

第 01 讲 行图像和列图像

Row picture & Column picture

线性方程的几何图像 The geometry of linear equations

线性代数的基本问题就是解 n 元一次方程组。例如：二元一次方程组

$$\begin{aligned} 2x - y &= 0 \\ -x + 2y &= 3 \end{aligned}$$

写成矩阵形式就是

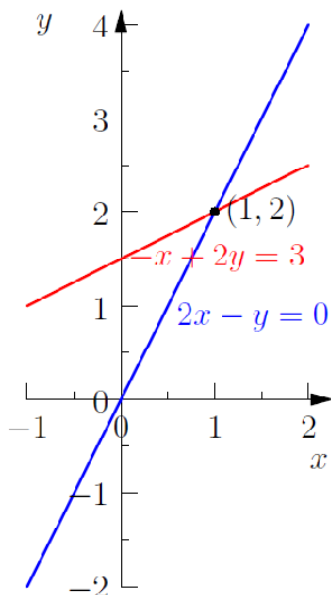
$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

其中 $\mathbf{A} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ 被称为系数矩阵 (coefficient matrix)。

未知数向量通常记为 $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$

而等号右侧的向量记为 \mathbf{b} 。线性方程组简记为 $\mathbf{Ax} = \mathbf{b}$ 。

行图像 Row Picture



行图像遵从解析几何的描述，每个方程在平面上的图像为一条直线。找到符合方程的两个数组，就可以确定出 x - y 平面上的两个点，连接两点可以画出该方程所代表的直线。两直线交点即为方程组的解 $x=1, y=2$ 。

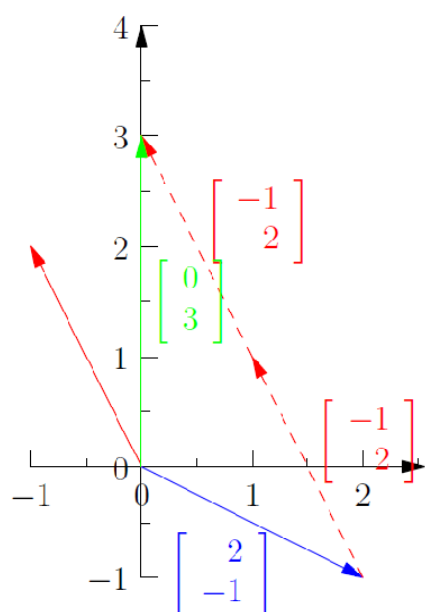
列图像 Column Picture

在列图像中，我们将系数矩阵写成列向量的形式，则求解原方程变为寻找列向量的线性组合 (linear combination) 来构成向量 \mathbf{b} 。

$$x \begin{bmatrix} 2 \\ -1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

向量线性组合是贯穿本课程的重要概念。对于给定的向量 **c** 和 **d** 以及标量 x 和 y ，我们将 $x\mathbf{c}+y\mathbf{d}$ 称之为 **c** 和 **d** 的一个线性组合。

从几何上讲，我们是寻找满足如下要求的 x 和 y ，使得两者分别数乘对应的列向量之后相加得到向量 $\begin{bmatrix} 0 \\ 3 \end{bmatrix}$ 。其几何图像如下图。



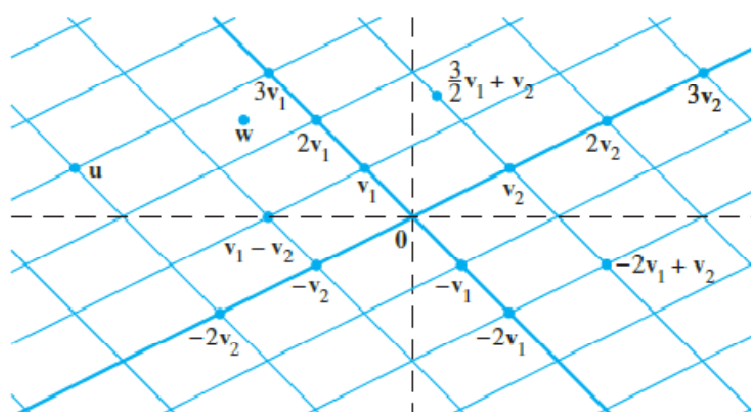
蓝色为向量 $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$ ；

红色为向量 $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ ；

可以看到当蓝色的向量乘以 1 与红色的向量乘以 2（红色虚线）后做加法（首尾相接）就可以得到绿色的向量 $\mathbf{b} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$ ，

由此可得到方程的解 $x=1, y=2$ 。

想象一下如果任意取 x, y ，则得到的线性组合又是什么？其结果就是以上两个列向量的所有线性组合将会布满整个坐标平面。



D.C.Lay 的《线性代数及其应用》中，绘制向量 $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ 和 $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 的线性组合充满整个平面的图像，节点处为向量的整数倍线性组合。

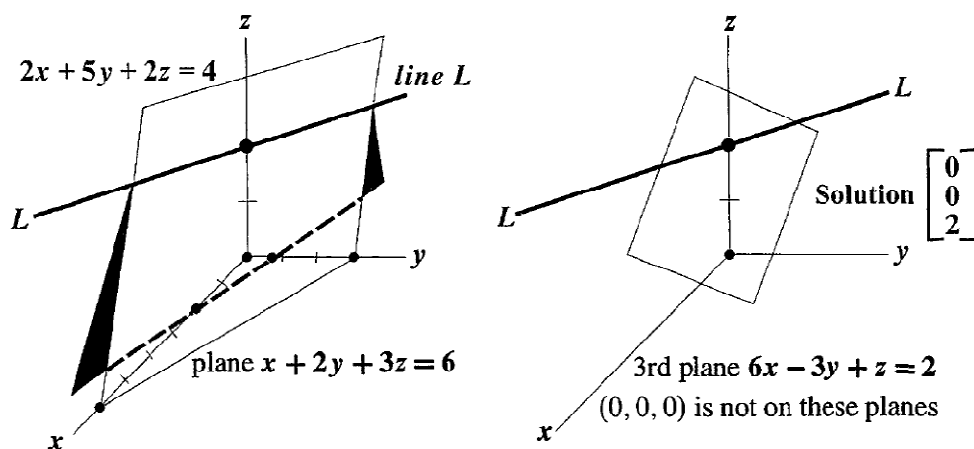
这本书也是难得的好书，作者喜欢利用几何图像来帮助读者理解线性代数中的

概念，英文版出到第 5 版了，华章出过中译本。（是不是觉得上面那个图片有点斜，是斜线造成的错觉哟！）

将以上讨论扩展到三元。图不好弄，所以用了 GS 书里的另一个方程，没有用视频中的那个 !!! 这样方程和配图是吻合的。

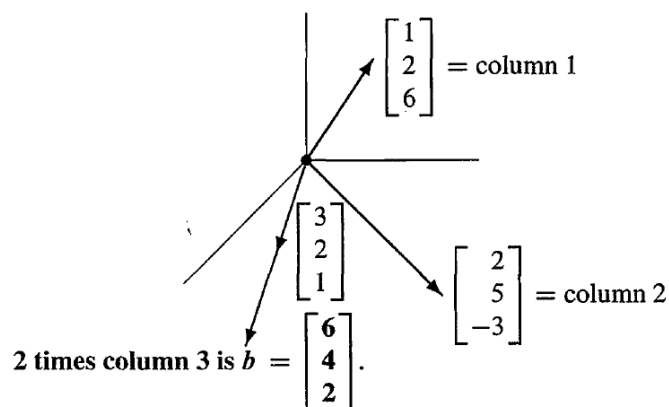
$$\begin{aligned} x + 2y + 3z &= 6 \\ 2x + 5y + 2z &= 4 \\ 6x - 3y + z &= 2 \end{aligned} \quad \text{矩阵形式} \quad \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 6 & -3 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}$$

方程的行图像比较复杂，每一个方程都是三维空间内的一个平面，方程组的解为三个平面的交点。



画图真不是 GS 的长项，在视频里画的就比较 shi，他自己也承认了。在课本里他用两个面相交于一条直线画了一个图，然后让这条直线和第三个平面相交画了第二个图。同样的事，D.C.Lay 一张图分分钟搞定。

$$\text{方程组列图像为 } x \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix} + y \begin{bmatrix} 2 \\ 5 \\ -3 \end{bmatrix} + z \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}$$



如果改变等号右侧的 **b** 的数值，那么对于行图像而言三个平面都改变了，而对于列图像而言，三个向量并没有发生变化，只是需要寻找一个新的组合。

那么问题来了，是否对于所有的 \mathbf{b} ，方程 $\mathbf{Ax}=\mathbf{b}$ 都有解？

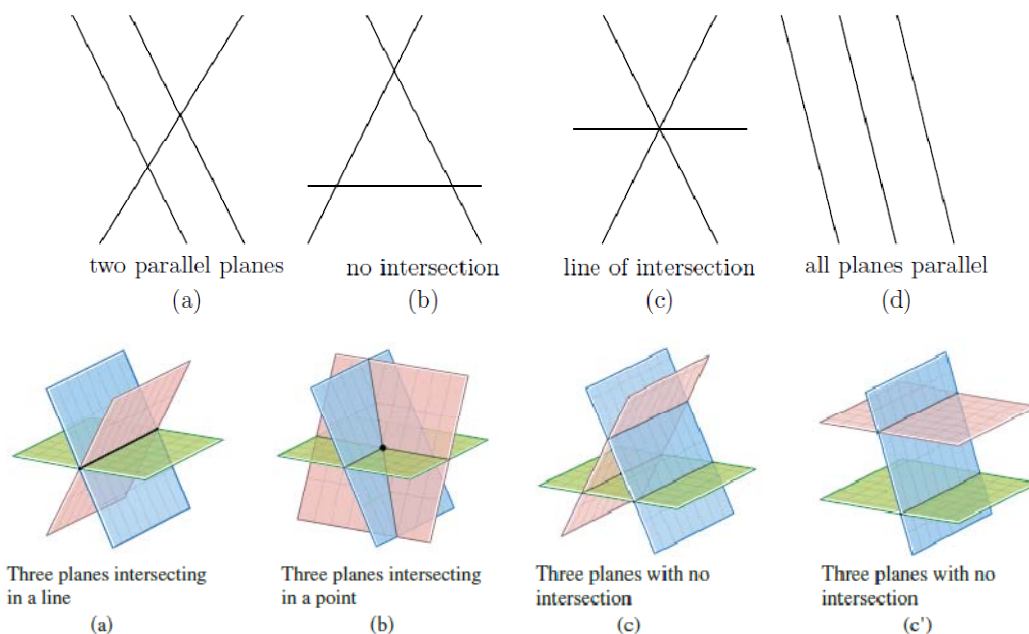
从列图像上看，问题转化为“列向量的线性组合是否覆盖整个三维空间？”

反例：若三个向量在同一平面内——比如“列 3”恰好等于“列 1”加“列 2”，而若 \mathbf{b} 不在该平面内，则三个列向量无论怎么组合也得不到平面外的向量 \mathbf{b} 。此时矩阵 \mathbf{A} 为奇异阵或称不可逆矩阵。在矩阵 \mathbf{A} 不可逆条件下，不是所有的 \mathbf{b} 都能令方程 $\mathbf{Ax}=\mathbf{b}$ 有解。

对 n 维情形则是， n 个列向量如果相互独立——“线性无关”，则方程组有解。否则这 n 个列向量起不到 n 个的作用，其线性组合无法充满 n 维空间，方程组未必有解。

从行图像的角度来看，三元方程组是否有解意味着什么？当方程所代表的三个平面相交于一点时方程有唯一解；三个平面中至少两个平行则方程无解；平面的两两交线互相平行方程也无解；三个平面交于一条直线则方程有无穷多解。

都是示意图，来看看 GS 和 Lay 的作图差异有多大吧.....



