另一方面,要结合导师的项目在某个专业领域进行更深层次的创新研究,推动着该学科的发展。但是在很多情况下,矩阵论课程在训练学生的实践思维能力方面稍有欠缺,首先,囿于学时的限制,在规定的学时里差不多把基本内容讲完,没有更多的学时进行实践教学。其次,这门课需要学生对先修课程内容有较好的把握,比如特征值理论、线性空间和线性变换、矩阵运算、多项式理论等,但是这些往往又是学生的薄弱环节,需要占用一些课时进行回顾复习。本科阶段的线性代数[3][4]的学习都是侧重于定理的推导和基本的矩阵计算,较少涉及知识点的应用案例,在概念的引入时,没有恰当的应用背景,只有“野蛮的被定义”,让学生觉得突兀而生涩。即使整个内容学完,大部分学生仅仅停留在行列式的计算、矩阵的初等变换以线性方程组解的判定和解的结构等枯燥无味的理论层面,如果研究生阶段仍然延续这种授课方式,那将不利于创新型、应用型人才培养。并且在学习过程中,学生被动学习,容易厌学这门高度抽象的数学课程。

这个现象引起了广大教育工作者的注意,近年来研究生矩阵论课程的教学改革也在如火如荼的进行着(比如文献[5][6]),这些文献都是基于教学内容、教学方法等方面的改革,大大地调动学生的学习积极性,激发学生的学习热情,达到事半功倍的效果。除此之外,新工科背景下的矩阵论课程,更应该突出其实用性和可操作性,除了引入贴近生活和工程技术的典型案例,还要把一些数值计算软件应用进去,训练学生的科学计算能力。因此在教学过程中,我们秉承实践思维理念,将内容的抽象性与实用性有机结合起来,使学生学完这门课程之后理解课程的内容,又明白这些内容可以用在什么地方,让学生体会到学有所用、用有所成的快乐!近年来,课程组将建模思想、实践思维引入教学,将数值计算软件 matlab

引入到课堂教学中,大大提高了课堂教学效果,在教育教学改革中取得了显著的成效。而学生的学习也经历了由实际问题到建立数学模型,然后构建算法,利用科学计算方法最终解决实际问题的整个过程,由知识的记忆、理解等低阶水平达到分析、评价和创造的高阶层次,为后续专业发展和研究打下良好基础。本文就以矩阵的特征值与特征向量理论为例来阐述矩阵论课程的实践应用研究。

众所周知,方阵的特征值与特征向量是一个重要的数学概念,在数据处理的统计方法、通信工程中的信息检索、图像压缩与恢复、机械振动等多个方面都有广泛的应用,例如,工程技术中的振动问题和稳定性问题,在数值上大都归结为矩阵的特征值与特征向量的问题。

2. 方阵的特征值与特征向量的内涵

矩阵的特征值和特征向量定义高度抽象,在本科阶段对特征值和特征向量的求法已经做了深入的分析,学生对这个概念的理解仅仅停留在计算方面,因此,在矩阵论课程中对于特征值和特征向量的求法进行弱化,关键是理解内涵和探讨应用。

定义[3] 设 A 是 n 阶方阵,若存在数 λ 和 n 维非零向量 x ,使得 $Ax = \lambda x$ 成立,则称数 λ 是方阵 A 的特征值,非零向量 x 为方阵 A 的属于特征值 λ 的特征向量。

谱分解定理[1] 设 $A = P\Lambda P^{-1}$, 矩阵 P 的列是 A 的单位正交特征向量 u_1, u_2, \dots, u_n , 相应的特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 组成对角阵 Λ , 且 $P^{-1} = P^T$, 那么

$$A = P\Lambda P^T = (u_1, u_2, \dots, u_n) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} u_1^T \\ \vdots \\ u_n^T \end{pmatrix} = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \dots + \lambda_n u_n u_n^T. \quad (1)$$

由上述定义可知,方阵的特征向量是经过矩阵变换后,保持方向不变,只是进行长度扩大或者缩小的向量,而特征值反映了特征向量在矩阵变换时的扩大或者缩小的倍数。结合谱分解定理可得,一个方阵完全可以由它的特征向量表示,特征值即是方阵在对应特征方向上的贡献率大小,即一个方阵可由特征值与特征向量组成的“特征”来表示,特征向量的几何直观如图1所示。

