DOI:10.19789/j.1004-9398.2018.05.002

线性代数课程中的特征值与特征向量教学研究*

刘智慧 付丽华 李宏伟 李超群
(中国地质大学数理学院,湖北 武汉 430074)

摘 要

针对特征值与特征向量的抽象性,本文从特征值和特征向量的概念及特征值分解定理出发,通过几何直观演示、出租车的调配及高维数据的降维两个具体实例,并结合 MATLAB 软件阐明特征值与特征向量的几何直观性和实际应用,以期学生多角度深入理解特征值与特征向量,激发学生的学习兴趣,培养学生运用数学解决实际问题的能力.

关键词: 线性代数,特征值和特征向量,几何直观,出租车调配,数据降维.
中图分类号: O151

特征值与特征向量是《线性代数》中的重点内容之一,在众多领域有着广泛的应用[1-8]. "特征"一词来自德语的 eigen,翻译为"自身的""有特征的""特定于……的",强调了特征值与特征向量对特定矩阵的重要性. 由于特征值与特征向量涉及的概念、定理较为抽象,在一定程度上阻碍了学生对其理解与认识. 如果在特征值与特征向量教学中,融入几何直观并通过实际生活中的具体例子阐明特征值与特征向量的作用,将在一定程度上调动学生的学习积极性,使其更好地掌握特征值与特征向量. 本文借助 MATLAB 软件从几何直观以及出租车调配和高维数据降维的具体实例说明特征值与特征向量,注重应用特征值与特征向量解决实际中的相关问题,加强学生对特征值与特征向量的深入理解,领会特征值与特征向量的数学意义和应用价值.

1 矩阵的特征值与特征向量

定义 1[7,9] 设 A 为 n 阶矩阵,x 为非零向量,若存在数 λ,使得

$$Ax = \lambda x$$

成立,则称 λ 为 A 的特征值,x 为 A 的属于 λ 的一个特征向量.

从定义 1 可知,特征向量为一些特定的向量,这些向量在方阵 A 的作用下保持方向不变,只进行长度上的伸缩(当 $\lambda < 0$ 时,伸缩值为负值;当 $\lambda = 0$ 时,伸缩值为零). 若 x 为 A 的特征向量,则 $kx(k \neq 0)$ 也是 A 的特征向量,所以特征向量不是一个向量而是一个向量族.

定理 1 特征值分解定理[7,9] 设 A 为 n 阶对称矩阵,则必存在一个正交矩阵 $P = [p_1, p_2, \cdots, p_n]$,使得

$$A = P\Lambda P^{\mathrm{T}}$$

其中,$(\cdot)^{\mathrm{T}}$ 表示转置运算,$\Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$,$\lambda_i$ 为 A 的特征值,p_i 为 A 的属于 λ_i 的特征向量.

从定理 1 中可知,任意一个实对称矩阵总可以用它的特征值和特征向量完全描述.

通过一个具体的例子演示矩阵的特征值和特征向量.

例 1 求矩阵 $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ 的特征值和特征向量,绘制特征向量图,并分析其几何意义.

利用 MATLAB 命令求出矩阵的特征值和特征向量,相应的 MATLAB 程序为:

```
A = [1 0;0 2];
[P,D] = eig(A); % 求矩阵 A 的特征值和特征向量
eigshow(A); % 演示矩阵 A 的特征值和特征向量
```

收稿日期: 2017-06-20
* 中国地质大学(武汉)本科教学工程(ZL201739,ZL201801);中国地质大学(武汉)教学研究项目(2016A32).

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

例 1 的 MATLAB 程序运行结果如下:

```
P =
     1     0
     0     1
D =
     1     0
     0     2
```

图 1 描述了蓝色向量 Ax 随绿色向量 x 的变换关系,x 表示二维单位向量,当用鼠标拖动 x 围绕原点旋转时,Ax 随着 x 旋转,向量 x 的轨迹是一个绿色单位圆,向量 Ax 的轨迹为一个蓝色椭圆. 在图 1 中,当 x 旋转到水平方向和竖直方向时(见图 1(c)(d)),Ax 恰好与 x 重合,即 $Ax = \lambda x$,由定义 1 知,矩阵 A 的两个特征值为椭圆的短轴和长轴,对应的特征向量分别为 $k_1(1,0)$,$k_2(0,1)$,其中 $k_1 \neq 0$,$k_2 \neq 0$. 例 1 说明,对所有二维向量,只有形如 $k_1(1,0)$ 和 $k_2(0,1)$ 的二维向量在二阶对角矩阵 A 的作用下保持方向不变.

