特征值与特征向量

Dezeming Family

2021 年 5 月 16 日

DezemingFamily 系列书和小册子因为是电子书，所以可以很方便地进行修改和重新发布。如果您获得了 DezemingFamily 的系列书，可以从我们的网站 [https://dezeming.top/] 找到最新版。对书的内容建议和出现的错误欢迎在网站留言。

目录

一 矩阵与线性变换

我们首先需要明确的是，一个 n 维空间里可以由 n 个不相关的 n 维向量张成。例如三维空间中，$(2,3,4)$ 和 $(3,2,7)$ 和 $(1,8,2)$ 经过线性组合就可以构成整个空间。

但是，$(2,3,4)$ 和 $(3,2,7)$ 和 $(5,5,11)$ 却没法构成整个三维空间，因为向量 $(5,5,11)$ 与向量 $(2,3,4),(3,2,7)$ 线性组合构成的平面是平行的，而任何与该平面不平行的向量它们是无法线性组合得到的。

向量的线性变换

其实线性变换可以理解为（或者从线性函数来说，就是）将基进行了变换。我们假设对于二维空间中的基为 $\widehat{i}$ 和 $\widehat{j}$，向量 $\widehat{w}=(a,b)$，可以表示为 $a\widehat{i}+b\widehat{j}$。我们对向量进行线性变换，即乘以一个矩阵 A，但本质上是作用于了基。我们设 $a=2.5$，$b=1.5$，$\widehat{i}=(1,0)$，$\widehat{j}=(0,1)$，矩阵 A 表示为：

$$A=\begin{bmatrix}3 & -2\\ 1 & 2\end{bmatrix} \tag{一.1}$$

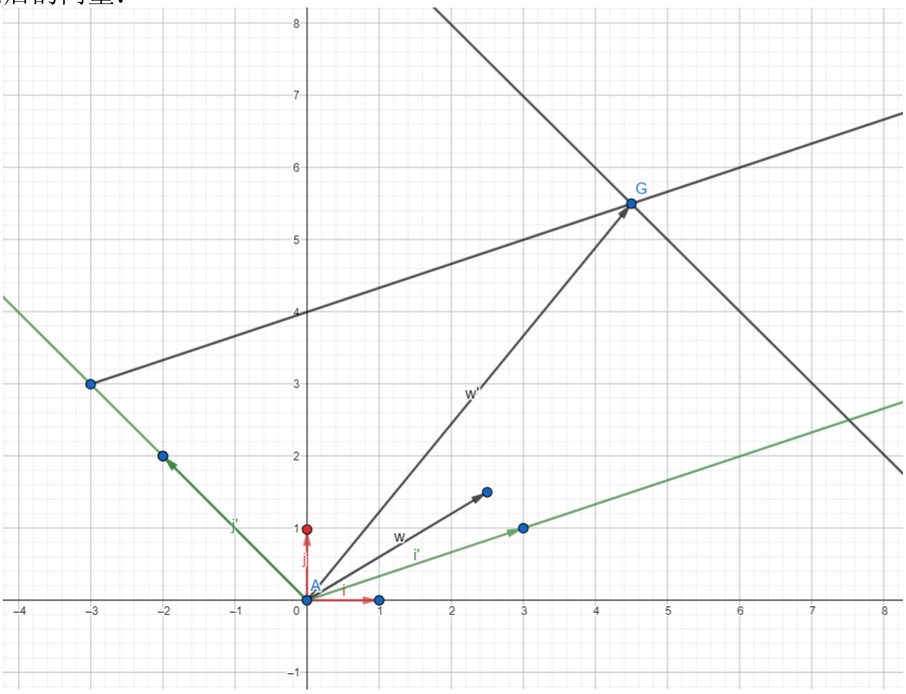
我们可以看到 $A\widehat{w}^T$：

$$A\begin{bmatrix}2.5\\ 1.5\end{bmatrix}=2.5A\widehat{i}^T+1.5A\widehat{j}^T \tag{一.2}$$

