机器学习的线性代数基础与Python增强理解

王星
WANGXING@RUC.EDU.CN

参考
HTTPS://BLOG.CSDN.NET/BRYANT_MENG/ARTICLE/DETAILS/99010621

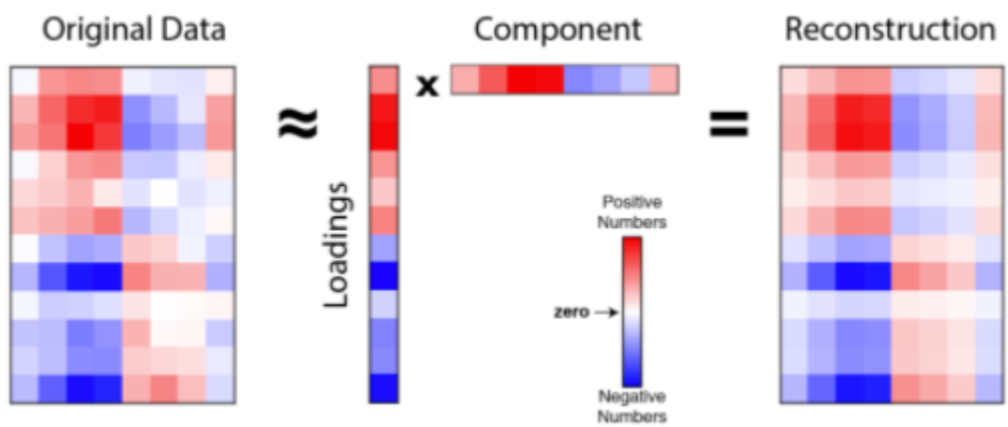
HTTPS://ZHUANLAN.ZHIHU.COM/P/327042762

HTTPS://GITHUB.COM/MACROANALYST/LINEAR_ALGEBRA_WITH_PYTHON/BLOB/MASTER/CHAPTER%202%20-%20BASIC%20MATRIX%20ALGEBRA.IPYNB

线性代数----矩阵家族

1. 基本符号和基本概念---内积和外积
2. 矩阵的乘法和线性方程组的求解---2个视角4个空间：行视角和列视角，零空间和左零空间）
3. 矩阵的运算和性质
   - 转置(transpose)、对称矩阵(symmetric)、迹(trace)、范数(norm)
   - 行列式(deterministic)
   - 二次型和正定矩阵(Defiite)
   - ✓特征值和特征向量(eigen,eigenvector)
   - 对称矩阵的特征值和特征向量
4. 可视化

1.内积和外积



内积

最常见的"教科书式"的矩阵乘法，可以看成2个向量的点积:

$$c_1 = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \sum_{i=1}^{3} a_i b_i$$

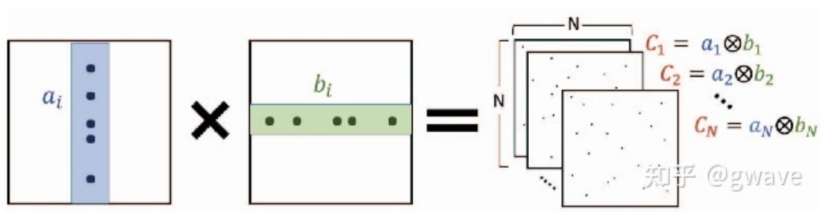
$$\begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \begin{bmatrix} b_1 & b_2 & b_3 \\ b_4 & b_5 & b_6 \\ b_7 & b_8 & b_9 \end{bmatrix} = \begin{bmatrix} c_1 & c_2 & c_3 \\ c_4 & c_5 & c_6 \\ c_7 & c_8 & c_9 \end{bmatrix}$$

外积

矩阵乘法也可看成是向量外积之和。

$$u_{[m\times 1]}v_{[1\times n]}=b_{[m\times n]}$$

$$uv^T=\begin{bmatrix}2\\2\\1\end{bmatrix}\begin{bmatrix}3 & 4 & 5\end{bmatrix}=\begin{bmatrix}6 & 8 & 10\\6 & 8 & 10\\3 & 4 & 5\end{bmatrix}$$



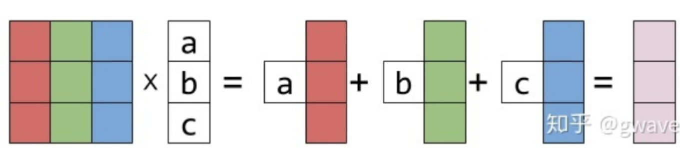
uv^T 的列空间是1维的，向量方向都与 u 相同； 行空间也是类似的，向量方向都与 v^T 相同。所有非零 uv^T 都是秩1矩阵，它们是构建任何矩阵的完美基础砖块。上图体现了：矩阵乘法可以看成是 n 个秩1矩阵叠加之和。