矩阵与向量的乘法

列图像： \mathbf{Ax} 是矩阵 \mathbf{A} 列向量的线性组合：

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 12 \\ 7 \end{bmatrix}$$

也可以通过将矩阵 \mathbf{A} 的行向量和 \mathbf{x} 向量进行点积来计算：

$$\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \times 2 + 2 \times 5 \\ 1 \times 1 + 2 \times 3 \end{bmatrix} = \begin{bmatrix} 12 \\ 7 \end{bmatrix}$$

第 02 讲 矩阵消元

Elimination with matrices

消元法 Method of Elimination

消元法是计算机软件求解线性方程组所用的最常见的方法。任何情况下，只要是矩阵 A 可逆，均可以通过消元法求得 $Ax=b$ 的解。

此处给出的线性方程组为

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 12 \\ 2 \end{bmatrix}$$

高斯消元法 (Gauss elimination) 就是通过对方程组中的某两个方程进行适当的数乘和加 (jian) 和 (fa) , 以达到将某一未知数系数变为零, 从而削减未知数个数的目的。

我们将矩阵左上角的 1 称之为 “主元一” (the first pivot), 第一步要通过消元将第一列中除了主元之外的数字均变化为 0。操作方法就是用之后的每一行减去第一行的适当倍数, 此例中第二行应减去第一行的 3 倍。之后应对第三行做类似操作, 本例中三行第一列数字已经为 0, 故不用进行操作。

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} \xrightarrow{(2,1)} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix} \xrightarrow{(3,2)} U = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 5 \end{bmatrix}$$

处在第二行第二列的主元二为 2, 因此用第三行减去第二行的两倍进行消元, 得到第三个主元为 5。

矩阵 A 为可逆矩阵, 消元结束后得到上三角阵 U (Uppertriangular matrix), 其左侧下半部分的元素均为 0, 而主元 1,2,5 分列在 U 的对角线上。主元之积即行列式的值。

需要说明的是, 主元不能为 0, 如果恰好消元至某行, 0 出现在了主元的位置上, 应当通过与下方一行进行 “行交换” 使得非零数字出现在主元位置上。如果 0 出现在了主元位置上, 并且下方没有对等位置为非 0 数字的行, 则消元终止, 并证明矩阵 A 为不可逆矩阵, 且线性方程组没有唯一解。

例如消成这样

$$\begin{bmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & 0 \end{bmatrix}$$

回代 Back-Substitution

做方程的高斯消元时，需要对等式右侧的 **b** 做同样的乘法和加减法。手工计算时比较有效率的方法是应用“增广矩阵” (augmented matrix)，将 **b** 插入矩阵 **A** 之后形成最后一列，在消元过程中带着 **b** 一起操作。(Matlab 是算完系数矩阵再处理 **b** 的。)

$$\begin{bmatrix} 1 & 2 & 1 & 2 \\ 3 & 8 & 1 & 12 \\ 0 & 4 & 1 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & 2 & -2 & 6 \\ 0 & 4 & 1 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & 2 & -2 & 6 \\ 0 & 0 & 5 & -10 \end{bmatrix}$$

此时我们将原方程 **Ax=b** 转化为了新的方程 **Ux=c**，其中 $\mathbf{c} = \begin{bmatrix} 2 \\ 6 \\ -10 \end{bmatrix}$ 。

从最后一行得到 $z = -2$ ，依次回代可以得到 $y = 1$ 和 $x = 2$ 。

以上高斯消元法的内容基本是回忆求解线性方程的步骤，是我们很熟悉的东西。在线性代数中比较重要的就是将之前所说的“第二行减去第一行的 3 倍”这种操作条例变为矩阵化的数学语言。

消元矩阵 Elimination Matrices

矩阵运算的核心内容就是对“行”或者“列”进行独立操作。

如前一节课“列图像”部分所言，系数矩阵乘以未知数向量，相当于对系数矩阵的列向量进行线性组合。

例如 $\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} = 3 \begin{bmatrix} * \\ * \\ * \end{bmatrix} + 4 \begin{bmatrix} * \\ * \\ * \end{bmatrix} + 5 \begin{bmatrix} * \\ * \\ * \end{bmatrix} = \begin{bmatrix} \otimes \\ \otimes \\ \otimes \end{bmatrix}$

列 1 列 2 列 3

与之相对称，矩阵左乘行向量则是对矩阵的行向量进行线性组合。

例如 $\begin{bmatrix} 1 & 2 & 7 \end{bmatrix} \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix} = 1 \begin{bmatrix} * & * & * \end{bmatrix} + 2 \begin{bmatrix} * & * & * \end{bmatrix} + 7 \begin{bmatrix} * & * & * \end{bmatrix} = \begin{bmatrix} \oplus & \oplus & \oplus \end{bmatrix}$

行 1 行 2 行 3

我学这部分的时候，列向量的线性组合非常容易就接受了，左乘行向量这个总觉得很别扭，偏偏行向量在下面介绍消元矩阵时比较重要。后来想想觉得“列”操作就像是把向量开进矩阵，而“行操作”这个就像把向量倒车进入矩阵（如图中箭头所示）。-_-! 虽然挺扯淡的，但是我反正突然就觉得习惯了。

矩阵消元的第一步是通过左乘矩阵 **E₂₁** 来实现原矩阵 **A** 的第二行减去第一行的

3 倍这一过程。 E_{21} 的第二行使矩阵 A 的行向量进行前述的线性组合，而其它两行为了保持与原矩阵相同，采用同阶单位阵 I 的行向量。左乘的这个矩阵为“初等矩阵” (Elementary Matrix)，因此记做 E 。我以为是消元矩阵，所以记做 E 呢。因为所乘行向量的倍数-3 出现在 E 矩阵的第二行第一列，因此将之标注为 21。完成操作后矩阵变为 $E_{21}A$ 。

$$\begin{aligned}
 \text{第一行: } [1 \ 0 \ 0] & \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} = [1 \ 2 \ 1] \\
 \text{第三行: } [0 \ 0 \ 1] & \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} = [0 \ 4 \ 1] \\
 \text{关键第二行: } [-3 \ 1 \ 0] & \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} = \begin{matrix} -3 & [1 & 2 & 1] \\ + & 1 & [3 & 8 & 1] \\ + & 0 & [0 & 4 & 1] \end{matrix} = [0 \ 2 \ -2] \\
 \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} &= \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix} \\
 \mathbf{E}_{21} \quad \mathbf{A} \quad \mathbf{E}_{21}\mathbf{A}
 \end{aligned}$$

矩阵消元的第二步是完成矩阵 $E_{21}A$ 的第三行减去第二行的 2 倍，通过左乘矩阵 E_{32} 来实现这一过程。

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 5 \end{bmatrix} \\
 \mathbf{E}_{32} \quad \mathbf{E}_{21}\mathbf{A} \quad \mathbf{E}_{32}(\mathbf{E}_{21}\mathbf{A})$$

3×3 矩阵的消元本来应该分三步完成，最终得到 $E_{32}(E_{31}(E_{21}A))$ 。本例中 $E_{31}=I$ ，所以结果变为 $E_{32}(E_{21}A) = U$ ，因为矩阵运算符合结合律，也可写作 $(E_{32}E_{21})A = U$ 。可以记作 $EA=U$ 。

方程 $Ax=b$ 的解也满足方程 $Ux=EAx=Eb=c$ ，因此我们将问题转化为 $Ux=c$ 。

置换矩阵 Permutation

左乘置换矩阵可以完成原矩阵的行变换，右乘置换矩阵则为列变换。例如

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} c & d \\ a & b \end{bmatrix} \\
 \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} b & a \\ d & c \end{bmatrix}$$

$$\text{对于三阶矩阵, } \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \underline{a} & \underline{b} & \underline{c} \\ d & e & f \\ \underline{g} & \underline{h} & \underline{i} \end{bmatrix} = \begin{bmatrix} \underline{g} & \underline{h} & \underline{i} \\ d & e & f \\ \underline{a} & \underline{b} & \underline{c} \end{bmatrix}$$

构造 P 矩阵是通过对 I 矩阵进行“行交换”来实现的。

左右乘效果不同也展示了矩阵运算不符合交换律的性质。

逆矩阵 Inverse

这里主要讨论消元矩阵的逆矩阵。消元矩阵之逆矩阵的实施效果就是抵消原矩阵的消元操作。消元矩阵实现了对原矩阵 A 的操作，使第二行行向量[3,8,1]减掉了第一行[1,2,1]的 3 倍变为[0,2,-2]，则逆向操作就应该是把现在的第二行行向量[0,2,-2]加上第一行[1,2,1]的 3 倍，从而变回原来的第二行[3,8,1]。

$$\text{所以对于 } E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ 有 } E_{21}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}。$$

$$\text{满足 } E_{21}^{-1} E_{21} = I \quad \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

第 03 讲 矩阵的乘法和逆矩阵

Multiplication & inverse matrices

矩阵乘法 Matrix multiplication

我们通过四种方法讨论如何使矩阵 A 与 B 相乘得到矩阵 C 。其中 A 为 $m \times n$ (m 行 n 列) 矩阵, 而 B 为 $n \times p$ 矩阵, 则 C 为 $m \times p$ 矩阵, 记 c_{ij} 为矩阵 C 中第 i 行第 j 列的元素。

1) 标准方法 (行乘以列)

矩阵乘法的标准计算方法是通过矩阵 A 第 i 行的行向量和矩阵 B 第 j 列的列向量点积得到 c_{ij} 。

举例

$$AB = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ a_{31} & a_{32} & \dots & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \begin{bmatrix} * & * & * & b_{14} & * \\ * & * & * & b_{24} & * \\ * & * & * & \dots & * \end{bmatrix} = \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & c_{34} & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}$$

矩阵 A $m \times n$ 矩阵 B $n \times p$ 矩阵 C $m \times p$

$$c_{34} = \text{row}3 \cdot \text{col}4 = \sum_{k=1}^n a_{3k} b_{k4} = a_{31}b_{14} + a_{32}b_{24} + a_{33}b_{34} + \dots$$

所谓标准方法就是矩阵乘法的运算规则, 下面扯几句矩阵代数运算的淡。念书的时候, 学到矩阵代数, 同桌问我: “矩阵运算怎么这么奇怪, 只有同样型号的矩阵 (指同为 $m \times n$) 可以做加法, 但是不同型号的矩阵却可以做乘法, 话说这不同型号的矩阵为啥要做这种复杂的乘法? 数乘一个矩阵为啥其中每一个元素都要乘以这个数, 而行列式则不是?” 我后来找到了 A.D.亚历山大洛夫编著的《数学: 它的内容, 方法和意义》第三卷中找到了答案。

先补一下矩阵的加法和数乘规则:

Addition $A + B$ $\begin{bmatrix} 2 & 1 \\ 3 & 0 \\ 0 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ -3 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 0 & 1 \\ 1 & 6 \end{bmatrix}$ $2 \begin{bmatrix} 2 & 1 \\ 3 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 6 & 0 \\ 0 & 8 \end{bmatrix}$
Multiplication $2A$

本质上矩阵的代数运算就是线性方程组的运算：

$\begin{cases} y_1 = a_{11}x_1 + \dots + a_{1n}x_n \\ \dots\dots\dots \\ y_m = a_{m1}x_1 + \dots + a_{mn}x_n \end{cases}$	<p>可以简写做矩阵形式 $\mathbf{y} = \mathbf{A}\mathbf{x}$ \mathbf{A} 为 $m \times n$ 矩阵</p>
$\begin{cases} z_1 = b_{11}x_1 + \dots + b_{1n}x_n \\ \dots\dots\dots \\ z_m = b_{m1}x_1 + \dots + b_{mn}x_n \end{cases}$	<p>$\mathbf{z} = \mathbf{B}\mathbf{x}$ \mathbf{B} 为 $m \times n$ 矩阵</p>
$\begin{cases} y_1 + z_1 = (a_{11} + b_{11})x_1 + \dots + (a_{1n} + b_{1n})x_n \\ \dots\dots\dots \\ y_m + z_m = (a_{m1} + b_{m1})x_1 + \dots + (a_{mn} + b_{mn})x_n \end{cases}$	<p>$\mathbf{y} + \mathbf{z} = (\mathbf{A} + \mathbf{B})\mathbf{x}$ 所以矩阵加法运算规则就只能形状相同的矩阵进行加和</p>
$\begin{cases} \alpha y_1 = \alpha a_{11}x_1 + \dots + \alpha a_{1n}x_n \\ \dots\dots\dots \\ \alpha y_m = \alpha a_{m1}x_1 + \dots + \alpha a_{mn}x_n \end{cases}$	<p>$\alpha \mathbf{y} = \alpha \mathbf{A}\mathbf{x}$ 数乘要在矩阵每个元素上乘以 α</p>