$$A\widehat{i}^T=(3,1)^T \tag{一.3}$$

$$A\widehat{j}^T=(-2,2)^T \tag{一.4}$$

如下图，红色向量表示原来的基。绿色向量表示经过线性变换以后的新的基，w 是初始向量，w' 是变换以后的向量：



顺便提一句，我们可以举一反三，当 $\widehat{i}$ 和 $\widehat{j}$ 不是坐标轴上的单位向量时，上面的变换也依然是成立的。

当矩阵的行向量或者列向量是线性相关的时候，例如：

$$A_1 = \begin{bmatrix} 1 & -1 \\ -2 & 2 \end{bmatrix} \tag{一.5}$$

$$A_2 = \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix} \tag{一.6}$$

A_1 将单位基 $(1,0)$ 和 $(0,1)$ 变换为 $(1,-2)$ 和 $(-1,2)$，注意变换后的两个基是平行的，它们已经无法张成整个二维空间了。A_2 也是同理，将单位基 $(1,0)$ 和 $(0,1)$ 变换为 $(1,-1)$ 和 $(-2,2)$。

我们总结一下矩阵乘以向量的值：

$$\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} b_1 \begin{bmatrix} a_{1,1} \\ a_{2,1} \end{bmatrix} & b_2 \begin{bmatrix} a_{1,2} \\ a_{2,2} \end{bmatrix} \end{bmatrix} \tag{一.7}$$

连续线性变换

对于连续相乘的矩阵，我们可以理解为，将基进行了连续的变换。我们假设两个变换矩阵：

$$A_1 = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \tag{一.8}$$

$$B_1 = \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \tag{一.9}$$

方便起见还是使用单位基 $(1,0)$ 和 $(0,1)$。我们目标是计算 $A_1 B_1 \widehat{i}^T$ 和 $A_1 B_1 \widehat{j}^T$，可以看到，先与 B_1 相乘后，基变为了：

$$(b_{1,1}, b_{2,1}) \tag{一.10}$$

$$(b_{1,2}, b_{2,2}) \tag{一.11}$$

然后再与 A_1 相乘，之后基就变为了我们求解矩阵相乘以后得到的结果：

$$(a_{1,1}b_{1,1} + a_{1,2}b_{2,1}, a_{2,1}b_{1,1} + a_{2,2}b_{2,1}) \tag{一.12}$$

$$(a_{1,1}b_{1,2} + a_{1,2}b_{2,2}, a_{2,1}b_{1,2} + a_{2,2}b_{2,2}) \tag{一.13}$$

二　特征值与特征向量

对于某个线性变换，用矩阵 A 表示（A 为方阵）。特征值和特征向量表示如下，若对于矩阵 A：

$$A\widehat{x} = \lambda\widehat{x} \tag{二.1}$$

则 $\widehat{x}$ 称为 A 的特征向量，λ 称为矩阵 A 的特征值。

这意味着什么呢？意味着矩阵 A 对向量 $\widehat{x}$ 进行了线性变换以后，仅仅是将 $\widehat{x}$ 进行了放缩，而没有任何其他的变换。

我们从上面的内容进行理解，可以认为矩阵变换了基，然后由基组合形成的该向量只是被进行了缩放。

例如对于三维旋转矩阵来说，因为它的特征向量变换后没有改变方向，因此三维旋转矩阵的特征向量就是其旋转轴。而且二维旋转矩阵是没有特征向量的，因为二维空间上除了 $(0,0)$ 以外的任何向量旋转以后它的方向都会发生改变（有趣的是这个时候的特征值为复数，即复数对应了旋转操作）。

特征值与特征向量的求解可以写为（其中 I 表示单位矩阵）：

$$(A - \lambda I)\widehat{x} = \widehat{0} \tag{二.2}$$

这样我们就能立即为，我们令矩阵 A 减去 λ 倍的单位矩阵，然后再作用于向量 $\widehat{x}$。我们变换 λ 时，说一句无关紧要的话。思考一个问题，什么变换可以将一个向量变换为 $\widehat{0}$ 向量呢？我们会在末尾给一个形象一点的解释