2.1列空间

这种**以行为主**的**内积**运算(也称点积)，将左矩阵($m\times n$)第1行和右矩阵($n\times p$)第1列（上例是向量，可视为1列的矩阵， $p=1$ ）进行点积，进行 n 次乘法和 $n-1$ 次加法才得到结果矩阵的第1行第1列元素 a_{11} ： $(2，3)\cdot(x_1,x_2)=2x_1+3x_2$ ，这种计算方式颗粒度太细，适合机器运算，却不利于理解；我们应该拥有更高维度更清晰的视角 —— **列向量**:

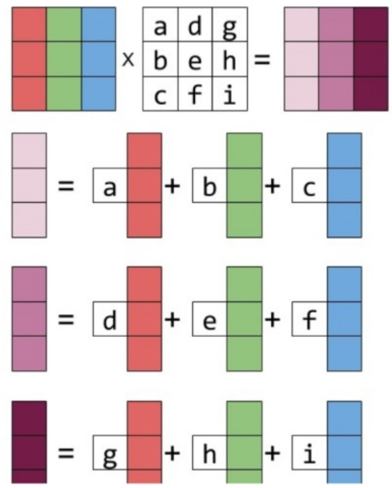
$$\begin{bmatrix}2 & 3\\2 & 4\\3 & 7\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=x_1\begin{bmatrix}2\\2\\3\end{bmatrix}+x_2\begin{bmatrix}3\\4\\7\end{bmatrix}$$

以列为主的视角将矩阵乘法视为**对**矩阵的**列向量**进行**线性组合**，这是线性代数非常重要的本质和基础，并很自然地引出了**列空间**的概念。



矩阵乘法-列视角：列向量对矩阵进行线性组合，得到结果矩阵中的一列

扩展到矩阵的乘法

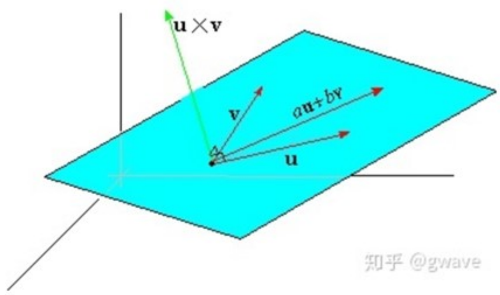


我们知道：两条相交的直线确定一个平面。x_1, x_2 为任意实数，$x_1\begin{bmatrix}2\\2\\3\end{bmatrix}+x_2\begin{bmatrix}3\\4\\7\end{bmatrix}$ 能组合出一个2维平面空间内的任何向量，该平面只是3维空间 R^3 中无数平面中的一个（子空间），但对于位于该平面内的任意向量 $b=\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}$ ，一定存在 (x_1, x_2) ，能使得 $x_1\begin{bmatrix}2\\2\\3\end{bmatrix}+x_2\begin{bmatrix}3\\4\\7\end{bmatrix}=b$ ，该平面是这两个向量所张成（Span）的**列空间**（Column Space），这两个向量是列空间的**基**（Basis)，通过对基的**线性组合**，可以得到空间内的任意向量，难怪有人说线代是"搞基"。

基的选择可以很任性，只要不平行就行，俗称线性无关，要尽量选择垂直的，称为正交基，最好还要是长度为1的单位向量。一旦两个向量平行了，其中任一向量是另一向量的倍数，没有贡献新的信息量，两个向量线性相关，选基时与一个向量没有区别。

列空间由矩阵 A 的列向量线性组合（Linear Combination）填充而成，记为 $C(A)$ 。

线性代数的初心和核心是求解 $Ax=b$ ， A 是 $m\times n$ 的矩阵，以列向量表示: $[a_1，\ a_2]$ ， x 和 b 分别是 $n\times 1$ 和 $m\times 1$ 的向量， b 位于 A 的列空间时， x 一定存在解 (x_1,x_2) 将 b 表达为 A 的列向量线性组合 $x_1a_1+x_2a_2$ 。



u,v通过线性组合张成的空间

考虑 $m\times n$ 的矩阵 A ，它的列为 $\overrightarrow{a_1},\cdots,\overrightarrow{a_n}$ ，列空间:

$$C(A)=\left\{\vec{b}|\vec{b}=A\vec{x},\vec{x}\in\mathbb{R}^n\right\}$$

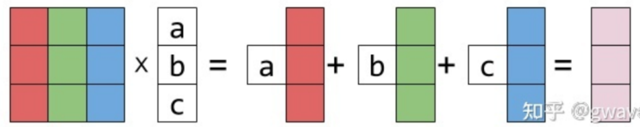
$$=\left\{\vec{b}|\vec{b}=x_1\overrightarrow{a_1}+\cdots+x_n\overrightarrow{a_n},\vec{x}\in R^n\right\}$$