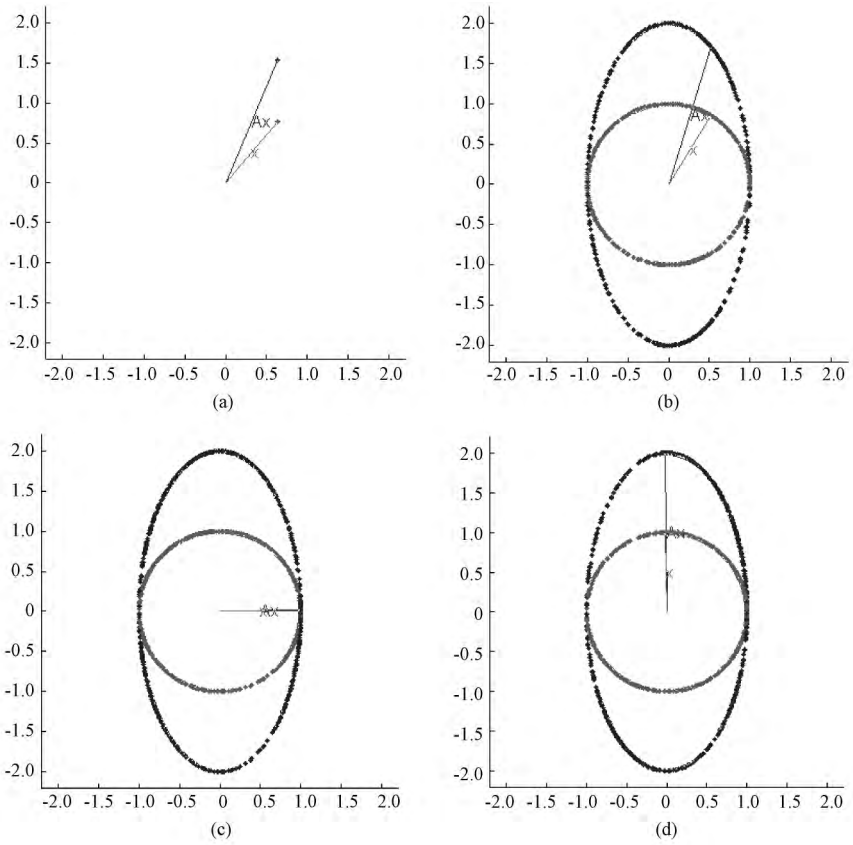


图 1　矩阵特征值和特征向量的图形演示

2　特征值与特征向量的应用

2.1　出租车的调配问题

假设一家出租车公司在甲地和乙地各有一家分支机构,专门负责为旅游公司提供出租车. 由于甲地和乙地距离不远,出租车每天可以往返两地. 根据公司统计的历史数据,每天甲地的车辆有 60% 前往乙地后返回甲地,余下 40% 前往乙地并留在乙地分支机构; 而每天乙地的车辆有 70% 前往甲地后返回乙地,余下 30% 前往甲地并留在甲地分支机构. 假设出租车公司出租车数量是固定不变的,问: 当经过足够长时间后,甲地和乙地出租车数量是否会越来越不平衡? 如果甲乙两地车辆数量不平衡,公司如何调配?