而矩阵乘法可以视为给线性方程组做变量替换

$\begin{cases} y_1 = a_{11}x_1 + \dots + a_{1n}x_n \\ \dots\dots\dots \\ y_m = a_{m1}x_1 + \dots + a_{mn}x_n \end{cases}$ <p>$\mathbf{y} = \mathbf{A}\mathbf{x}$ \mathbf{A} 为 $m \times n$ 矩阵</p>	$\begin{cases} w_1 = d_{11}y_1 + \dots + d_{1m}y_m \\ \dots\dots\dots \\ w_k = d_{k1}y_1 + \dots + d_{km}y_m \end{cases}$ <p>$\mathbf{w} = \mathbf{D}\mathbf{y}$ \mathbf{D} 为 $k \times m$ 矩阵</p>
<p>将左侧 y_1 至 y_m 的表达式代入右侧的方程，则有</p> $\begin{cases} w_1 = d_{11}(a_{11}x_1 + \dots + a_{1n}x_n) + \dots + d_{1m}(a_{m1}x_1 + \dots + a_{mn}x_n) \\ \dots\dots\dots \\ w_k = d_{k1}(a_{11}x_1 + \dots + a_{1n}x_n) + \dots + d_{km}(a_{m1}x_1 + \dots + a_{mn}x_n) \end{cases}$ <p>重新整理一下</p> $\begin{cases} w_1 = (d_{11}a_{11} + \dots + d_{1m}a_{m1})x_1 + \dots + (d_{11}a_{1n} + \dots + d_{1m}a_{mn})x_n = c_{11}x_1 + \dots + c_{1n}x_n \\ \dots\dots\dots \\ w_k = (d_{k1}a_{11} + \dots + d_{km}a_{m1})x_1 + \dots + (d_{k1}a_{1n} + \dots + d_{km}a_{mn})x_n = c_{k1}x_1 + \dots + c_{kn}x_n \end{cases}$ <p>$\mathbf{w} = \mathbf{D}\mathbf{y} = \mathbf{D}\mathbf{A}\mathbf{x} = \mathbf{C}\mathbf{x}$</p> <p>$\mathbf{C} = \mathbf{D}\mathbf{A}$ 为 $k \times n$ 矩阵，从上式可得到 c_{ij} 的公式，即矩阵的乘法运算规则，可以看到这是变量代换的结果。所以矩阵乘法中两矩阵形状可以不同，但要求左侧矩阵的列数要等于右侧矩阵的行数，即 $\mathbf{w} = \mathbf{D}\mathbf{y}$ 中 \mathbf{y} 的分量个数要等于 $\mathbf{y} = \mathbf{A}\mathbf{x}$ 中 \mathbf{y} 的分量数，不如此则线性方程的变量代换无法进行。</p>	

矩阵乘法符合结合律、分配率，但是不遵守交换律，这些通常通过定义加以证

明,即将等号两侧矩阵的元素乘开,比较对应元素,从而得出相等或者不等的结论。我们也可以从变量替换的角度思考一下矩阵乘法运算定律。比如结合律,若方程组有如下关系 $u = Ew$, $w = Dy$, $y = Ax$, 则做变量替换可有 $u = Ew = EDy = ED Ax$ 。结合律为 $(ED)A = E(DA)$, 它表达的意思就是替换环节无论是先把 u 表示成 EDy , 再表示成 x 的表达式,又或者先把 w 表示成 DAx ,再把公式带入 $u = Ew$ 都是等效的。分配率也一样,先做加法再进行变量替换,或者先做变量替换再相加也是等效的。但是交换律就不行了,别说矩阵的尺寸可能导致交换顺序后不能进行乘法运算,即使能进行乘法,变量替换的关系也完全不对了。

2)列操作

列操作是指矩阵 C 的第 j 列是通过矩阵 A 乘以矩阵 B 第 j 列的列向量得到的。这表明矩阵 C 的列向量是矩阵 A 列向量的线性组合,组合的“权”就是矩阵 B 第 j 列的各个分量。

$$AB = A \begin{bmatrix} b_1 & b_2 & \dots & b_p \end{bmatrix} = \begin{bmatrix} Ab_1 & Ab_2 & \dots & Ab_p \end{bmatrix}$$

上式中 b_i , Ab_i 均为列向量,其中

$$Ab_j = \begin{bmatrix} \left| \right| & \left| \right| & \dots & \dots & \left| \right| \\ a_1 & a_2 & & & a_n \\ \left| \right| & \left| \right| & & & \left| \right| \end{bmatrix} \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix} = b_{1j} \begin{bmatrix} \left| \right| \\ a_1 \\ \left| \right| \end{bmatrix} + b_{2j} \begin{bmatrix} \left| \right| \\ a_2 \\ \left| \right| \end{bmatrix} + \dots + b_{nj} \begin{bmatrix} \left| \right| \\ a_n \\ \left| \right| \end{bmatrix}$$

为矩阵 A 中列向量的线性组合

3)行操作

行操作是指矩阵 C 的第 i 行是通过矩阵 A 的第 i 行乘以矩阵 B 得到的。这表明矩阵 C 的行向量是矩阵 B 行向量的线性组合。

$$AB = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_m \end{bmatrix} B = \begin{bmatrix} a_1 B \\ a_2 B \\ \vdots \\ \vdots \\ a_m B \end{bmatrix} \quad \text{此处 } a_i \text{ 和 } a_i B \text{ 为行向量}$$

$$a_i B = \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix} \begin{bmatrix} \text{--- } b_1 \text{ ---} \\ \text{--- } b_2 \text{ ---} \\ \vdots \\ \vdots \\ \text{--- } b_n \text{ ---} \end{bmatrix} = \begin{matrix} a_{i1} [\text{--- } b_1 \text{ ---}] \\ + \\ a_{i2} [\text{--- } b_2 \text{ ---}] \\ \vdots \\ + \\ a_{in} [\text{--- } b_n \text{ ---}] \end{matrix}$$

4)列乘以行

矩阵 A 的第 k 列是一个 $m \times 1$ 的向量，而矩阵 B 第 k 行是一个 $1 \times p$ 的向量，两向量相乘会得到一个矩阵 C_k ，将所有的 n 个矩阵相加记得到 C 。

$$AB = \sum_{k=1}^n \begin{matrix} \begin{bmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{bmatrix} \\ m \times 1 \end{matrix} \begin{matrix} \begin{bmatrix} b_{k1} & \cdots & b_{kp} \end{bmatrix} \\ 1 \times p \end{matrix}$$

可以从矩阵乘法对加法的分配率推导出来。

分块乘法 block multiplication

如果将矩阵 A 和矩阵 B 划分为严格匹配的区块，则矩阵乘法可以通过分块的乘法加以实现。

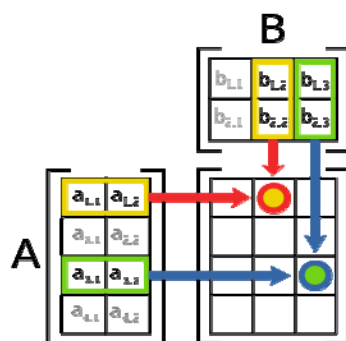
$$\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \end{bmatrix} = \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix}$$

其中 $C_1 = A_1 B_1 + A_2 B_3$ ，计算方法与标准算法中矩阵里元素的操作方式相同。

之前没有提到一点，就是因为矩阵乘法规则有点小复杂，在手算计算过程中有可能会串行，我们不可能每一次都如下图这样认真标注。

$$AB = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \\ b_{41} & b_{42} \end{bmatrix} = \begin{bmatrix} * & * \\ * & * \\ * & * \end{bmatrix} = (AB)_{32}$$

一个比较好的小技巧就是把 B 矩阵的位置改变一下，这样矩阵 C 中每一个元素所对应 A 和 B 中的行与列就变得非常清楚了。我在 wiki 百科中看到了这张图，MIT 多元微积分课程上老师也提到了这种方法。



这里我们可以利用这种小技巧来帮助理解矩阵的分块乘法：

$$\begin{bmatrix} \square & \square & * & * \\ \square & \square & * & * \\ \hline \Delta & \Delta & \Delta & \Delta \\ \Delta & \Delta & \Delta & \Delta \end{bmatrix} \quad \begin{bmatrix} \square & \square & \Delta \\ \square & \square & \Delta \\ \square & \square & \Delta \\ \hline * & * & \Delta \\ * & * & \Delta \end{bmatrix}$$

可以看到 C_1 就是经过 $C_1 = A_1 B_1 + A_2 B_3$ 的运算计算出来的， A_1 中的元素只和 B_1 的元素进行运算， A_2 只和 B_3 进行运算， C_1 为两者加和。

逆矩阵 Inverse

如果矩阵 A 是方阵，若存在逆矩阵 A^{-1} ，使得 $A^{-1}A = I = AA^{-1}$ （左逆矩阵等于右逆矩阵）。我们称矩阵 A 可逆（invertible）或者矩阵 A 非奇异（nonsingular）。反之，如果 A 为奇异（singular），则其没有逆矩阵。它的行列式为 0。另一个等价的说法是， A 为奇异阵，则方程 $Ax=0$ 存在非零解 x 。例如：

$$\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

在这个二阶矩阵的例子中，两个列向量排列在同一方向上。不可逆矩阵中总有列向量对生成线性组合没有贡献，等价的说法还有：不可逆矩阵的列向量可以通过线性组合得到 0。

换言之，若矩阵 A 存在逆矩阵，则方程 $Ax=0$ 只有零解。证明：反设其存在非零解 x ，则有 $x = (A^{-1}A)x = A^{-1}(Ax) = A^{-1}0 = 0$ ，矛盾。

对于可逆矩阵，求它的逆矩阵是一个重要的问题。

$$\begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$A \quad A^{-1} \quad I$

从“列操作”的角度来看，求逆矩阵过程其实和求 $Ax=b$ 相同，只是这里 x 为矩阵 A^{-1} 的第 j 列，而 b 为单位阵 I 的第 j 列。

高斯-若尔当消元法 Gauss-Jordan Elimination

对于前面的二阶矩阵，求逆相当于两组方程：

$$\begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ 和 } \begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Gauss-Jordan 消元法可以同时处理两个方程：

$$\begin{array}{c} \left[\begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 2 & 7 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{array} \right] \rightarrow \left[\begin{array}{cc|cc} 1 & 0 & 7 & -3 \\ 0 & 1 & -2 & 1 \end{array} \right] \\ \mathbf{A} \quad \mathbf{I} \qquad \qquad \mathbf{I} \quad \mathbf{A}^{-1} \end{array}$$

在用高斯消元法的到上三角矩阵之后，按照若尔当的做法继续消元，用第一行减去第二行的若干倍，最后原矩阵变为单位阵，这时右侧的矩阵即为逆矩阵。

对 \mathbf{A} 进行一系列消元操作，相当于左乘消元矩阵 \mathbf{E} ，此时消元的结果为 $\mathbf{EA}=\mathbf{I}$ ，因为右侧矩阵也进行了同样消元操作，也等于左乘矩阵 \mathbf{E} ，则右侧矩阵为 $\mathbf{EI}=\mathbf{E}$ 。由 $\mathbf{EA}=\mathbf{I}$ 可知这里的 \mathbf{E} 就是 \mathbf{A} 的逆矩阵，因此右侧矩阵给出的就是逆矩阵 $\mathbf{E}=\mathbf{A}^{-1}$ 。

$$\mathbf{E}[\mathbf{A} \mid \mathbf{I}] = [\mathbf{EA} \mid \mathbf{EI}] = [\mathbf{I} \mid \mathbf{A}^{-1}]$$