$$=Span(\left\{\overrightarrow{a_1},\cdots,\overrightarrow{a_n}\right\})$$

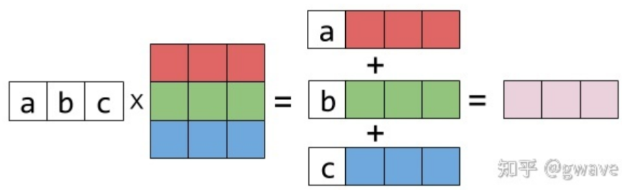
$$\subset\mathbb{R}^m$$

2.2行空间

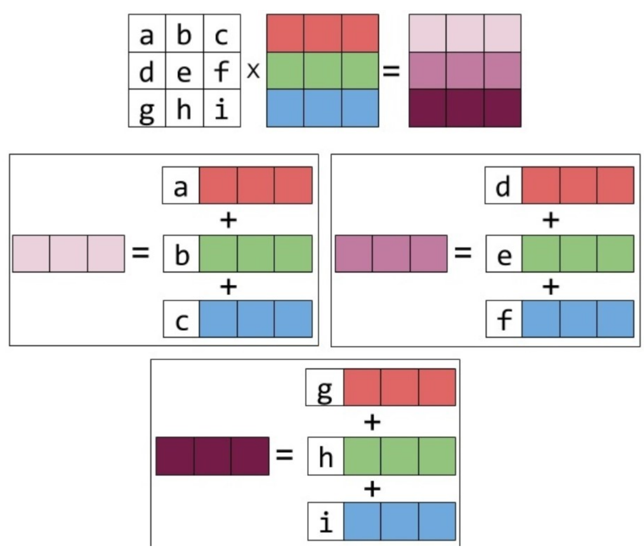
- 矩阵右乘列向量，向量对矩阵的列进行线性组合，结果向量位于矩阵的列空间 $C(A)$；
- 矩阵左乘行向量，向量对矩阵的行进行线性组合，结果向量位于矩阵的行空间 $C(A^T)$。



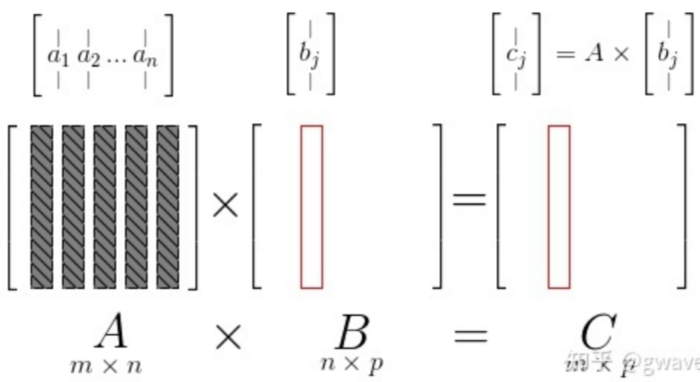
矩阵-向量乘法-列视角：矩阵右乘列向量，向量对矩阵的列进行线性组合

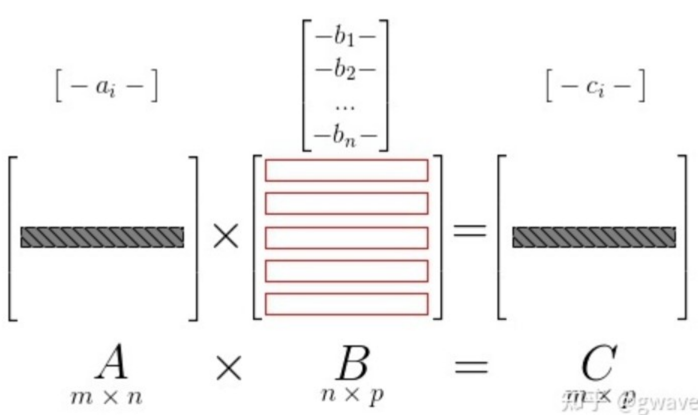


扩展到一般的矩阵



最后，总结一下以列视角和行视角的矩阵乘法：共同点都是以向量对矩阵的向量进行线性组合，不同点在于：列视角是右乘向量，行视角是左乘向量。





2.3零空间

与**行空间**紧密相关的空间———**零空间**(Null Space，也称线性映射的Kernel)，但不同于行空间/列空间的关注是 $Ax=b$ 中的 b ，零空间 $N(A)$ 关注的是 x 的空间，该空间中的任何向量都是**齐次线性方程组**(Homogeneous linear equations) $Ax=0$ 的解，而这与线性代数的核心问题 $Ax=b$ 的求解密切相关。