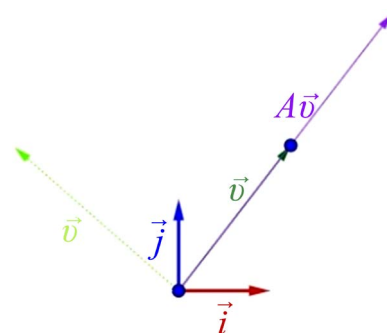


Figure 1. The effect of matrix on eigenvector
图 1. 矩阵对特征向量的作用

3. 特征值与特征向量在专业背景中的应用

3.1. 特征值与特征向量在 PageRank 算法中的应用

特征值理论在网络社区发现和数据挖掘中的应用,最经典的案例就是谷歌搜索的网页排序算法,也就是 PageRank 算法,谷歌用它来体现网页的相关性和重要性,这一算法构成了谷歌搜索引擎的核心技术,方阵的特征值和特征向量理论为其提供了理论支撑。该算法的简单模型如图2所示。

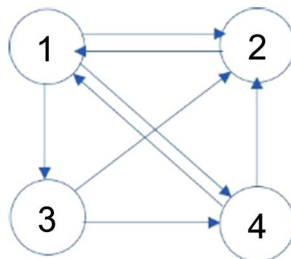


Figure 2. Network link model

图 2. 网络链接模型

设 n_j 表示第 j 个网页出去的链接数, 用 L_k 表示进入第 k 个网页的网页集合, 则重新定义网页等级为 $x_k = \sum_{j \in L_k} \frac{x_j}{n_j}$, 由图 2 不难建立关系式如下

$$\begin{cases} x_1 = \frac{1}{3}x_3 + \frac{1}{2}x_4 \\ x_2 = \frac{1}{3}x_1 \\ x_3 = \frac{1}{3}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_4 \\ x_4 = \frac{1}{3}x_1 + \frac{1}{2}x_2 \end{cases} \Leftrightarrow X_{k+1} = AX_k, A = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix} \quad (2)$$

(2)式建立了马尔科夫链的一阶差分方程, 由盖尔圆定理[1]知其概率转移矩阵的最大特征值为 1, 所对应的特征向量为 x , 由文献[7]知道该马尔科夫链的稳态向量就是最大特征值所属的特征向量。依据稳态向量的各个分量的大小可以给出网页的排序, 在课堂讲授中给学生大概介绍一下理想化的模型, 有兴趣的同学可以课外查找文献, 仔细研究具体算法, 激发学习兴趣。

3.2. 特征值与特征向量在机械振动中的应用

在讲解特征值和特征向量的概念时, 一定要结合它的物理意义进行阐述, 一方面加深对抽象概念的理解, 另一方面, 便于解释在工程技术中的实现。如果把矩阵看作是对运动的描述, 而矩阵的特征值就是运动的速度, 所对应的特征向量就是运动的方向, 所以特征值和特征向量就构成了运动的“特征”。因此有人说, 有振动的地方就有特征值和特征向量, 这话有一定的理论依据。比如音乐厅的结构设计或者桥梁的构造, 建筑学家都会结合特征值进行详细推导计算, 在合唱或者器乐齐鸣时达到更好的回响效果。例如历史上著名的塔科马悬索桥风毁事件, 法国和俄国都曾出现过部队齐步过桥时, 桥毁人亡的事件, 究其原因是桥梁的机械频率和运动的振动频率基本一致导致的共振造成的。再比如 2020 年 5 月发生的虎门大桥晃动事件, 就是由于增加了水马, 改变了桥的重量, 从而导致共振, 引起桥身晃动, 拿走了水马, 桥面就逐渐不再晃动。这些与振动相关的事件都可以通过特征值估计理论[4]予以详细解释。

为简单起见, 假设矩阵 A 可对角化, 特征向量 p_1, p_2, \dots, p_n 是 R^n 的基, 并且已经排列, 使得对应的特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 的绝对值递减, 具有主特征值 λ_1 。假设 $x \in R^n$, 且 $x = c_1 p_1 + c_2 p_2 + \dots + c_n p_n$ ($c_1 \neq 0$), 则

$$A^k x = c_1 \lambda_1^k p_1 + c_2 \lambda_2^k p_2 + \dots + c_n \lambda_n^k p_n \quad (k=1, 2, \dots)$$

等式两边同除以 λ_1^k 得,

$$\frac{1}{\lambda_1^k} A^k x = c_1 p_1 + c_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k p_2 + \dots + c_n \left(\frac{\lambda_n}{\lambda_1}\right)^k p_n \quad (k=1, 2, \dots) \quad (3)$$

由于(3)式右端的所有分数 $\frac{\lambda_2}{\lambda_1}, \dots, \frac{\lambda_n}{\lambda_1}$ 的值都小于 1, 因此它们的幂趋于零, 即当 $k \rightarrow \infty$ 时,