解: 设向量 $x_n = \begin{bmatrix} J_n \\ Y_n \end{bmatrix}$ 表示第 n 天甲地和乙地出租车的数量,其中 J_n 表示甲地的出租车数量,Y_n 表示乙地的出租车数量,则第 $n+1$ 天甲地和乙地出租车的数量为

$$\begin{cases} J_{n+1} = 0.6J_n + 0.3Y_n, \\ Y_{n+1} = 0.4J_n + 0.7Y_n. \end{cases} \tag{1}$$

式(1)用矩阵表示为

$$x_{n+1} = \begin{bmatrix} J_{n+1} \\ Y_{n+1} \end{bmatrix} = \begin{bmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{bmatrix} \begin{bmatrix} J_n \\ Y_n \end{bmatrix} \triangleq Ax_n. \tag{2}$$

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

由式(2)可知,每天出租车的调配情况由矩阵 $A=\begin{bmatrix}0.6 & 0.3\\ 0.4 & 0.7\end{bmatrix}$ 来决定. 进一步,由式(2)可得,$x_n = Ax_{n-1} = A^2x_{n-2} = \cdots = A^nx_0$.

利用 MATLAB 命令求出矩阵 A 的特征值为 $\lambda_1=1$,$\lambda_2=0.3$,对应的特征向量为 $p_1=[-0.600\,0,-0.800\,0]^{\mathrm{T}}$,$p_2=[-0.707\,1,0.707\,1]^{\mathrm{T}}$.

显然特征向量 p_1 和 p_2 是 $\mathbf{R}^2$ 的基,故初始向量 $x_0=\begin{bmatrix}J_0\\ Y_0\end{bmatrix}$ 可以惟一表示为:$x_0=c_1p_1+c_2p_2$.

又 p_i 为特征向量,故 $x_n=A^nx_0=c_1(\lambda_1)^np_1+c_2(\lambda_2)^np_2$. 若 $n\to\infty$,则 $(\lambda_2)^n=(0.3)^n\to 0$. 于是,当 $c_1>0$ 时,

$$x_n\approx c_1(\lambda_1)^np_1, \tag{3}$$

以及

$$x_{n+1}\approx c_1(\lambda_1)^{n+1}p_1=\lambda_1c_1(\lambda_1)^np_1\approx\lambda_1x_n. \tag{4}$$

式(3)表明,x_n 的 2 个分量之比(甲地和乙地出租车的数量之比)近似等于 p_1 对应分量之比,即甲乙两地出租车的数量之比只与最大特征值 $\lambda_1=1$ 对应的矩阵特征向量 p_1 有关,而与甲乙两地初始出租车的数量 x_0 无关. 由于 p_1 分量之比为一个常数,所以当经过足够长时间后,甲地和乙地出租车数量之比趋于一个稳定数,即甲乙两地车辆数量趋于平衡. 式(4)表明,x_n 最终以近似 λ_1 的倍数增长,故矩阵 A 的最大特征值 λ_1 确定了最终增长率.

图 2 表示出租车公司总计 7 000 辆出租车,甲地和乙地出租车数量之比随着时间的变化规律,其中图 2(a)表示甲乙两地初始车辆数分别为 2 000 辆和 5 000 辆所对应的图像,图 2(b)表示甲乙两地初始车辆数分别为 7 000 辆和 0 辆所对应的图像. 从图 2 中可知,当出租车公司车辆数固定不变时,无论甲乙两地初始出租车数量为多少,随着时间的增加,甲地和乙地车数量之比趋于一个稳定数 $\frac{3}{4}$,即矩阵 A 的最大特征值 $\lambda_1=1$ 对应的特征向量分量之比 $\frac{-0.600\,0}{-0.800\,0}$.

以上例子说明,利用矩阵的特征值和特征向量,选择矩阵较大特征值及对应的特征向量可以分析出租车的调配问题.