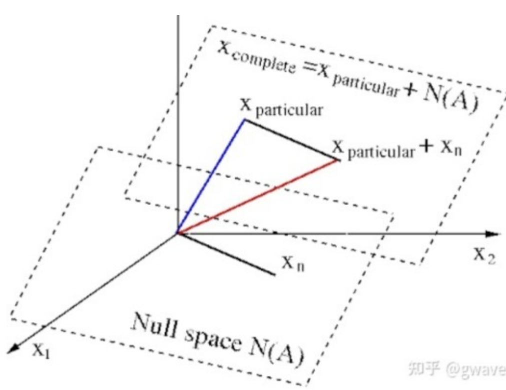
这里不得不吐槽一下**齐次线性方程组**术语的翻译，Homogeneous的含义是‘同种类的’，同种类的含义不仅是‘齐次’，而且更重要的**方程等号右侧（RHS）都等于0**！教科书的例子大多都是‘齐次’的，在术语中强调齐次意义不大，齐次**没有**体现出RHS**等于0**的重要特点。

看一个 $Ax=0$ 的零空间的例子：

$$\begin{bmatrix}1&1&2\\1&2&3\\1&3&4\\1&4&5\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}0\\0\\0\\0\end{bmatrix}$$

$C\begin{bmatrix}1\\1\\-1\end{bmatrix}$ ，C 是常数。还有一个平凡解 $\begin{bmatrix}0\\0\\-0\end{bmatrix}$ ，可以认为是 $C=0$ 的情况。

零空间在求解**非齐次线性方程组（Non-homogeneous systems of linear equations）** $Ax=b$ 中起到重要作用，假设 u 和 v 是非齐次线性方程组 $Ax=b$ 的两个解，那么：$A(u-v)=Au-Av=b-b=0$ ，即 $(u-v)$ 位于 A 的零空间中。因此，$Ax=b$ 完整的解 $X_{complete}$ 可以表述为两部分之和 $(v+z)$ ， v 是使 $Ax=b$ 成立的任意解，称为特解（ $X_{particular}$ ）， z 是 $Ax=0$ 的零空间 $N(A)$ 。用几何语言来表述：$Ax=b$ 的解是在特解 v 的基础上平移 A 的零空间。



最后，我们将上面的例子完整的计算下零空间，以展示完整过程，以及与行空间的关系。

- 考虑矩阵：$A=\begin{bmatrix}1&1&2\\1&2&3\\1&3&4\\1&4&5\end{bmatrix}$
- 该矩阵的零空间 $N(A)$ 包括所有向量 $(x,y,z)\in\mathbb{R}^3$ ，使得：

$$\begin{bmatrix}1&1&2\\1&2&3\\1&3&4\\1&4&5\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}0\\0\\0\\0\end{bmatrix}$$

- 以线性方程组的形式来表示：

$$\begin{cases}x+y+2z=0\\x+2y+3z=0\\x+3y+4z=0\\x+4y+5z=0\end{cases}$$

- 以增广矩阵来表示：

$$\left(\begin{array}{ccc|c}1&1&2&0\\1&2&3&0\\1&3&4&0\\1&4&5&0\end{array}\right)$$

- 高斯-乔丹消元法（Gauss–Jordan elimination）将方程组约减为RREF(Reduced Row Echelon Form，约减的行阶梯形式)：

$$\left(\begin{array}{ccc|c}1&0&1&0\\0&1&1&0\\0&0&0&0\\0&0&0&0\end{array}\right)$$

- 重写方程组

$$\begin{cases}x=-z\\y=-z\end{cases}$$

- 令 $z=1$,得到零空间，

$$\begin{bmatrix}x\\y\\z\end{bmatrix}=c\begin{bmatrix}-1\\-1\\1\end{bmatrix}\quad where\quad c\in\mathbb{R}$$