第 04 讲 矩阵的 LU 分解

Factorization into $A=LU$

本节的主要目的是从矩阵的角度理解高斯消元法，最后找到所谓的 L 矩阵，使得矩阵 A 可以转变为上三角阵 U 。即完成 LU 分解得到 $A=LU$ 。

首先继续了解一些矩阵乘法和逆矩阵的相关内容。

矩阵乘积的逆矩阵 Inverse of a product

$$(AB)^{-1} = B^{-1}A^{-1}$$

$ABB^{-1}A^{-1} = A(BB^{-1})A^{-1} = I$ 比较等式两端可得 $(AB)^{-1} = B^{-1}A^{-1}$ 。

矩阵乘积的转置 Transpose of a product

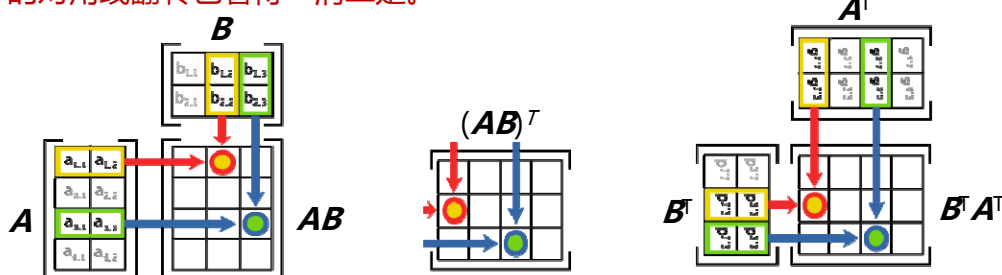
(本讲要用到的只是转置矩阵求逆的公式，把相关内容放在这里做个简介，方便大家理解。GS 在下一讲还会讲到转置。)

矩阵 A 的转置矩阵记为 A^T ，对矩阵进行转置就是将 A 矩阵的行变为 A^T 的列，则完成后 A 的列也就成为了 A^T 的行，看起来矩阵如同沿着对角线进行了翻转。

$$A = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 0 & 3 \end{bmatrix} \quad \text{则} \quad A^T = \begin{bmatrix} 2 & 0 \\ 1 & 0 \\ 4 & 3 \end{bmatrix}$$

其数学表达式为 $(A^T)_{ij} = A_{ji}$ ，即 A^T 的第 i 行 j 列的元素为原矩阵 A 中第 j 行 i 列的元素。

$(AB)^T = B^T A^T$ 可以用定义证明。把上次课介绍的那个乘法小技巧的图片按照乘积矩阵的对角线翻转也看得一清二楚。



转置矩阵的逆矩阵 Inverse of a transpose

$$(A^T)^{-1} = (A^{-1})^T$$

$AA^{-1} = I$ 两边取转置，并应用积的转置公式，得到 $(A^{-1})^T A^T = I$ 。根据逆矩阵定义得到 $(A^T)^{-1} = (A^{-1})^T$

矩阵的 LU 分解

我们已经看到了如何应用消元法将矩阵 A 转变为上三角阵 U ，这就引出了矩阵

的 LU 分解，它是理解矩阵 A 性质的重要方法。可以将矩阵的分解类比为多项式的因式分解，分解后的结果可以让我们更容易看清“解”的状态。

在没有行交换的情况下，矩阵 A 通过左乘一系列消元矩阵 E_j 可以转化为 U 。在二阶矩阵中，进行一次消元操作即可达到这一效果。

$$\begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 8 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$$

$E_{21} \quad A \quad U$

在等式两侧左乘，得到 $E_{21}^{-1}E_{21}A = E_{21}^{-1}U$ 即 $A = E_{21}^{-1}U$ 。就得到了矩阵 A 的 LU 分解结果。 E_{21}^{-1} 是 E_{21} 的逆向操作，即左乘 E_{21} 中使得矩阵 A 第二行[8,7]减去第一行的 4 倍得到“新第二行”[0,3]，那么再左乘 E_{21}^{-1} 可以使得“新第二行”[0,3]加上“第一行”的 4 倍又变回原第二行的数值[8,7]。

$$\begin{bmatrix} 2 & 1 \\ 8 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$$

$A \quad L \quad U$

其中 U 为上三角阵 (Upper triangular matrix)，主元依次排列于它的对角线上， E_{21}^{-1} 即 L 为下三角阵 (Lower triangular matrix)。有时我们也通过分解得到对角阵 D (diagonal matrix)，例如

$$\begin{bmatrix} 2 & 1 \\ 8 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1/2 \\ 0 & 1 \end{bmatrix}$$

$A \quad L \quad D \quad U'$

对于三阶矩阵不需要换行进行消元的情况则有：

$$E_{32}(E_{31}(E_{21}A)) = U, \text{ 左乘逆矩阵可得 } A = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}U = LU$$

设定一组消元矩阵，其中 E_{31} 为单位阵 I ，其它两个消元矩阵如下：

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -5 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 10 & -5 & 1 \end{bmatrix}$$

$E_{32} \quad E_{21} \quad E$

可以看到在消元矩阵 E 的左下角出现了数 10。它的出现是由于第一步操作 E_{21} 中“第二行”减去了 2 倍的“第一行”得到了“新第二行”。 $row_2 - 2row_1 = newrow_2$

而在第二步操作 E_{32} 中第三行减去了 5 倍的“新第二行”，这相当于减去 5 倍的“原第二行”并减去 5 倍的“(-2) 倍原第一行”，10 就出现在这里。

$$row_3 - 5newrow_2 = row_3 - 5(row_2 - 2row_1) = row_3 - 5row_2 + 10row_1$$

在右侧操作则不会有这种情况发生，运算顺序会发生变化， $L = E^{-1} = E_{21}^{-1}E_{32}^{-1}$ 。

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 5 & 1 \end{bmatrix}$$

$E_{21}^{-1} \quad E_{32}^{-1} \quad L$

如果没有行交换操作，则消元矩阵的因子可以直接写入矩阵 L 。没有多余的交叉项出现是 LU 分解要优于 $EA=U$ 这种形式的原因之一。

消元法所需运算量

在一些应用中我们需要处理超大型矩阵，即使用计算机来处理这一问题，也需要评估所需的计算量。

如果我们把“先乘后减”大致记为一次运算，那么对于一个 $n \times n$ 矩阵，对于一行进行消元要进行 n 次运算，由于有 n 行所以进行了 n^2 次运算，结果得到了第一列除第一主元外都消成 0 的矩阵。随后开始对除第一行第一列之外的剩余部分进行消元，这相当于一个 $(n-1) \times (n-1)$ 的矩阵，那么就要进行 $(n-1)^2$ 次运算，以此类推。

最后需要的运算次数为 $1^2 + 2^2 + \dots + n^2$ ，利用积分公式可以估算其数值。

$$1^2 + 2^2 + \dots + n^2 = \sum_{i=1}^n i^2 \approx \int_0^n x^2 dx = \frac{1}{3}n^3$$

算是个灵活使用数学工具的例子，我看到完全平方的求和，就一直在回忆公式，但其实作为估算采用积分公式就很方便了。有时候应用数学工具就像开车，保证你正常行驶主要靠驾驶技能，而不是你手边的一本《交规》，熟练驾驶但是不遵守交规很容易出错，交规倒背如流却没有驾驶技能则往往寸步难行。

再扯几句淡。在统计学中经常用到斯特林公式 (Stirling's approximation)，它用来评估和近似阶乘的大小 $n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$ ，如果你去阅读 Stirling 公式的发展历程和证明过程，就会意识到级数、极限和微积分中间的紧密联系，然而在大多数同学那里这些概念都是割裂开的，我觉得这就是普通人和学霸有差距的地方，知识点互相联系不起来。Stirling 公式的证明过程也用到了积分，感兴趣可以看一看。

等号右侧向量 b 的行变换大致需要 n^2 次运算。

行转换 Row exchanges

如果主元的位置出现了 0，就需要进行“行交换”。我们可以通过左乘一个置换矩阵 (Permutation Matrix) 实现“行交换”的操作。例如

$$P_{21} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

可以实现 3×3 矩阵的第一行与第二行的交换。所有的 3×3 的置换矩阵包括：

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

对于 $n \times n$ 矩阵存在着 $n!$ 个置换矩阵。置换矩阵每一行或者每一列只有一个元素是 1，其它都是 0，从第一行选一个位置设定为 1 有 n 个选择，第二行则只剩下 $n-1$ 个选择，以此类推，最终有 $n!$ 种可能。

对于某阶的置换矩阵集合而言，置换矩阵的两两乘积仍在这个集合中，置换矩阵的逆矩阵也在此集合中。置换矩阵的逆矩阵即为它的转置 $P^{-1} = P^T$ 。

可以用前面介绍过的乘法运算小技巧来理解这件事情

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad P^{-1} = \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}$$

P 的第 i 行和 P^{-1} 的第 i 列相乘会得到 1，与其他列相乘都得到 0，所以 P^{-1} 第 i 列只能是 P 的第 i 行行向量的转置（让分量 1 出现在向量里的相同位置），其他各列以此类推，则 P 的每一个行向量转置就得到 P^{-1} 的列向量，这也就是矩阵转置 P^T 的定义。因此可得 $P^{-1} = P^T$ 。

第 05 讲 转置、置换和空间

Transposes, permutations, spaces \mathbf{R}^n

本节的将引入向量空间 (vector spaces) 和子空间 (subspaces)。

置换 Permutations

当应用消元法求解方程组的时候我们需要通过行交换将 0 从主元位置移走。左乘一个置换阵可以实现行交换的操作。(为了满足数值计算的要求, Matlab 甚至会对接近于 0 的非零主元做行交换) 因此我们的 LU 分解由 $A=LU$ 变为 $PA=LU$ 。其中的 P 就是对 A 的行向量进行重新排序的置换矩阵。

我之前看到这里的时候有个疑惑, 如果在消元的过程中进行 “行交换”, 那么不是意味着 P 矩阵存在于一堆 E 矩阵之间么 $EEEEPEEA=U$ 矩阵乘法不符合交换律, 你怎么能把 P 提出来变成 $PA=LU$ 呢。同桌说可以假想在消元过程中你已经完全知道需要怎么进行 “行交换” 之后, 我们重新开始做矩阵分解, 这一次先对 A 进行 “行交换” 得到 A^* , 这时候对 A^* 消元, 就不用再进行 “行交换” 了, 于是有 $PA=A^*=LU$ 。

置换矩阵 P 是通过对单位阵进行 “行交换” 得到的。对于 $n \times n$ 矩阵存在着 $n!$ 个置换矩阵。置换矩阵具有特殊性质 $P^1=P^T$ 即 $P^T P=I$ 。

转置 Transposes

矩阵 A 的转置矩阵记为 A^T , 对矩阵进行转置就是将 A 矩阵的行变为 A^T 的列, 则完成后 A 的列也就成为了 A^T 的行, 看起来矩阵如同沿着对角线进行了翻转。

其数学表达式为 $(A^T)_{ij}=A_{ji}$, 即 A^T 的第 i 行 j 列的元素为原矩阵 A 中第 j 行 i 列的元素。

$$\begin{bmatrix} 1 & 3 \\ 2 & 3 \\ 4 & 1 \end{bmatrix}^T = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 3 & 1 \end{bmatrix}$$

若 A 是对称矩阵则有 $A^T=A$ 。

矩阵乘积的转置 $(AB)^T=B^T A^T$

给定一个矩阵 R , R 可以不是方阵, 则乘积 $R^T R$ 一定是对称阵。

$$(R^T R)^T = R^T (R^T)^T = R^T R$$

向量空间 Vector spaces

我们可以对向量进行所谓 “线性运算”, 即通过加和 ($v+w$) 与数乘运算 ($3v$) 得到向量的线性组合。向量空间对线性运算封闭, 即空间内向量进行线性运算得到的向量仍在空间之内。