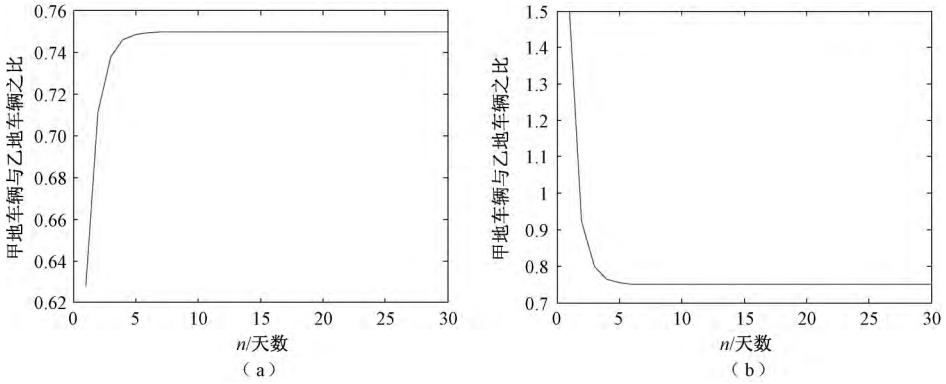


图 2 甲地和乙地出租车数量之比

矩阵 A 的特征值和特征向量及图 2(a)相应的 MATLAB 程序为:

```
A = [0.6 0.3;0.4 0.7];
x0 = [2 000;5 000];
[P,D] = eig(A);
for i = 1:30
x(:,i) = A^i* x0;
end
plot(x(1,:)./x(2,:));
xlabel('n/天数');
ylabel('甲地车辆与乙地车辆之比');
```

2.2 高维数据的降维问题

现有 m 组数据,每一组数据为 n 维向量,在机器学习中,若对这 n 维 m 组数据点直接进行分析和挖掘,那么算法的复杂度往往与维数 n 呈指数级关系. 能否将 n 维向量组降为 k 维向量组($k<n$),同时又能保留原始的数据点信息?回答是肯定的.

将此数据看作 $n\times m$ 的矩阵 $X_{n\times m}$,欲使 n 维向量组降为 k 维向量组,即寻找矩阵 $P_{k\times n}$,使得 $P_{k\times n}X_{n\times m}=Y_{k\times m}$,则矩阵 $Y_{k\times m}$ 为降维的 k 维向量组.

下面用简单的例子说明基于矩阵特征值和特征向量的高维数据降维问题. 图 3(a)表示五个不

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

同的二维数据点,若将此数据点向 x 轴投影,则右边的两个点会重叠在一起,中间的两个点也会重叠在一起,于是五个不同的二维数据点投影后只剩下三个不同的投影值,这是一种严重的信息丢失. 同理,若将此数据点向 y 轴投影,则投影后也只有三个不同的投影值. 因此 x 轴和 y 轴不是数据点最好的投影方向. 但若将此数据点向通过第一象限和第三象限的斜线投影(见图 3(b)),则五个二维数据点投影后为五个不同的投影值,即数据点向该斜线投影后,二维数据点降为了一维数据点,同时保留了五个原始数据信息. 此例说明,将高维数据进行降维,只需将数据点向某些方向进行投影,使得在这些方向上的投影值尽可能分散,问: 如何找到这些投影方向?