2.4左零空间

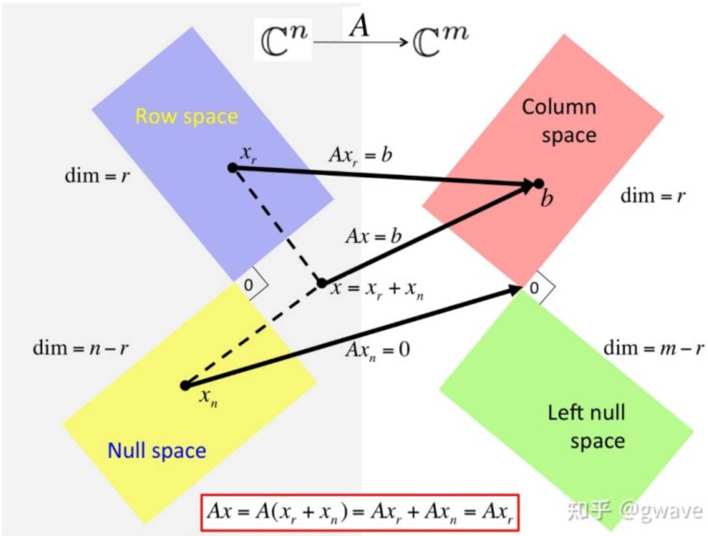
A 右乘 x 即 $Ax=0$ 得到了零空间，那么让 A 左乘 x^T ，使 $x^TA=0^T$ ，会得到什么呢？

对！你没猜错：**左零空间**(Left Null Space)！

之前提到，行空间没有属于自己的符号[1]，只能使用列空间的符号加转置 $C(A^T)$ 来表示；而 $x^TA=0^T$ 中的 x^T 不仅没有属于自己的符号，甚至连像样的名字的都没有，被嫌弃地称为**左零空间**，已经卑微到不仅没有大名 $N(A^T)$ ，就连小名也起的很随意，只是在别人名字前加了个“左”，这和“二狗”有什么差别？难道左零是沦落天涯的二等公民，不配拥有自己的名字？？

- 左边 n 维空间，由于 A 是 n 列的，故 A 的行向量和 x 向量都是 n 维，行空间 $C(A^T)$ 与零空间 $N(A)$ 这两个空间也都位于 $\mathbb{R}^n$ 维空间中；两者正交： $C(A^T)\perp N(A)$ 。
- 右边 m 维空间，由于 A 是 m 行的，故 A 的列向量和 x^T 向量都是 m 维，列空间 $C(A)$ 与左零空间 $N(A^T)$ 这两个空间也都位于 $\mathbb{R}^m$ 维空间中；两者正交： $C(A)\perp N(A^T)$ 。
- $Ax=b$ 的解等于特解 x_r + 零空间 x_n 。

零空间也被称为线性映射的核(**kernel** of Linear Map)，简称Kernal或Ker；相应的，左零空间被称为Cokernel，或者Coker[2]。列空间也被称为Picture，类似的，行空间则为Co-Picture；还有Variance和Co-Variance。



考虑 $Ax = 0$ ，行空间与零空间正交： $C(A^T) \perp N(A)$ ；同样的，因 $x^T A = 0^T$ ，左零空间与列空间正交： $C(A) \perp N(A^T)$ 。于是，这四个子空间构成了线性代数的 Big Picture，

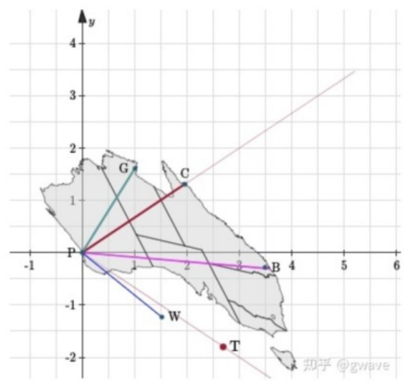
$Ax = 0$ 可以看成是 A 的 m 个行向量与 x 点积的结果等于0，这意味所有的行向量与 x 都是正交的，即：**行空间与零空间正交**： $C(A^T) \perp N(A)$ 。（题图所示）

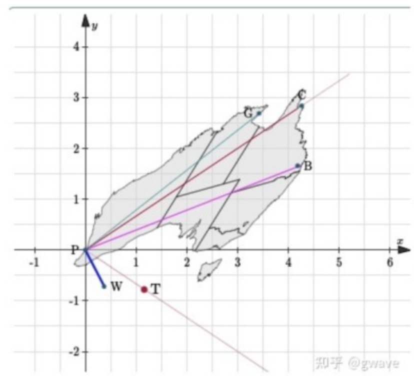
零空间在求解**非齐次线性方程组**（**Non-homogeneous systems of linear equations**） $Ax = b$ 中起到重要作用，假设 u 和 v 是非齐次线性方程组 $Ax = b$ 的两个解，那么： $A(u - v) = Au - Av = b - b = 0$ ，即 $(u - v)$ 位于 A 的零空间中。因此， $Ax = b$ 完整的解 $X_{complete}$ 可以表述为两部分之和 $(v + z)$ ， v 是使 $Ax = b$ 成立的任意解，称为特解（ $X_{particular}$ ）， z 是 $Ax = 0$ 的零空间 $N(A)$ 。用几何语言来表述： $Ax = b$ 的解是在特解 v 的基础上平移 A 的零空间。