\mathbf{R}^2 即为向量空间, 它是具有两个实数分量的所有向量 (二维实向量) 的集合。

例如 $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} \pi \\ e \end{bmatrix}$

向量 $\begin{bmatrix} a \\ b \end{bmatrix}$ 的图像是从原点出发到点(a,b)的箭头，其中第一分量 a 为横轴坐标，第二分量 b 为纵轴坐标。空间 \mathbf{R}^2 的图像为整个 x-y 平面。

所有向量空间必然包含零向量，因为任何向量数乘 0 或者加上反向量都会得到零向量，而因为向量空间对线性运算封闭，所以零向量必属于向量空间。

\mathbf{R}^3 是向量空间，它是具有三个实数分量的所有向量的集合。

\mathbf{R}^n 是向量空间，它是具有 n 个实数分量的所有向量的集合。

反例： \mathbf{R}^2 中的第一象限则不是一个向量空间。

子空间 Subspaces

包含于向量空间之内的一个向量空间称为原向量空间的一个子空间。例如用实数 c 数乘 \mathbf{R}^2 空间中向量 \mathbf{v} 所得到的向量集合就是 \mathbf{R}^2 空间的一个子空间，其图像为二维平面上穿过原点的一条直线，它对于线性运算封闭。

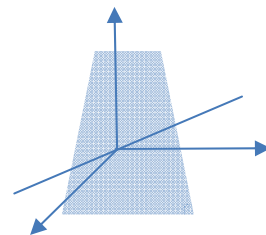
反例： \mathbf{R}^2 中不穿过原点的直线就不是向量空间。子空间必须包含零向量，原因就是数乘 0 的到的零向量必须处于子空间中。

\mathbf{R}^2 的子空间包括：

- \mathbf{R}^2 空间本身
- 过原点的一条直线 (这是 \mathbf{R}^2 空间中的一条直线，与 \mathbf{R}^1 空间有区别)
- 原点 (仅包含 0 向量)

\mathbf{R}^3 的子空间包括：

- \mathbf{R}^3 空间本身 (3 维)
- 过原点的一个平面 (2 维)
- 过原点的一条直线 (1 维)
- 原点 (仅包含 0 向量)



列空间 Column spaces

给定矩阵 \mathbf{A} ，其列向量属于 \mathbf{R}^3 空间，这些列向量和它们的线性组合张成了 \mathbf{R}^3 空间中的一个子空间，即矩阵 \mathbf{A} 的列空间 $C(\mathbf{A})$ 。

如果 $\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 3 \\ 4 & 1 \end{bmatrix}$ ，则 \mathbf{A} 的列空间是 \mathbf{R}^3 空间中包含向量 $\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$ 和 $\begin{bmatrix} 3 \\ 3 \\ 1 \end{bmatrix}$ 并穿过原点的

平面，空间内包含两向量的所有线性组合。

下面课程的任务就是在列空间和子空间的基础上理解 $\mathbf{Ax}=\mathbf{b}$ 。

第 06 讲 列空间和零空间

Column space & Nullspace

本节继续研究子空间，特别是矩阵的列空间 (column space) 和零空间 (nullspace)。

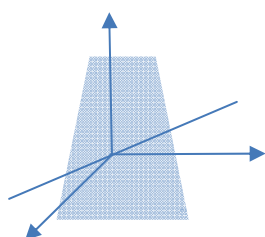
子空间综述

所谓的“向量空间”是对于线性运算封闭的向量集合。即对于空间中的任意向量 \mathbf{v} 和 \mathbf{w} ，其和 $\mathbf{v}+\mathbf{w}$ 和数乘 $c\mathbf{v}$ 必属于该空间；换言之对于任何实数 c 和 d ，线性组合 $c\mathbf{v}+d\mathbf{w}$ 必属于该空间。

$\mathbf{R}^1, \mathbf{R}^2, \mathbf{R}^3, \dots$ 都是重要的向量空间， \mathbf{R}^n 代表的空间包含所有具有 n 个分量的向量。其中字母 \mathbf{R} 表明分量均为实数 (real)。

“子空间”为包含于向量空间内的一个向量空间。它是原向量空间的一个子集，而且本身也满足向量空间的要求。但是“子空间”和“子集”的概念有区别，所有元素都在原空间之内就可称之为子集，但是要满足对线性运算封闭的子集才能成为子空间。

例如



通过原点的平面 P 是 \mathbf{R}^3 的一个子空间。

通过原点的直线 L 也是 \mathbf{R}^3 的子空间。

P 并 L (union) 通常并不是 \mathbf{R}^3 的子空间。

P 交 L (intersection) 是 \mathbf{R}^3 子空间的特例——0 空间，只有零向量。

任意子空间 S 和 T 的交集都是子空间，可以通过 S 和 T 本身对线性组合封闭来证明。

列空间 Column space

矩阵 \mathbf{A} 的列空间 $C(\mathbf{A})$ 是其列向量的所有线性组合所构成的空间。

求解 $\mathbf{Ax}=\mathbf{b}$ 的问题，对于给定的矩阵 \mathbf{A} ，对于任意的 \mathbf{b} 都能得到解么？

$$\text{如果 } \mathbf{A} = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix}$$

显然并不是所有的 \mathbf{b} 都能保证 $\mathbf{Ax}=\mathbf{b}$ 有解，因为它有 4 个线性方程而只有 3 个未知数，矩阵 \mathbf{A} 列向量的线性组合无法充满 \mathbf{R}^4 ，因此如果 \mathbf{b} 不能被表示为 \mathbf{A} 列向量的线性组合时，方程是无解的。只有当 \mathbf{b} 在矩阵 \mathbf{A} 列空间 $C(\mathbf{A})$ 里时， \mathbf{x} 才有解。

对于我们所给定的矩阵 A ，由于列向量不是线性无关的，第三个列向量为前两个列向量之和，所以尽管有 3 个列向量，但是只有 2 个对张成向量空间有贡献。矩阵 A 的列空间为 \mathbf{R}^4 内的一个二维子空间。

零空间 (或化零空间) Nullspace

矩阵 A 的零空间 $N(A)$ 是指满足 $Ax=0$ 的所有解的集合。对于所给定这个矩阵 A ，其列向量含有 4 个分量，因此列空间是空间 \mathbf{R}^4 的子空间， x 为含有 3 个分量的向量，故矩阵 A 的零空间是 \mathbf{R}^3 的子空间。对于 $m \times n$ 矩阵，列空间为 \mathbf{R}^m 的子空间，零空间为 \mathbf{R}^n 空间的子空间。

本例中矩阵 A 的零空间 $N(A)$ 为包含 $\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$ 的任何倍数的集合，因为很容易看到

第一列向量 (1) 和第二列向量 (1) 相加减去第三列向量 (-1) 为零。此零空间为 \mathbf{R}^3 中的一条直线。

为了验证 $Ax=0$ 的解集是一个向量空间，我们可以检验它是否对线性运算封闭。若 v 和 w 为解集中的元素，则有：

$$A(v+w) = Av + Aw = 0 + 0 = 0,$$

$$A(cv) = cAv = 0,$$

因此得证 $N(A)$ 确实是 \mathbf{R}^n 空间的一个子空间。

b 值的影响 Other values of b

若方程变为

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

则其解集不能构成一个子空间。零向量并不在这个集合内。解集是空间 \mathbf{R}^3 内过

$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ 和 $\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$ 的一个平面，但是并不穿过原点 $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ 。

本讲给出了关于矩阵的两种子空间，同时给出了两种构造子空间的方法。对于列空间，它是由列向量进行线性组合张成的空间；而零空间是从方程组出发，通过让 x 满足特定条件而得到的子空间。

第 07 讲 求解 $Ax=0$: 主变量, 特解

Solving $Ax=0$: pivot variables, special solutions

我们定义了矩阵的列空间和零空间, 那么如何求得这些子空间呢? 本节课的内容即从定义转到算法。

计算零空间 Nullspace

矩阵 A 的零空间即满足 $Ax=0$ 的所有 x 构成的向量空间。

$$\text{取 } A = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{bmatrix}$$

(A 的列空间并不是线性无关的。) 无论矩阵 A 是否可逆, 我们都采用消元法作为计算零空间的算法。

对于矩阵 A 进行“行操作”并不会改变 $Ax=b$ 的解, 因此也不会改变零空间。(但是会改变列空间。) 此处不需要应用增广矩阵, 因为等号右侧的向量 $b=0$ 。

第一步消元得到:

$$A = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{bmatrix} \rightarrow \begin{bmatrix} \boxed{1} & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 2 & 4 \end{bmatrix}$$

第二列没有主元, 因此主元二是第二行第三列的 2。

$$\begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 2 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} \boxed{1} & 2 & 2 & 2 \\ 0 & 0 & \boxed{2} & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} = U$$

主元列 $r=2$

自由列 $n-r=4-2$

矩阵 U 为梯形矩阵。其第三行变为零, 是因为第三行的行向量本身就是第一行和第二行行向量的线性组合。

矩阵的秩 (rank) 就是矩阵的主元的个数。本例中矩阵 A 和 U 的秩均为 2。矩阵中包含主元的列为主元列 (pivot column), 不包含主元的列称为自由列 (free column)。

特解 Special solutions

当我们将系数矩阵变换为上三角阵 U 时, 就可以用回代求得方程 $Ux=0$ 的解。本例中, 包含主元的矩阵第 1 列和第 3 列为主元列, 而不包含主元的第 2 列和第 4 列为自由列。对自由变量 (free variable) x_2 和 x_4 我们可以进行赋值。例如令 $x_2=1$

而 $x_4=0$ 。则有

$$2x_3 + 4x_4 = 0 \Rightarrow x_3 = 0$$

$$x_1 + 2x_2 + 2x_3 + 2x_4 = 0 \Rightarrow x_1 = -1$$

因此可得一解 $\mathbf{x} = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ ，其任意倍数均在矩阵的零空间之内。

取自由变量中 $x_2=0$ 而 $x_4=1$ ，则可得到另一解 $\mathbf{x} = \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}$

矩阵 \mathbf{A} 的零空间就是这些“特解”向量的线性组合所构成的向量空间。

矩阵的秩 r 等于其主元列的数目，因此自由列的数目就等于 $n-r$ ，即列的数目减去主元列的数目。这个数值等于特解的数目和零空间的维数。

主元列和自由列的一个重要区别就是，自由列可以表示为其左侧所有主元列的线性组合，而主元列则不可以。

$$\mathbf{U} = \begin{bmatrix} \boxed{\square} & * & * & * & * & * & * & * \\ 0 & 0 & \boxed{\square} & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & \boxed{\square} & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{\square} \end{bmatrix}$$

例如，我们得到一个消元完成后的梯形矩阵 \mathbf{U} ，其包含四个主元列。观察它的第五列，这是自由列，其左侧有两个主元列，这两个主元列显然线性无关，第五列也显然可以写成前两个主元列的线性组合。这里求的是 $\mathbf{U}\mathbf{x}=\mathbf{0}$ 的解，如果令对应第二列，第四列，第六列这三个自由列以及第五列右侧的两个主元列的 x 分量都为 0，而对应第五列的自由变量 $x_5=1$ ，则方程变为：

$$x_1 \begin{bmatrix} \square \\ 0 \\ 0 \\ 0 \end{bmatrix} + 0 + x_3 \begin{bmatrix} * \\ \square \\ 0 \\ 0 \end{bmatrix} + 0 + 1 \begin{bmatrix} * \\ * \\ 0 \\ 0 \end{bmatrix} + 0 + 0 + 0 = 0$$

相当于求第五列如何用前两个主元列进行线性组合，所得解 $\mathbf{x}^T = [x_1, 0, x_3, 0, 1, 0, 0, 0]$ 即为原方程的特解之一。对所有的自由列都进行此操作，就是上文求解过程中对自由变量的赋值过程。在本例中，四个自由变量分别取 1 会得到零空间的四个特解。如果把自由变量都赋值为 0 会怎么样？答案是求得的解为 $\mathbf{0}$ 向量。