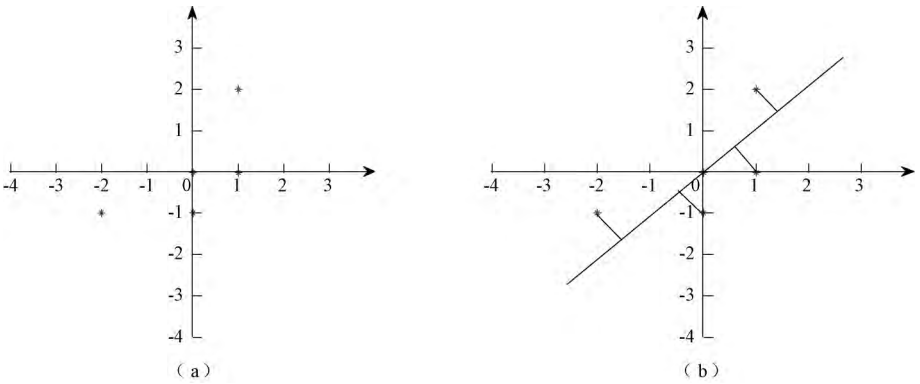


图 3　二维数据点分布情况

为讨论的方便,二维数据点表示为 $X=\begin{bmatrix}a_1 & a_2 & a_3 & a_4 & a_5\\ b_1 & b_2 & b_3 & b_4 & b_5\end{bmatrix}$,其中 $(a_i,b_i)^{\mathrm{T}}$ 表示第 i 个数据点,将 X 的每一行进行零均值化处理,得 $\overline{X}=\begin{bmatrix}a_1-u & a_2-u & a_3-u & a_4-u & a_5-u\\ b_1-v & b_2-v & b_3-v & b_4-v & b_5-v\end{bmatrix}$,其中 $u=\frac{1}{5}\sum_{i=1}^{5}a_i$, $v=\frac{1}{5}\sum_{i=1}^{5}b_i$. 计算 X 的协方差矩阵

$$A=\frac{1}{5}\overline{X}\,\overline{X}^{\mathrm{T}}=\begin{bmatrix}\frac{1}{5}\sum_{i=1}^{5}(a_i-u)^2 & \frac{1}{5}\sum_{i=1}^{5}(a_i-u)(b_i-v)\\ \frac{1}{5}\sum_{i=1}^{5}(a_i-u)(b_i-v) & \frac{1}{5}\sum_{i=1}^{5}(b_i-v)^2\end{bmatrix}.$$

显然, A 为对角矩阵, A 对角元表示二维向量的方差(刻画分散程度); A 对角元外的元素表示维度间的协方差(刻画维度间的不相关性). 对五个数据点进行降维,同时尽可能保留五个原始数据点,即满足五个数据点投影后投影值尽可能分散(二维向量的方差尽可能大),同时维度间尽可能不存在相关性(维度间的协方差尽可能小),即使得协方差矩阵 A 对角化(对角元的元素从大到小排列,对角元外的元素为零).

设 $Y=P\overline{X}$, Y 的协方差矩阵为 B,则

$$B=\frac{1}{5}YY^{\mathrm{T}}=\frac{1}{5}(P\overline{X})(P\overline{X})^{\mathrm{T}}=P\left(\frac{1}{5}\overline{X}\,\overline{X}^{\mathrm{T}}\right)P^{\mathrm{T}}=PAP^{\mathrm{T}}. \quad (5)$$

由定理 1 和式(5)知,对协方差矩阵 A,必有正交矩阵 P,使得 A 对角化,其中对角矩阵 B 的对角元为矩阵 A 的特征值, P 由特征值对应的特征向量构成. 由式(5),有 $AP^{\mathrm{T}}=P^{\mathrm{T}}B$,即 P 的行向量为 A 的特征向量. 于是,将二维数据点降为一维数据点,即找出 A 的最大特征值对应的特征向量 p_1 作为 P 的第一行,使得 $p_1^{\mathrm{T}}X=y$,其中 y 为一维的五个数据点. 由于 p_1 为正交矩阵 P 的第一行,故 p_1 为单位向量,于是 $y=p_1^{\mathrm{T}}X$ 可理解为五个数据点在最大特征值对应的特征向量 p_1 上的五个投影值. 具体求解步骤如下:

将五个二维数据点表示为 $X=\begin{bmatrix}-2 & 0 & 0 & 1 & 1\\ -1 & -1 & 0 & 0 & 2\end{bmatrix}$,由于 X 的每一行均值为 0,则 X 的协方差矩阵为 $A=\frac{1}{5}\begin{bmatrix}-2 & 0 & 0 & 1 & 1\\ -1 & -1 & 0 & 0 & 2\end{bmatrix}\begin{bmatrix}-2 & 0 & 0 & 1 & 1\\ -1 & -1 & 0 & 0 & 2\end{bmatrix}^{\mathrm{T}}=\begin{bmatrix}\frac{6}{5} & \frac{4}{5}\\ \frac{4}{5} & \frac{6}{5}\end{bmatrix}$. 利用 MATLAB 命令求出矩阵 A 的特征值为 $\lambda_1=2$, $\lambda_2=0.4$,对应的特征向量为 $p_1=$

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

[0.707 1,0.707 1]^{T},p_2 = [-0.707 1,0.707 1]^{T}. 利用矩阵A最大特征值$\lambda_1 = 2$对应的特征向量p_1作为行向量乘以数据矩阵X,得投影值

$$y = p_1^{T}X = [0.707\ 1, 0.707\ 1]\begin{bmatrix} -2 & 0 & 0 & 1 & 1 \\ -1 & -1 & 0 & 0 & 2 \end{bmatrix}$$

= [-2.121 3, -0.707 1,0,0.707 1,2.121 3].