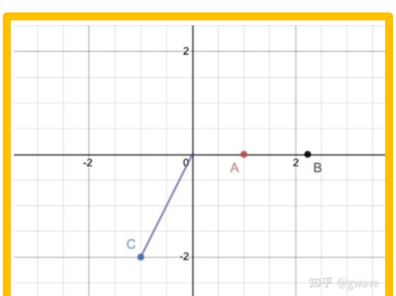
$m \times n$ 的矩阵的左乘与右乘，以及4个核心的子空间

	Ax (m维空间)	x'A / A'y (n维空间)
b	列空间 C(A)	行空间 C(A')
x	左零空间 N(A')	零空间 N(A)

3.特征根和特征空间的几何意义







A延伸到B旋转到C

矩阵乘法的作用是线性变换(Linear Transformation)，将一个向量变换为另一个向量，效果与函数相同：输入向量 x ，输出向量 Ax ，即 $f(x)=Ax$ ，或 $f:x\to Ax$ ，比如：

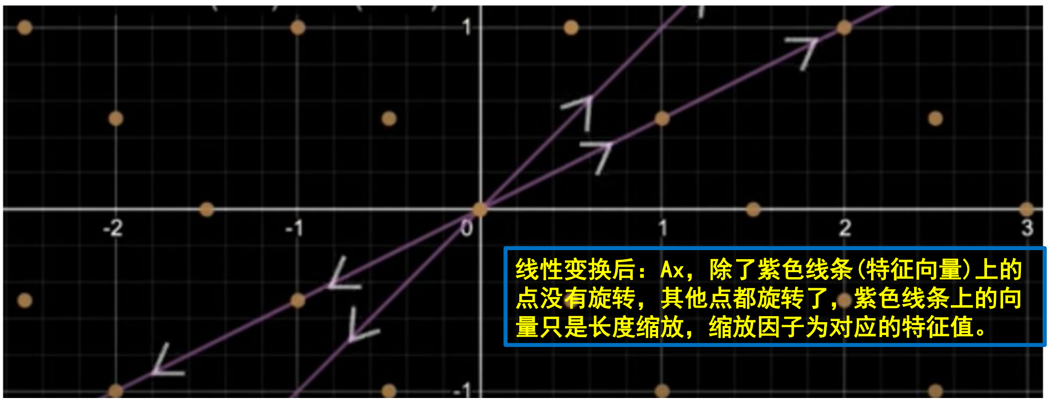
$A=\begin{bmatrix}-1 & 4\\ -2 & 5\end{bmatrix}$ ， $x=\begin{bmatrix}1\\0\end{bmatrix}$ ， $Ax=\begin{bmatrix}-1\\-2\end{bmatrix}$ (了解列向量的同学看一眼便知，对列向量线性组合： $Ax=\begin{bmatrix}-1 & 4\\ -2 & 5\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}-1\\-2\end{bmatrix}\cdot 1+\begin{bmatrix}4\\5\end{bmatrix}\cdot 0=\begin{bmatrix}-1\\-2\end{bmatrix}$)，从函数的角度看： $f(x)=f(\begin{bmatrix}1\\0\end{bmatrix})=\begin{bmatrix}-1\\-2\end{bmatrix}$ 。

从几何的角度来看，此过程可分解为2步：

1. 缩放：从A $\begin{bmatrix}1\\0\end{bmatrix}$ 延伸到B $\begin{bmatrix}\sqrt{5}\\0\end{bmatrix}$
2. 旋转：从B $\begin{bmatrix}\sqrt{5}\\0\end{bmatrix}$ 旋转到C $\begin{bmatrix}-1\\-2\end{bmatrix}$

是否存在矩阵乘法后，只长度缩放，不旋转的点呢？这里的点和向量是同义词。

平面内，除了2条直线上的点没有旋转之外(仅长度缩放)，所有其他点都是既缩放又旋转(不考虑原点)。这两条直线对应的就是**特征向量**，缩放比例就是**特征值**。



计算 A 的特征值和特征向量：

$A=\begin{bmatrix}-1 & 4\\ -2 & 5\end{bmatrix}$ ，令行列式 $|\lambda I-A|=0$ ， I 是单位矩阵：

$|\lambda I-A|=\begin{vmatrix}\lambda+1 & 4\\ -2 & \lambda-5\end{vmatrix}=\lambda^2-4\lambda+3=0$ ， $\lambda_1=1$； $\lambda_2=3$