注意我们在求特征值的时候，上式其实是一个齐次方程组（见 DezemingFamily 的《矩阵与方程组的解》），要使得其次方程组存在非 0 解（即使得 $\widehat{x}$ 是非 0 向量），则其秩必须要小于未知数个数，即对于此方阵来说，行列式要等于 0（见 DezemingFamily 的《矩阵与行列式》），因此需要计算：$\det(A - \lambda I) = 0$ 从而得到特征值。

我们以下面的矩阵为例：

$$A_1 = \begin{bmatrix} 6 & 1 \\ 4 & 3 \end{bmatrix} \tag{二.3}$$

$$A_1 - 2I = \begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix} \tag{二.4}$$

我们可以看到，$A_1 - 2I$ 对线性空间进行了降维，即它作用于基 $(1,0)$ 和 $(0,1)$ 以后，两个基变为了共线，它们将不再能线性组合出一个二维空间，因为它们互为另一个的倍数关系（注意这跟我们选取的基没有任何关系，比如我们选择 $(2,3)$ 和 $(1,6)$ 作为构成二维空间的基，$A_1 - 2I$ 作用于基之后同样进行了降维）。

我们再举个例子：

$$A_2 = \begin{bmatrix} 5 & 4 \\ 1 & 2 \end{bmatrix} \tag{二.5}$$

$$A_2 - 1I = \begin{bmatrix} 4 & 4 \\ 1 & 1 \end{bmatrix} \tag{二.6}$$

A_2 作用于二维空间的两个基以后，这两个基将不再能线性组合成整个二维空间。

同理，假设有一个三维方阵 A_3，它的特征值为 λ_3，则 $A_3 - \lambda_3 I$ 作用于三维空间的三个基之后，这三个基也将不再能线性组合出整个三维空间。

再回到这个问题，究竟什么变换可以将一个向量变换为 $\widehat{0}$ 向量呢？我们考虑二维向量，我们知道变换作用于向量其实就是作用于它的基。我们就可以理解为，变换以后，该向量垂直于已经被变换为共线的基了，而因为该空间任何向量经过该线性变换以后都会和基共线，由垂直且共线，即为 $\widehat{0}$ 向量。

三　基的变换

关于更详细的与基变换有关的内容参考 DezemingFamily 的《线性空间基变换》。

我们由第一章可知，不同的向量用不同的基表示，系数是不同的。例如分别有两组基：

$$basis_1 : (1, 2), (3, 1) \tag{三.1}$$

$$basis_2 : (2, 3), (1, 4) \tag{三.2}$$

那么，用第二组基表示的 $(1, 3)$ 如何用第一组基表示呢？我们先描述一下 $(1, 3)$，有第一章可得矩阵列向量就是基：

$$\begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \tag{三.3}$$

因此很容易就能得到：

$$\begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \tag{三.4}$$

求逆矩阵就能继续求出 (x_1, y_1) 来了。这些过程对理解基和矩阵的关系具有重要作用。

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \tag{三.5}$$

我们可以总结：如果想用一组基 $basis_2$ 来表示用 $basis_1$ 表示的向量，则需要将 $basis_2$ 基构成的矩阵求其逆，并作用于 $basis_1$ 表示的向量上。

四　矩阵的对角化

当一个矩阵只有对角线上有元素的时候，称为对角矩阵。对角矩阵作用于基之后，只是将基进行了放缩，而没有任何旋转操作。

$$A_1 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \tag{四.1}$$

$$A_1(1, 0)^T = (2, 0)^T \tag{四.2}$$

$$A_1(0, 1)^T = (0, 3)^T \tag{四.3}$$

$$A_1(2, 3)^T = (4, 9)^T \tag{四.4}$$

因此我们注意到，对角矩阵的特征值都在对角线上。

我们知道矩阵 A 作用于特征向量只会对其进行拉伸，我们还是以二维矩阵为例，假设它有特征向量 $\widehat{x_1}$ 和 $\widehat{x_2}$ 以及特征值 λ_1 和 λ_2：

$$A_1 \begin{bmatrix} \widehat{x_1} & \widehat{x_2} \end{bmatrix} = \begin{bmatrix} \lambda_1 \widehat{x_1} & \lambda_2 \widehat{x_2} \end{bmatrix} \tag{四.5}$$