行最简阶梯矩阵 Reduced row echelon form (rref)

通过继续消元我们可以将矩阵 \mathbf{U} 转变为行最简阶梯矩阵形式 \mathbf{R} 其中主元为 1，

而主元列除主元外皆为 0。在 Matlab 中用命令 $\text{rref}(\mathbf{A})$ 实现这一过程。

$$\mathbf{U} = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 0 & -2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} \boxed{1} & 2 & 0 & -2 \\ 0 & 0 & \boxed{1} & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathbf{R}$$

在矩阵中主元行和主元列的交汇处存在一个单位阵。通过“列交换”，可以将矩阵 \mathbf{R} 中的主元列集中在左侧，从而在左上角形成这个单位阵，而将自由列集中在矩阵的右侧。如果矩阵 \mathbf{A} 中的某些行是线性相关的，则在矩阵 \mathbf{R} 的下半部分就会出现一些完全为 $\mathbf{0}$ 的行向量。

$$\mathbf{R} = \begin{bmatrix} \mathbf{I} & \mathbf{F} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{array}{l} r \text{ 主元行} \\ m-r \text{ } \mathbf{0} \text{ 向量行} \end{array}$$

$r \text{ 主元列} \quad n-r \text{ 自由列}$

这里的 \mathbf{I} 是一个 $r \times r$ 的方阵。 \mathbf{F} 即自由列消元后组成的部分。

原方程 $\mathbf{Ax}=\mathbf{0}$ 变为求解 \mathbf{R} 的主元行乘以 \mathbf{x} ， $[\mathbf{I} \quad \mathbf{F}] \begin{bmatrix} x_{pivot} \\ x_{free} \end{bmatrix} = \mathbf{0}$ 。我们将 $\mathbf{Ax}=\mathbf{0}$

的特解作为列向量写成一个矩阵 \mathbf{N} ，即零空间矩阵。则其形式为 $\mathbf{N} = \begin{bmatrix} \mathbf{F} \\ \mathbf{I} \end{bmatrix}$ 。这里的 \mathbf{I}

为一个 $n-r \times n-r$ 的矩阵，就是对 $n-r$ 个自由变量分别赋值为 1 所构造出来的，零空间矩阵满足 $\mathbf{RN}=\mathbf{0}$ ， $\mathbf{0}$ 矩阵是一个 $m \times (n-r)$ 的矩阵。则从矩阵分块乘法运算可知零空间矩阵上半部分为 $-\mathbf{F}$ ，即 \mathbf{N} 最终形式为

$$\mathbf{N} = \begin{bmatrix} -\mathbf{F} \\ \mathbf{I} \end{bmatrix} \begin{array}{l} r \text{ 主变量} \\ n-r \text{ 自由变量} \end{array}$$

对于矩阵 \mathbf{R} 而言，求零空间特解就变得非常简单，只需要将消元的到的 \mathbf{F} 部分拼接上单位阵就可以得到所有的通解。注意如果在变换出 \mathbf{R} 左上角的单位阵的过程中采用了列交换，则最后的解要完成逆变换。

第 08 讲 求解 $Ax=b$: 可解性与结构

Complete Solution of $Ax=b$

可解的条件 Solvability conditions on b

我们仍取 $A = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{bmatrix}$ 为例, 则方程为 $\begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ 。

矩阵 A 的第三行为第一行和第二行的加和, 因此 $Ax=b$ 中 b 的第 3 个分量也要等于其第 1 和第 2 个分量的和。若 b 不满足 $b_3=b_1+b_2$ 则方程组无解。

检验 $Ax=b$ 是否可解的方法是对增广矩阵进行行消元。如果矩阵 A 的行被完全消去的话, 则对应的 b 的分量也要得 0。在本例中, 矩阵 A 的第三行被消去:

$$\begin{bmatrix} 1 & 2 & 2 & 2 & b_1 \\ 2 & 4 & 6 & 8 & b_2 \\ 3 & 6 & 8 & 10 & b_3 \end{bmatrix} \rightarrow \dots \rightarrow \begin{bmatrix} 1 & 2 & 2 & 2 & b_1 \\ 0 & 0 & 2 & 4 & b_2 - 2b_1 \\ 0 & 0 & 0 & 0 & b_3 - b_2 - b_1 \end{bmatrix}$$

如果 $Ax=b$ 有解, 则 $b_3-b_1-b_2=0$ 。在本例中我们令 $b = \begin{bmatrix} 1 \\ 5 \\ 6 \end{bmatrix}$

前几讲讨论过, 只有当 b 处于矩阵的列空间 $C(A)$ 之中时, 方程才有解。本讲推导出矩阵 A 的行向量若经过线性组合成为了零向量, 则对应的 b 经同样的线性组合后也要等于 0。因此看起来我们有了两条关于 b 的限制条件, 但实际上这两点是等价的。

通解 Complete solution

为求得 $Ax=b$ 的所有解, 我们首先检验方程是否可解, 然后找到一个特解。将特解和矩阵零空间的向量相加即为方程的通解。

特解 A particular solution

求 $Ax=b$ 特解的方法是将自由变量均赋值为 0, 求解其主变量。

本例中, 令 $x_2=x_4=0$ 得到方程组:

$$\begin{aligned} x_1 + 2x_3 &= 1 \\ 2x_3 &= 3 \end{aligned}$$

可解得 $x_3=3/2$, $x_1=-2$,

因此特解为 $\mathbf{x}_p = \begin{bmatrix} -2 \\ 0 \\ 3/2 \\ 0 \end{bmatrix}$ 。

上一讲说了主元列和自由列的一个重要区别就是，自由列可以表示为其左侧所有主元列的线性组合，而主元列则不可以，主元列是线性无关的。

$$U = \begin{bmatrix} \boxed{1} & * & * & * & * & * & * & * \\ 0 & 0 & \boxed{1} & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & \boxed{1} & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} \end{bmatrix}$$

我们仍以这个消元完成后的梯形矩阵 U 为例，其包含四个主元列。对于 $A\mathbf{x}=\mathbf{b}$ 的求解转变为 $U\mathbf{x}=\mathbf{c}$ ，其中 \mathbf{c} 是向量 \mathbf{b} 经过与左侧矩阵相同的行操作得到的向量。明显可以看到，此时四个主元列的线性组合可以组成任何 \mathbf{R}^4 中的向量，我们将 \mathbf{x} 中的自由变量赋值为 0 就可以去掉自由列列向量的干扰，求得方程的特解 \mathbf{x}_p 。如果消元得到 U 最后 i 行为 0，就要求 \mathbf{c} 的最后 i 个分量为 0，这时主元列才可以通过线性组合得到 \mathbf{c} ，否则无解。

与零空间进行线性组合 Combined with nullspace

$A\mathbf{x}=\mathbf{b}$ 的通解为 $\mathbf{x}_{\text{complete}}=\mathbf{x}_p+\mathbf{x}_n$ ，其中 \mathbf{x}_n 为矩阵零空间中的一般向量。将 $A\mathbf{x}_p=\mathbf{b}$ 和 $A\mathbf{x}_n=\mathbf{0}$ 相加可得 $A(\mathbf{x}_p+\mathbf{x}_n)=\mathbf{b}$ 。

上一讲我们得到了矩阵的零空间 $N(A)$ 就是其特解 $\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ 和 $\begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}$ 的线性组合的集

合。因此方程 $A\mathbf{x}=\begin{bmatrix} 1 \\ 5 \\ 6 \end{bmatrix}$ 的通解即为：

$$\mathbf{x}_{\text{complete}} = \begin{bmatrix} -2 \\ 0 \\ 3/2 \\ 0 \end{bmatrix} + c_1 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}$$

式中 c_1 和 c_2 为任意实数。

矩阵的零空间 $N(A)$ 是 \mathbf{R}^4 空间中的二维子空间，方程的解 $A\mathbf{x}=\mathbf{b}$ 构成了穿过 \mathbf{x}_p 点并和矩阵零空间平行的“平面”。但该“平面”并不是 \mathbf{R}^4 空间的子空间。

前面求取 $A\mathbf{x}=\mathbf{b}$ 的特解过程中，我们令所有自由变量赋值为 0。如果不赋值为 0，则等于带着自由列进行计算，但自由列其实也就是主元列的线性组合，这样求的

特解 \mathbf{x}_p' 只不过是 \mathbf{x}_p 与零空间特解的一个加和。如此例中若令 $x_2=1, x_4=0$ ，则求特

解的方程变为：
$$\begin{cases} x_1 + 2 + 2x_3 = 1 \\ 2x_3 = 3 \end{cases}$$
，求得 $x_1 = -4, x_3 = 3/2$ 。因此特解为 $\mathbf{x}_p' = \begin{bmatrix} -4 \\ 1 \\ 3/2 \\ 0 \end{bmatrix}$ 。而

实际上 $\mathbf{x}_p' = \mathbf{x}_p + \mathbf{x}_{n1} = \begin{bmatrix} -2 \\ 0 \\ 3/2 \\ 0 \end{bmatrix} + \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ 。

秩 Rank

矩阵的秩等于矩阵的主元数。如果 $m \times n$ 矩阵的秩为 r ，则必有 $r \leq m$ 且 $r \leq n$ 。

讨论满秩 (full rank) 的情形：

- 列满秩： $r=n$ 。每列都有主元， \mathbf{x} 的每一个分量都是主变量，没有自由变量。

零空间 $N(\mathbf{A})$ 之内只有零向量。方程无解或者有唯一解 \mathbf{x}_p 。

$$\text{例如 } \mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 1 \\ 6 & 1 \\ 5 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \mathbf{R}$$

- 行满秩： $r=m$ 。每行都有主元，无论 \mathbf{b} 取何值，方程 $\mathbf{Ax}=\mathbf{b}$ 都有解。主变量 r 个，自由变量 $n-r$ 个。

$$\text{例如 } \mathbf{A} = \begin{bmatrix} 1 & 2 & 6 & 5 \\ 3 & 1 & 1 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & * & * \\ 0 & 1 & * & * \end{bmatrix} = \mathbf{R}$$

- 满秩 $r=m=n$ ，矩阵可逆。零空间只有零向量，无论 \mathbf{b} 取何值，方程 $\mathbf{Ax}=\mathbf{b}$ 都有唯一解。

$$\text{例如 } \mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{R} \text{ 为单位阵。}$$

总结：

$r=m=n$	$r=n < m$	$r=m < n$	$r < n, r < m$
$\mathbf{R} = \mathbf{I}$	$\mathbf{R} = \begin{bmatrix} \mathbf{I} \\ 0 \end{bmatrix}$	$\mathbf{R} = [\mathbf{I} \quad \mathbf{F}]$	$\mathbf{R} = \begin{bmatrix} \mathbf{I} & \mathbf{F} \\ 0 & 0 \end{bmatrix}$
唯一解	无解或唯一解	无穷多解	无解或无穷多解

秩决定了方程组解的数量。

$m \times n$ 给出了矩阵的尺寸，但是秩 r 给出的是矩阵的实际“大小”。关于这个随后会有讨论。

第 09 讲 线性无关，基和维数

Independence, basis, and dimension

向量的线性无关意味着什么？如何用线性无关的概念来帮助我们描述包括零空间在内的子空间。

线性无关 Independence

矩阵 A 为 $m \times n$ 矩阵，其中 $m < n$ （因此 $Ax = b$ 中未知数个数多于方程数）。则 A 中具有至少一个自由变量，那么 $Ax = 0$ 一定具有非零解。 A 的列向量可以线性组合得到零向量，所以 A 的列向量是线性相关的。

若 $c_1x_1 + c_2x_2 + \dots + c_nx_n = 0$ 仅在 $c_1 = c_2 = \dots = c_n = 0$ 时才成立，则称向量 x_1, x_2, \dots, x_n 是线性无关的。若这些向量作为列向量构成矩阵 A ，则方程 $Ax = 0$ 只有零解 $x = 0$ ，或称矩阵 A 的零空间只有零向量。换言之，若存在非零向量 c ，使得 $Ac = 0$ ，则这个矩阵 A 的列向量线性相关。