其中,y为降维后的五个一维数据点,p_1为二维数据点的投影方向(见图3(b)).

二维数据点降为一维数据点的相应 MATLAB 程序为:

```
X = [-2 0 0 1 1; -1 -1 0 0 2];
A = (1/5) * X* X'; % 数据点的协方差矩阵 A
[P,D] = eig(A);
y = P(:,1)'* X; % 降维后的一维数据点
```

对n维m组数据点$X_{n\times m}$,若将其降维为k维m组数据点$Y_{k\times m}$($k < n$),则只需利用协方差矩阵$A = \frac{1}{m}\overline{X}_{n\times m}\overline{X}_{m\times n}^{T}$前面$k$个较大特征值对应的特征向量$p_1, p_2, \cdots, p_k$,使得$P_{k\times n}X_{n\times m} = Y_{k\times m}$,其中$P_{k\times n} = [p_1, p_2, \cdots, p_k]^{T}$.

以上例子说明,矩阵较大特征值对应的特征向量上包含了较多的信息量,利用矩阵前面部分较大特征值对应的特征向量可实现高维数据的降维.

3 小 结

本文对《线性代数》课程中特征值和特征向量两个抽象的概念运用 MATLAB 软件,从几何直观演示、出租车的调配及高维数据降维等实际问题进行了阐述,使得学生能更好地理解和认识特征值与特征向量,同时,对学生的学习兴趣及学生运用数学知识解决实际问题的能力的培养有一定的帮助.

参 考 文 献

[1] Lay D C. 线性代数及其应用[M]. 刘深泉,洪毅,马东魁,等译. 北京: 机械工业出版社,2005.
[2] 林蔚,周双红,国萃,等. 线性代数的工程案例[M]. 哈尔滨: 哈尔滨工程大学出版社,2012.
[3] 贤锋. 最大特征值及其特征向量的应用[J]. 闽江学院学报:自然科学版,2006,27(5): 35-41.
[4] 王海侠,孙和军. 几何直观在特征值问题中的应用[J]. 高等数学研究,2014,17(1): 105-108.
[5] 阮春蕾. 特征值与特征向量的两例应用[J]. 河南科技大学学报:自然科学版,2015,36(6): 87-91.
[6] 朱玲. 线性代数中的特征值和特征向量的教学应用案例[J]. 兰州教育学院学报,2016,32(12): 86-90.
[7] 王亮,冯国臣,王兵团. 基于 MATLAB 的线性代数实用教程[M]. 北京: 科学出版社,2008.
[8] 陈怀琛,龚杰民. 线性代数实践及 MATLAB 入门[M]. 北京: 电子工业出版社,2008.
[9] 同济大学应用数学系. 线性代数(第六版)[M]. 北京: 高等教育出版社,2014.

Teaching of Eigenvalue Problems in Linear Algebra

Liu Zhihui Fu Lihua Li Hongwei Li Chaoqun

(School of Mathematics and Physics, China University of Geosciences, Wuhan Hubei 430074)

Abstract

Considering the abstract of the eigenvalue and eigenvector, in this paper, first the concept of the eigenvalue and eigenvector and eigenvalue decomposition theorem are presented, then the geometric visualization, the application of the eigenvalue and eigenvector in taxi allocation and data dimension reduction are addressed by MATLAB mathematical software. This paper emphasizes on applying the eigenvalue and eigenvector to solve the practical problems which deepen the understanding of the students about the eigenvalue and eigenvector.

Key words: linear algebra, eigenvalue and eigenvector, geometric visualization, taxi allocation, data dimension reduction.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net