从基变换角度下考虑，我们假设某个向量 $\widehat{y}$ 可以表示成上面的特征向量的组合 (y_1, y_2)：

$$\begin{bmatrix} \widehat{x_1} & \widehat{x_2} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{四.6}$$

即矩阵 A 作用于该向量 $\widehat{y}$ 不会有旋转变换，只是进行了拉长，我们可以将 (y_1, y_2) 表示为：

$$A_1 \begin{bmatrix} \widehat{x_1} & \widehat{x_2} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 y_1 \widehat{x_1} & \lambda_2 y_2 \widehat{x_2} \end{bmatrix} \tag{四.7}$$

$$\begin{bmatrix} \widehat{x_1} & \widehat{x_2} \end{bmatrix}^{-1} A_1 \begin{bmatrix} \widehat{x_1} & \widehat{x_2} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 y_1 \\ \lambda_2 y_2 \end{bmatrix} \tag{四.8}$$

因为上面的式子只会将 $\widehat{y}$ 进行拉伸，而不会有任何旋转操作，因此可知下面的矩阵 B_1 为对角矩阵：

$$B_1 = \begin{bmatrix} \widehat{x_1} & \widehat{x_2} \end{bmatrix}^{-1} A_1 \begin{bmatrix} \widehat{x_1} & \widehat{x_2} \end{bmatrix} \tag{四.9}$$

更多与矩阵对角化性质的有关内容参见 DezemingFamily 的《矩阵的相似对角化》。

五　二次型与特征值

关于矩阵与二次型的详细内容参见 DezemingFamily 的《矩阵二次型》，我们这里只是根据该书的二次型与椭圆继续往下讲。

首先这里简单提一句，实对称矩阵的特征向量是相互正交的，可以构成正交矩阵。正交变换不改变向量的长度，分别作用于两个向量也不改变它们的夹角也不变，因此正交变换不会改变形状。

我们考虑一下该二次型：

$$f(x,y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \frac{5}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{3}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{五.1}$$

$$= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{五.2}$$

我们可以从正交矩阵性质得到：

$$\left(\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \right) = \left(\begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right)^T \tag{五.3}$$

回想上一节的矩阵的对角化可知，二次型标准型的中间的对角矩阵只是对向量进行了缩放，而：

$$\left(\begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right)^T \tag{五.4}$$

只是对向量进行了旋转操作，这也是为什么该二次型矩阵为系数和直接以对角矩阵为系数的椭圆形状完全一样，只是进行了旋转：

$$f(x,y) = 2.5x^2 + \frac{3}{2}y^2 + \sqrt{3}xy = 9 \tag{五.5}$$

$$g(u,v) = 3u^2 + v^2 = 9 \tag{五.6}$$

六　总结

对于矩阵的描述一直都是非常困难的事情，因为矩阵的本质是线性映射 [3]，而有时候可以从几何的角度去理解矩阵变换到底发生了什么。我在几年前，大概是 2017 年看了一遍 [1] 中的描述，讲得很简单，也很有意思，但用空间变换来描述矩阵的作用貌似有些片面。矩阵的噩梦一直伴随着我，什么 Jorden 标准型、Hermite 等等，越是了解，越能感到里面的神奇；直到现在读博，线性代数对我来说也是一个神奇而且有趣的工具，里面还是有很多值得探索的内容。

总之，线性代数和线性泛函在应用数学领域一直都是非常重要的一个方面，不仅仅是做一些计算，更重要的是理解背后的意义。

参考文献

[1] https://www.bilibili.com/video/BV1ys411472E?p=1

[2] Strang G . Introduction to Linear Algebra[M]. Wellesley-Cambridge Press, 2003.

[3] Axler S J . Linear Algebra Done Right[J]. Undergraduate Texts in Mathematics, 2015.