$\frac{1}{\lambda_1^k} A^k x \rightarrow c_1 p_1$ 。因此对于足够大的 k , $A^k x$ 的倍数的方向几乎与特征向量 $c_1 p_1$ 的方向相同。由于正的倍数不会改变向量的方向, 若给定 $c_1 \neq 0$, 则 $A^k x$ 的方向几乎与 p_1 或 $-p_1$ 一致。可以通过二阶方阵让学生借助 matlab 快速计算向量的迭代, 直观演示如图 3 所示。一座桥梁可以视为一个物理系统, 其各种物理特性如质量、刚性、阻尼等可以用一个矩阵来描述。部队在桥梁上行走时, 人对地面的作用可以视为一个种机械振动, 桥梁的各种物理特性构成矩阵的特征值决定着行走中振动的频率和幅度减弱的衰退率。当人马齐步通过桥梁时, 每个人对桥梁的作用近似相同, 振动相互叠加, 如果出现叠加信号与特征向量相同, 而主特征值又大于 1, 则矩阵在特征向量所指的方向上对向量产生一个持续的增强作用, 从而可能形成强烈振动损坏桥梁, 导致桥毁人亡。如果部队分散通过桥梁, 那么矩阵对各振动信号的作用无法与矩阵的特征向量一致形成“共振”, 从而避免悲剧的发生。

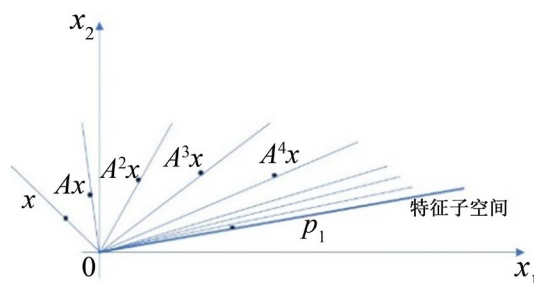


Figure 3. Direction determined by x, Ax, A^2x, \dots, A^7x

图 3. 由 x, Ax, A^2x, \dots, A^7x 确定的方向

3.3. 特征值与特征向量在图像压缩中的应用

矩阵的各种分解形式为矩阵的科学计算提供了强有力的理论支撑, 通过矩阵分解可以达到对矩阵进行降维的目的, 从而减小内存量, 简化运算。为了说明特征值与特征向量在图像压缩技术中的应用, 先来介绍一下奇异值分解的概念。

定义 2 [1] 设 $A \in R_r^{m \times n}$, $A^T A$ 的特征值为 $\lambda_i (i=1, 2, \dots, n)$, 且设 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \dots = \lambda_n = 0$, 称 $\sigma_i = \sqrt{\lambda_i} (i=1, 2, \dots, r)$ 为矩阵 A 的奇异值。如果存在 m 阶正交矩阵 U 和 n 阶正交矩阵 V , 满足 $A = UDV^T$, 其中 $m \times n$ 矩阵 D 的分块形式为

$$D = \begin{pmatrix} \Sigma & o \\ o & o \end{pmatrix}, \text{ 且 } \Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}, \sigma_i (i=1, 2, \dots, r) \text{ 为矩阵 } A \text{ 的奇异值, 那么矩阵的积分解式}$$

$A = UDV^T$ 称为 A 的奇异值分解。事实上, A 的奇异值是其法矩阵 $A^T A$ 的正特征值的平方根。若记 $U = (u_1, u_2, \dots, u_m), V = (v_1, v_2, \dots, v_n)$, 则矩阵 A 可以分解为如下秩 1 矩阵的和:

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T. \quad (4)$$

(4)式和谱分解定理中的(1)式形式相近, 但是要注意两者的差异。

假定一幅图像有 $m \times n$ 个像素, 如果将这 mn 个数据一起传送, 往往数据量会很大。因此, 我们考虑在信息的发送端传送比较少的数据, 并且在接收端利用这些传送数据对图像进行重构。这就是图像压缩的最初想法, 不过图像压缩要求较高的压缩比, 同时不产生失真。矩阵的奇异值分解可以将任意一个矩

阵和一个只包含几个(非零)奇异值的矩阵对应。把“大”的矩阵对应到“小”的矩阵,这就产生了“压缩”的思想,并且利用矩阵的计算可以恢复压缩前的数据[8]。矩阵 A 表示一个 536×498 的矩阵,需要保存 $536 \times 498 = 266928$ 个元素的值,(4)式中的 u, v 分别为 $536 \times 1, 498 \times 1$ 的列向量,每一项有 $1 + 536 + 498 = 1035$ 个元素,如果我们只保留奇异值分解的前 50 项,则需要存储的元素为 $1035 \times 50 = 51750$,和存储原始矩阵 A 相比,存储量仅为后者的 19.39%,大大节省了存储空间。一般地,如果从中选择 k 个较大的奇异值,则称比率 $\rho = \frac{mn}{(m+n+1)k}$ 为图像的压缩比,通过公式 $A_k = \sum_{i=1}^k \sigma_i u_i v_i^T$ 重构出原图像就可以满足人的视觉要求。