在 \mathbb{R}^2 中，两个向量只要不在一条直线上就是线性无关的。（在 \mathbb{R}^3 中，三个向量线性无关的条件是它们不在一个平面上。）若选定空间 \mathbb{R}^2 中的三个向量，则他们必然是线性相关的。例如，如下的三个向量 v_1, v_2 和 v_3 是线性相关的。

$$\begin{matrix} & v_1 & v_2 & v_3 \\ A = & \begin{bmatrix} 2 & 1 & 2.5 \\ 1 & 2 & -1 \end{bmatrix} \end{matrix}$$

由我们研究方程组得到的结论，此矩阵构成的方程 $Ax = 0$ 必有非零解，即三个向量线性相关。

如果矩阵 A 的列向量为线性无关，则 A 所有的列均为主元列，没有自由列，矩阵的秩为 n 。若 A 的列向量为线性相关，则矩阵的秩小于 n ，并且存在自由列。

张成空间 Spanning a space

当一个空间是由向量 v_1, v_2, \dots, v_k 的所有线性组合组成时，我们称这些向量张成了这个空间。例如矩阵的列向量张成了该矩阵的列空间。

如果向量 v_1, v_2, \dots, v_k 张成空间 S ，则 S 是包含这些向量的最小空间。

基与维数 Basis & Dimension

向量空间的基是具有如下两个性质的一组向量 v_1, v_2, \dots, v_d ：

- v_1, v_2, \dots, v_d 线性无关
- v_1, v_2, \dots, v_d 张成该向量空间

空间的基告诉了我们空间的一切信息。

例： \mathbb{R}^3 空间

\mathbf{R}^3 空间有一组基 $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, 因为 $c_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ 只有零解。

并且这三个向量可以张成 \mathbf{R}^3 空间。

而 $\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 8 \end{bmatrix}$ 则不能构成一组基, 因为以它们为列向量组成的矩阵, 有两个相同的行, 消元肯定有自由列存在, 因此这三个向量并非线性无关。

GS 在这里犯了一个错误, 他以为这三个是线性无关的, 最后在第 10 讲中才进行了更正。当判定线性相关性时, 可以随时在矩阵、空间和方程组的概念之间切换, 哪个判据更容易判定就用哪个, 这里显然矩阵不可逆更容易看出来, 因为存在行向量重复的情况。从这里也可以看到行向量线性相关则列向量不可能线性无关, 矩阵行空间的维数等于列空间的维数。

若以 \mathbf{R}^n 空间中的 n 个向量为列向量构成的矩阵为可逆矩阵, 则这些向量可以构成 \mathbf{R}^n 空间中的一组基。

子空间的基 Basis for a subspace

向量 $\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 5 \end{bmatrix}$ 可以张成 \mathbf{R}^3 中的一个平面, 但是它们无法成为 \mathbf{R}^3 空间的一组基。

空间的每一组基都具有相同的向量数, 这个数值就是空间的维数 (dimension)。所以 \mathbf{R}^n 空间的每组基都包含 n 个向量。

列空间和零空间的基 Basis of a column space and nullspace

$$\text{取 } \mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{bmatrix}$$

讨论列空间: 矩阵 \mathbf{A} 的四个列向量张成了矩阵 \mathbf{A} 的列空间, 其中第 3 列和第 4 列与前两列线性相关, 而前两个列向量线性无关。因此前两列为主元列。他们组成了列空间 $C(\mathbf{A})$ 的一组基。矩阵的秩为 2。

实际上对于任何矩阵 \mathbf{A} 均有:

矩阵的秩 r = 矩阵主元列的数目 = 列空间的维数

注意: 矩阵具有秩 rank 而不是维数 dimension, 而空间有维数而不是秩。

当知道了列空间的维数, 可以从矩阵列向量中随意选取足够数量的线性无关的向量, 它们每一组都可以构成列空间的一组基。

讨论零空间: 本例中矩阵的列向量不是线性无关的, 因此其零空间 $N(\mathbf{A})$ 不止包

含零向量。因为可以看出第 3 列是第 1 列和第 2 列的加和。所以向量 $\begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix}$ 必然在零

空间 $N(\mathbf{A})$ 之内。同样还可以对 x_4 赋值为 1，从而得到 $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ 也在零空间之内。它们就

是 $\mathbf{Ax}=\mathbf{0}$ 的两个特解。

零空间的维数=自由列的数目= $n-r$ ，因此本例中 $N(\mathbf{A})$ 的维数为 $4-2=2$ 。这两个特解就构成了零空间的一组基。

第 10 讲 四个基本子空间

The four fundamental subspaces

本讲讨论矩阵的四个基本子空间以及他们之间的关系。

四个子空间 Four subspaces

任意的 $m \times n$ 矩阵 A 都定义了四个子空间。

列空间 Column space $C(A)$

矩阵 A 的列空间是 A 的列向量的线性组合在 \mathbf{R}^m 空间中构成的子空间。

零空间 Nullspace $N(A)$

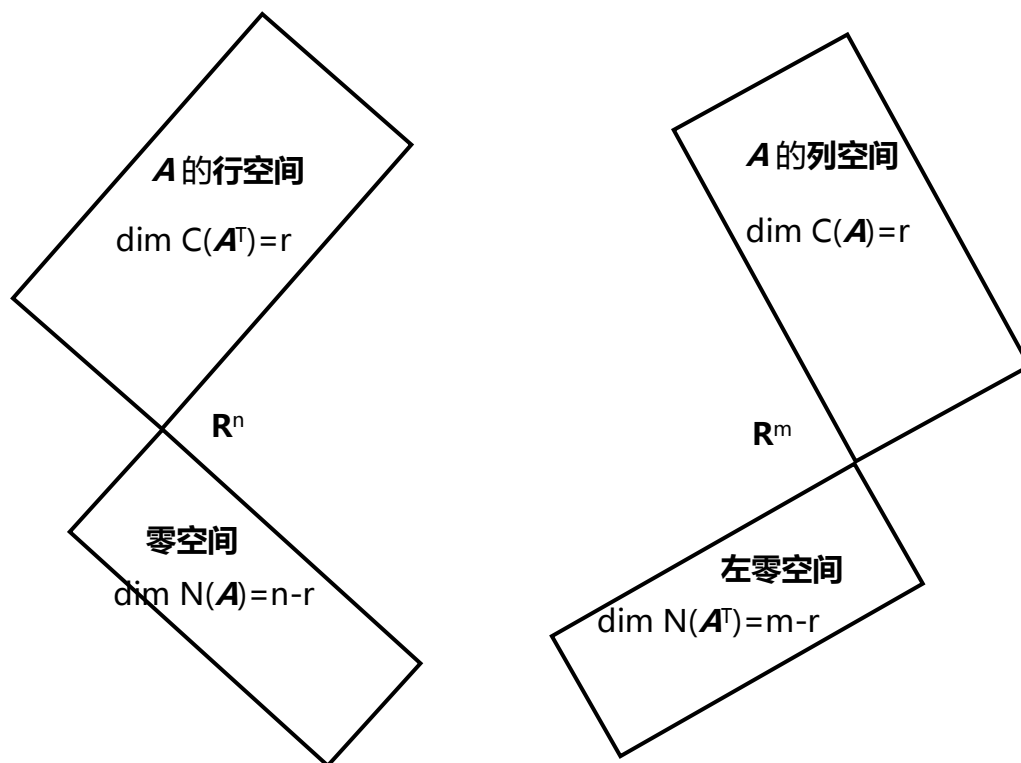
矩阵 A 的零空间是 $Ax=0$ 的所有解 x 在 \mathbf{R}^n 空间中构成的子空间。

行空间 Row space $C(A^T)$

矩阵 A 的行空间是 A 的行向量的线性组合在 \mathbf{R}^n 空间中构成的子空间，也就是矩阵 A^T 的列空间。

左零空间 Left nullspace $N(A^T)$

我们称矩阵 A^T 的零空间为矩阵 A 的左零空间，它是 \mathbf{R}^m 空间中的子空间。



基和维数 Basis & Dimension

列空间

矩阵 A 的 r 个主元列构成了列空间 $C(A)$ 的一组基。 $\dim C(A)=r$

零空间

$Ax=0$ 的一组特解对应于矩阵 A 的 $n-r$ 个自由列，并构成了零空间的一组基。

$\dim N(A)=n-r$

行空间

我们用矩阵 A 的化简的行阶梯矩阵 R 。

$$A = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{bmatrix} \rightarrow \dots \rightarrow \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix} = R$$

尽管矩阵 A 和矩阵 R 的列空间不同，但两者行空间相同。 R 的行向量来自于 A 的行向量的线性组合，因为消元操作是可逆的，所以 A 的向量也可以表示为 R 行向量的线性组合。

R 的前 r 行阶梯型“行向量”就是矩阵 A 行空间 $C(A^T)$ 的一组基。 $\dim C(A^T)=r$

左零空间

矩阵 A^T 有 m 列，而其秩为 r ，因此其自由列数目为 $m-r$ 。所以 $\dim N(A^T)=m-r$ 。

左零矩阵是满足 $A^T y = 0$ 的所有向量 y 的集合。称之为左零矩阵是因为该式可写作 $y^T A = 0$ ，而 y 出现在矩阵 A 左侧。

为找到左零空间的基，我们应用增广矩阵：

$$\begin{bmatrix} A_{m \times n} & I_{m \times n} \end{bmatrix} \rightarrow \begin{bmatrix} R_{m \times n} & E_{m \times n} \end{bmatrix}$$

我们将 A 通过消元得到矩阵 R ，其消元矩阵记为 E ，即 $EA=R$ 。若 A 为方阵，且 $R=I$ ，则有 $E=A^{-1}$ 。在本例中

$$EA = \begin{bmatrix} -1 & 2 & 0 \\ 1 & -1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = R$$

以“行操作”的观点来看矩阵 E 和 A 的乘法，则矩阵 E 最下面的 $m-r$ 个行向量使得矩阵 A 的行向量线性组合成为 0 ，也就是矩阵 R 最下面的 $m-r$ 个零向量。本例中， $m-r=1$ 。

矩阵 E 的这 $m-r$ 个行向量满足 $y^T A = 0$ ，它组成了矩阵 A 左零空间的一组基。

新向量空间 New vector space

所有 3×3 矩阵构成的集合是一个向量空间，符合对于线性运算封闭，称之为 M 。

M 的子空间包括：

- 所有的上三角阵

- 所有的对称阵
- 所有的对角阵

对角阵是前两个子空间的交集，其维数为 3，具有以下一组基：

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

当最初告诉我说，矩阵的列秩等于主元数，并且主元列构成了列空间的一组基时，其实我是拒绝的。主元这个东西不是行变换消元得来的么，消元过程列空间不是已经改变了么，为什么所得出 U 的主元数和主元列的位置还能够反映出矩阵 A 列空间的状态呢？

这里需要说明的是两点，其一是关于秩的定义有很多在数学上等价但是描述差异很大的说法，在这里我们把“秩”理解为行（列）向量中最大的线性无关向量组的向量数即可，在矩阵 A 行变换消元成梯形阵后，很容易看到行空间内极大无关组之一就是主元所在的那前 r 行，这 r 个行向量可以张成行空间，因此行空间的维数与主元数相等都是 r ，并且前 r 行构成了行空间的一组基。