斜率较小的那条线的 $\lambda=1$ ，说明此线上的向量长度没有改变；另一条直线对应的 $\lambda=3$ ，说明它的每个向量的长度延伸3倍。

接下来求特征向量：

$\lambda=1$ 时，令 $(\lambda I-A)x=0$

$$\lambda I-A=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}-\begin{bmatrix}-1 & 4\\ -2 & 5\end{bmatrix}=\begin{bmatrix}2 & -4\\ 2 & -4\end{bmatrix}=2\cdot\begin{bmatrix}1 & -2\\ 1 & -2\end{bmatrix}$$

令 $(\lambda I-A)x=0$ ， $x=\begin{bmatrix}2\\ 1\end{bmatrix}$ ，对应斜率较小的直线，线上所有向量长度没有变化。

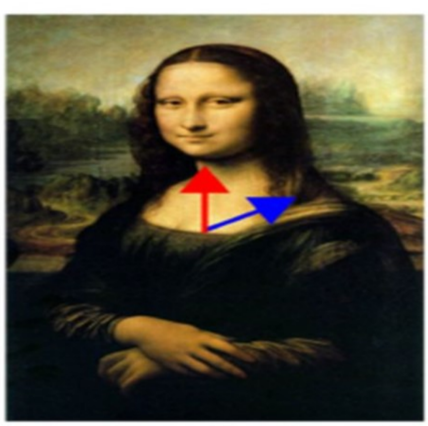
令 $(\lambda I - A)x = 0$ ， $x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ ，对应斜率较小的直线，线上所有向量长度没有变化。

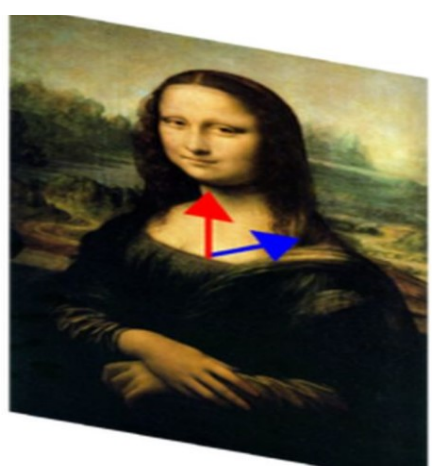
$\lambda = 3$ 时，令 $(\lambda I - A)x = 0$

$$\lambda I - A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} - \begin{bmatrix} -1 & 4 \\ -2 & 5 \end{bmatrix} = \begin{bmatrix} 4 & -4 \\ 2 & -2 \end{bmatrix} = 2 \cdot \begin{bmatrix} 2 & -2 \\ 1 & -1 \end{bmatrix}$$

令 $(\lambda I - A)x = 0$ ， $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ，对应斜率较大的直线，线上所有向量长度放大3倍。

P.S. 令 $|\lambda I - A| = 0$ 的原因是基于特征向量和特征值的定义 $Ax = \lambda x$ ，移项后求解零空间所致。





定理 1 实对称矩阵的特征多项式在复数域中的每一个根都是实数.

证明 设 A 是 n 级实对称矩阵. 设 λ_0 是 A 的特征多项式 $|\lambda I - A|$ 在复数域中的任意一个根, 于是 $|\lambda_0 I - A| = 0$. 从而齐次线性方程组 $(\lambda_0 - A)X = 0$ 有非零解. 取它的一个非零解

$$\alpha = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}.$$

则 $(\lambda_0 I - A)\alpha = 0$. 从而

$$A\alpha = \lambda_0\alpha. \tag{5}$$

想证 $\bar{\lambda}_0 = \lambda_0$, 于是在(5) 式两边取复数共轭, 得

$$\bar{A}\bar{\alpha} = \bar{\lambda}_0\bar{\alpha}, \tag{6}$$

由于 A 是实矩阵, 因此 $\bar{A} = A$. 从而(6) 式也就是

$$A\bar{\alpha} = \bar{\lambda}_0\bar{\alpha}. \tag{7}$$

(7) 式两边左乘 α', 得

$$\alpha' A\bar{\alpha} = \bar{\lambda}_0\alpha'\bar{\alpha}. \tag{8}$$

由于 A 是对称矩阵, 因此 $A' = A$. 在(5) 式两边取转置, 然后用 $\bar{\alpha}$ 右乘, 得

$$\alpha' A\bar{\alpha} = \lambda_0\alpha'\bar{\alpha}. \tag{9}$$

比较(8)、(9) 两式, 得

$$\bar{\lambda}_0\alpha'\bar{\alpha} = \lambda_0\alpha'\bar{\alpha}.$$

即

$$(\bar{\lambda}_0 - \lambda_0)\alpha'\bar{\alpha} = 0. \tag{10}$$

由于 $\alpha \neq 0$, 因此

$$\begin{aligned} \alpha'\bar{\alpha} &= c_1\bar{c}_1 + c_2\bar{c}_2 + \cdots + c_n\bar{c}_n \\ &= |c_1|^2 + |c_2|^2 + \cdots + |c_n|^2 \neq 0. \end{aligned}$$

于是从(10) 式得, $\bar{\lambda}_0 - \lambda_0 = 0$. 即 $\bar{\lambda}_0 = \lambda_0$. 因此 λ_0 是实数. ▮

定理 2 实对称矩阵 A 的属于不同特征值的特征向量是正交的.