容易理解:若数 k 越小,即压缩比越大,则重构的图像质量不高;反之, k 太大,则压缩比太小,降低了图像传输效率,因此,实际应用中需要兼顾传输效率和重构图像质量的同时选择合适的压缩比。图 4(a) 是一张 536×498 像素的原始黑白图像,通过 matlab 命令读取图像数据,然后进行奇异值分解,336 个非零奇异值的和为 785.2964。表 1 清晰直观的给出了压缩后的奇异值占比和压缩比,从图表不难看出,图 4(b) 中取 $k=1$,截取最大奇异值,重构图像模糊不清,看不出是什么图形。图 4(c) 中取 $k=5$,截取前 5 个奇异值,此时可以看出图像轮廓。图 4(d) 中取 $k=20$,此时图像已经可以辨认,但是还有点模糊。图 4(e) 中取 $k=50$,截取前 50 个奇异值,此时图像已经非常接近于原图,但是存储空间大大降低,不及原图像 20%。在图像处理领域,奇异值分解不仅可以用于图像压缩上,还可以对图像进行降噪处理。

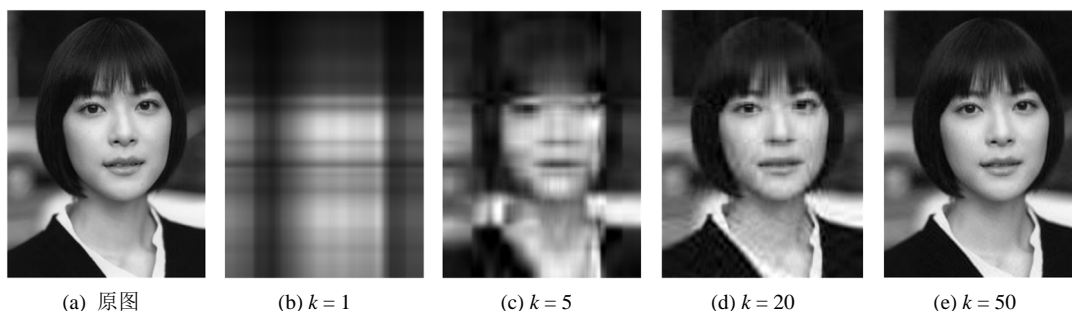


Figure 4. Original image and compressed image
图 4. 原图以及压缩后的图像

Table 1. Compressed data table of image
表 1. 图像的压缩数据表

项目	$k=1$	$k=5$	$k=20$	$k=50$
奇异值占比	46.03%	72.27%	87.64%	94.49%
压缩比	257.9	51.58	12.9	5.16

在教学过程中,这个案例作为大作业来执行的,学生通过实际动手操作,深入理解了矩阵的奇异值分解理论,熟练掌握了的 matlab 常用命令以及算法实现,极大的提高了学生对这门课程的学习兴趣。

4. 结语

教学案例的设计要结合专业背景,引导学生在探索专业问题解决方案的过程中逐步建构课程的知识体系,加深学生对课程内容的本质理解,促使他们在潜移默化间形成初步的科学研究意识[9]。在课程教学活动中,不一定将详细的操作过程演示给学生,可以借助数学软件分析具有强烈应用背景的案例,激发学生对课程的学习兴趣,打破矩阵论的“高冷”形象,这样“接地气”的教学设计能推动学生积极探索,培养学生分析和解决问题的能力。

基金项目

国家自然科学基金(41474009); 信息工程大学教育教学研究项目(JXYJ2019C006, JXYJ2020D012)。

参考文献

- [1] 程云鹏. 矩阵论[M]. 西安: 西北工业大学出版社, 2017.
- [2] 张贤达, 周杰. 矩阵论及其工程应用[M]. 北京: 清华大学出版社, 2015.
- [3] 同济大学数学系. 线性代数(第五版) [M]. 北京: 高等教育出版社, 2007.
- [4] David C. Lay, 著. 线性代数及其应用[M]. 刘深泉, 洪毅, 等, 译. 北京: 机械工业出版社, 2005.
- [5] 刘碧玉, 刘庆平, 唐先华. 工科研究生矩阵论课程教学改革的探索与实践[J]. 数学理论与应用, 2013, 33(1): 125-128.
- [6] 赵礼峰. 工科研究生矩阵论课程教学改革研究与实践[J]. 大学数学, 2013, 29(4): 1-3.
- [7] 李裕奇. 随机过程[M]. 北京: 国防工业出版社, 2003.
- [8] 邱森, 竹林生. 高等代数探究性课题精编[M]. 武汉: 武汉大学出版社, 2012.
- [9] 黄影, 张丽华. 基于“新工科”的线性代数案例式教学模式的研究与实践[J]. 沈阳师范大学学报(自然科学版), 2019, 37(5): 467-470.