证明 设 λ_1 与 λ_2 是 A 的不同特征值,α_i 是 A 的属于 λ_i 的特征向量,$i=1,2$. 由于

$$\lambda_1(\alpha_1,\alpha_2)=(\lambda_1\alpha_1,\alpha_2)=(A\alpha_1,\alpha_2)=(A\alpha_1)'\alpha_2=\alpha_1'A'\alpha_2=\alpha_1'A\alpha_2,$$
$$\lambda_2(\alpha_1,\alpha_2)=(\alpha_1,\lambda_2\alpha_2)=(\alpha_1,A\alpha_2)=\alpha_1'A\alpha_2,$$

因此 $\lambda_1(\alpha_1,\alpha_2)=\lambda_2(\alpha_1,\alpha_2)$. 于是 $(\lambda_1-\lambda_2)(\alpha_1,\alpha_2)=0$. 由于 $\lambda_1\neq\lambda_2$, 因此 $(\alpha_1,\alpha_2)=0$. 即 α_1 与 α_2 正交.

定理 3 实对称矩阵一定正交相似于对角矩阵.

证明 对于实对称矩阵的级数 n 作数学归纳法.

$n=1$ 时,(a) 已经是对角矩阵,且 $I_1^{-1}(a)I_1=(a)$.

假设任意一个 $n-1$ 级实对称矩阵都能正交相似于对角矩阵. 现在来看 n 级实对称矩阵 A.

取 A 的一个特征值 λ_1(这由定理 1 保证可取到),取 A 的属于 λ_1 的一个特征向量 η_1,且 $|\eta_1|=1$. η_1 可扩充成 $\mathbf{R}^n$ 的一个基(见第 3 章 §4 定理 1 后面的注),然后经过施密特正交化和单位化,可得到 $\mathbf{R}^n$ 的一个标准正交基:$\eta_1,\eta_2,\cdots,\eta_n$. 令

$$T_1=(\eta_1,\eta_2,\cdots,\eta_n),$$

则 T_1 是 n 级正交矩阵. 我们有

$$\begin{aligned}T_1^{-1}AT_1&=T_1^{-1}(A\eta_1,A\eta_2,\cdots,A\eta_n)\\&=(T_1^{-1}\lambda_1\eta_1,T_1^{-1}A\eta_2,\cdots,T_1^{-1}A\eta_n).\end{aligned}$$

因为 $T_1^{-1}T_1=I$,所以

$$T_1^{-1}(\eta_1,\eta_2,\cdots,\eta_n)=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n).$$

于是得,$T_1^{-1}\eta_1=\varepsilon_1$. 从而 $T_1^{-1}AT_1$ 的第 1 列是 $\lambda_1\varepsilon_1$. 因此可以设

$$T_1^{-1}AT_1=\begin{pmatrix}\lambda_1 & \alpha\\ 0 & B\end{pmatrix}.$$

由于 T_1 是正交矩阵，A 是实对称矩阵，因此 $T_1^{-1}AT_1$ 也是实对称矩阵. 从而得，$\alpha = 0$，并且 B 是 $n-1$ 级实对称矩阵. 据归纳假设，有 $n-1$ 级正交矩阵

T_2，使得

$$T_2^{-1}BT_2 = \operatorname{diag}\{\lambda_2, \cdots, \lambda_n\}.$$

令

$$T = T_1\begin{pmatrix} 1 & 0 \\ 0 & T_2 \end{pmatrix},$$

由于上式右端的两个矩阵都是 n 级正交矩阵，因此 T 是 n 级正交矩阵，并且有

$$\begin{aligned} T^{-1}AT &= \begin{pmatrix} 1 & 0 \\ 0 & T_2 \end{pmatrix}^{-1} T_1^{-1}AT_1 \begin{pmatrix} 1 & 0 \\ 0 & T_2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & T_2^{-1} \end{pmatrix}\begin{pmatrix} \lambda_1 & 0 \\ 0 & B \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & T_2 \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & 0 \\ 0 & T_2^{-1}BT_2 \end{pmatrix} \\ &= \operatorname{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_n\}. \end{aligned}$$