但是为什么列空间的维数也是 r ，并且主元列可以构成列空间的一组基呢？这就是要说明的第二点，初等行变换不会改变列向量的线性相关性。为了叙述方便起见，我们假定矩阵 A 列向量的极大无关组就是 A 前 r' 列的向量（若否可以通过列交换而达成，列交换不会改变线性关系）。则矩阵 A 经过行消元得到梯形矩阵 U 的过程，可以用左乘可逆矩阵 E 来实现，即 $EA=U$ 。对等式左侧的矩阵乘法采用列操作的形式进行理解， $E[a_1, a_2, \dots, a_n] = U$ ，其中 a_i 为矩阵 A 的第 i 列。则 U 的前 r' 列分别为 $Ea_1, \dots, Ea_{r'}$ 。第一步，证明之前线性无关的列向量，行变换之后还无关：反设这些列向量线性相关，即存在非零 $c_1, c_2, \dots, c_{r'}$ 使方程 $c_1 Ea_1 + c_2 Ea_2 + \dots + c_{r'} Ea_{r'} = 0$ ，等式等价于 $E(c_1 a_1 + \dots + c_{r'} a_{r'}) = 0$ ，等式两端同时乘以 E^{-1} ，则有 $c_1 a_1 + c_2 a_2 + \dots + c_{r'} a_{r'} = 0$ ，但已知前 r' 列的向量是线性无关的，所以 $c=0$ ，矛盾，因此经过左乘可逆矩阵 E ，即“行变换”，不会改变线性无关的状态。第二步证明类似，就是证明之前线性相关的还相关，没有生成新的线性无关项，此处略去。综上可知，经过行变换列向量极大无关组的向量数仍为 r' ，即 U 的列空间维度为 r' 。而从 U 的矩阵形态可知，它的 r 个主元列线性无关，其列空间的维度为 r ，因此有 $r'=r$ ，并且主元列经过逆变换回去得到 A 的主元列也不会改变线性无关的关系，因此 A 的主元列就是 A 列空间的基。

我们大致说明了一件事情，就是矩阵的行秩等于其列秩。如果将矩阵 A 消元得到 rref， $PA=R$ ，其中 $R = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$ ，对 R 再进行列变换，可得 $PAQ=RQ = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$ ，最后凹出的这个造型被称为矩阵 A 的相抵标准型，它和 A 具有相同的秩，具有同构

的行空间和列空间。你可以视这个标准型为“初代”，在对它进行了可逆的行变换或者列变换之后得到了一系列矩阵，其中就包括老 **A**，这个初代的行列空间维数同为 r ，所以它“繁衍”出来的一系列矩阵的行列空间维数就同为 r ，所有的矩阵都有它的相抵标准型，所有的矩阵行秩都等于列秩。所以之前我们说过 m 和 n 给出了矩阵的尺寸，但是 r 给出了矩阵的“大小”，它告诉你矩阵行向量和列向量中真正“管用的”有几个，它告诉你方程组里管用的方程有几个，解集里面不能随便取的变量是几个，能随便赋值的又是哪几个。

再扯几句淡，如果去网上搜索“矩阵的行秩等于列秩究竟意味着什么”，那么你会发现很多看上去完完全全不同的说法，甚至完全不像是在讨论一个问题。之所以有这种感受，其中一个原因是对于秩的定义出发点不同，比如从行列式开始讲起的线代是用行列式来定义“秩”的；而另一个原因是，大家是在线代的不同层面在讨论这件事情。总体而言，GS 所讲的这套线代，是偏重于应用数学这个方向，希望通过深入理解矩阵运算，在工程应用的方面能走得更远，后面他引入马尔科夫链、傅里叶、微分方程的内容都是朝着这个方向努力，还有他开设的“计算科学与工程”的课程也是秉持这个理念；与之不同，还有一种讲解线性代数的思路是强调线性算子的核心作用，实际上它是在一个更加抽象的层面来进行讨论和推演，Sheldon Axler 的《Linear algebra done right》（中译本：线性代数应该这样学）就是以这个出发点来写作的书，引领我们走向线代更深入的部分，方便与其他数学分支建立联系。

其实数学的学习过程中，我们一直在做着“抽象”这件工作，我们从 10 个苹果、2 头牛这种实际事物中，抽象出数字的概念放下了实物的概念，随后我们抽象出未知数 x, y 而放下了具体数字的概念，再后来到函数连 x, y 的系数都已经被字母替代，到了微积分直接讨论函数之间的联系已经可以暂时放下函数的解析式，再到泛函……每前进一步我们都用一个更加抽象的概念替代一族具象的概念，使得一切操作变得更加简洁（但不一定就好懂），而结果更富普遍意义，但是对于理解能力就有更高的要求。

把话说回来，GS 讲的线性代数更 low 么（有人说 GS 的所讲的不是线性代数的核心而是 matlab 的核心-_-!）？显然不是的，他将线代的触角已经延伸出去很广的范围，在应用的领域做了非常细致有效的工作，开启了科学家和工程师对线代的正确理解方式。打个比方，就好像外门与内功，GS 的线代有如老顽童评价的“降龙十八掌”——已经是外门功夫的顶尖了，郭靖一招“亢龙有悔”基本上 S 级以下的出场人物分分钟打跑。反观张无忌身负九阳神通，被灭绝在昆仑山景区当众虐菜。都开着主角光环怎么差距就这么大呢！以我工科学渣的身份，估计内功真是了解一下就好了，把外功练到极致可能是更好的选择。

第 11 讲矩阵空间、秩 1 矩阵和小世界图

Matrix Spaces ; Rank 1 ; Small World Graphs

扩展一下向量空间的含义。

新的向量空间 New vector spaces

3×3 矩阵空间 3 by 3 matrices

空间 **M** 是所有 3×3 矩阵所构成的空间，**M** 的部分子空间包括：

- 所有的上三角阵，记为 **U**。
- 所有的对称阵，记为 **S**。
- 所有的对角阵，记为 **D**，它是前两个子空间的交集。

空间 **M** 的维数为 9，与 \mathbf{R}^9 空间很类似。我们可以选定它的一组基：

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

对角阵构成的子空间 **S** 维数为 6，它的一组基为：

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

上三角阵构成的子空间 **U** 维数也为 6，它的一组基为：

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

对角阵构成的子空间 **D** 维数为 3，可以选定 **S** 和 **U** 的基的交集为 **D** 的基。

S 和 **U** 的并集，即 3×3 矩阵中或为上三角阵或为对称阵的矩阵，构成 **M** 的子空间么？答案是否定的，这就如同在 \mathbf{R}^2 空间中找出两条直线，询问它们的并集是否构成一个子空间。如果我们将 **S** 和 **U** 中所有元素可能构成的加和作为一个集合，可以称为和集 **S+U**，它是 **M** 的一个子空间。实际上 **S+U** 就是 **M** 本身，其维数为 9。

$$\dim \mathbf{S} + \dim \mathbf{U} = \dim \mathbf{S} \cap \mathbf{U} + \dim \mathbf{S} + \mathbf{U}$$

此处网上的笔记有误，等式右侧应是和集的维数与交集的维数相加。

微分方程 Differential equations

对于给定的微分方程 $\frac{d^2 y}{dx^2} + y = 0$ ，求解该方程可以视为求它的零空间。我们可以得到的解为： $y = \cos x$ ， $y = \sin x$ ， $y = e^{ix}$ 。

事实上，通解为 $y = c_1 \cos x + c_2 \sin x$ ，其中 c 可以取任意复数。也将解的线性组合构成的空间称为解空间，其维数为 2。 $\cos x$ 和 $\sin x$ 可以成为解空间的一组基。这

些并不像是向量，它们是函数，但是可以对其进行线性运算，在线性代数的范畴内讨论之。

秩 1 矩阵 Rank one matrices

矩阵 $A = \begin{bmatrix} 1 & 4 & 5 \\ 2 & 8 & 10 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 4 & 5 \end{bmatrix}$ ，矩阵的秩为 1。 $\dim C(A) = r = \dim C(A^T)$

所有的秩 1 矩阵都可分解 $A = UV^T$ ，其中 U 和 V 都是列向量。秩 1 矩阵的行列式和特征值都很简单，它可以被当作是构建其他矩阵的“积木”。**其实我们在矩阵乘法的第四种形式里面见过它的作用。**如果存在一个 5×17 的矩阵 M ，而其秩为 4，那么它可以由 4 个秩 1 矩阵组合而成。

若矩阵空间 M 为所有的 5×17 矩阵，那么 M 中所有的秩 4 矩阵所构成的集合是一个子空间么？答案是否定的，即使加入零矩阵也无法构成子空间，对于两个矩阵的加和，秩 4 矩阵集合并不封闭。

两个矩阵之和的秩小于等于两个矩阵的秩之和。

在 R^4 空间中，所有满足分量之和 $v_1 + v_2 + v_3 + v_4 = 0$ 的向量 $v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$ 构成了一个

子空间 S ，包含零向量并对线性运算封闭。它是矩阵 $A = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$ 的零空间，因为矩阵 A 的秩为 1，因此其零空间的秩为 $n - r = 3$ 。零空间的基就是 $Ax = 0$ 特解：

$$\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}。$$

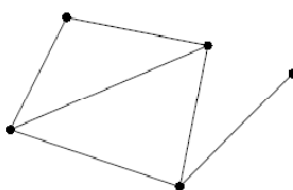
本例的一个意义就是寻找子空间维数的逆向思路，可以考虑它是不是某个方程的解空间，在这里它是矩阵 $A = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$ 的零空间，我们可以从矩阵推出这个空间的维数。

矩阵 A 的列空间为 R^1 。左零空间仅包含零向量，维数为 0。

小世界图 Small world graphs

介绍小世界图主要是引出图论和线性代数的联系。

在这里，“图” G 是结点和边的集合： $G = \{ \text{结点 (nodes)}, \text{边 (edges)} \}$



此图包含 5 个结点和 6 条边，我们可以利用一个 5×6 矩阵完全描述它。

我们可以用图来描述一个实际问题，如果每个人是一个结点，两个人互相认识为一个边，那么整个美国可以以此构成一张大图。我们可以通过这张图来确认两个人之间的最短距离是多少，即两个人需要通过最少几个朋友才能建立联系。G 本人和克林顿之间的距离为 2，他的一个朋友是参议员，他认识这个参议员朋友，那个人认识克林顿。班里的学生跟克林顿的距离因此不会大于 3。还可以继续算希拉里和莱温斯基……

所谓“六度分割理论”（six degrees of separation）猜想一个人和陌生人之间间隔的点不会超过六个。因此当陌生的两人聊起这种联系都会感叹：“世界真小啊！”这也是“小世界图”这个名字的由来。

尽管都在 $Ax=b$ 这个单元之下，但是第 11 和第 12 两讲的内容与其他几节课程明显有所差异。之前的课程偏重于介绍线性代数的基础知识，主要是将矩阵运算和消元法解线性方程联系起来，告诉我们矩阵不是你常规理解的用于存储的数组，它可以是一种操作或者是空间的缩影。而这两节课程告诉我们线性方程组也不是你常规理解的，从一个非常明显的线性关系抽象出来的方程，它可以是从微分方程或者从图论的角度建立起来的关系式，而应用我们在线性代数里面学到的理论，可以处理这些问题，甚至得到一些更加精深的理解。

在应用数学工具来解决实际问题的过程中有两个基本步骤：建模和解方程。可以说前面的课程主要在解方程这边，而最后两节课稍稍向建模那个方面靠近了一点。

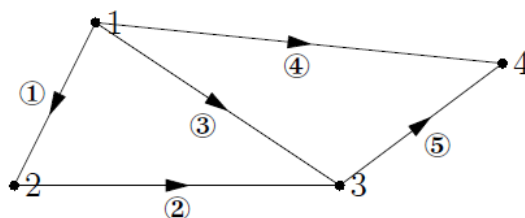
第 12 讲 图、网络、关联矩阵

Graphs , networks , incidence matrices

本讲讨论线性代数在物理系统中的应用。

图和网络 Graphs & Networks

“图” 就是 “结点” 和 “边” 的一个集合。



边线上的箭头代表从结点流出的正方向。

关联矩阵 (Incidence matrices)

构造一个矩阵来表示图的内在含义，此矩阵称为关联矩阵，图中每个结点代表一行，每边代表一行。则上图为 5×4 矩阵。反过来从这个矩阵出发我们也能画出图。

结点	1	2	3	4	
	-1	1	0	0	边 ①
	0	-1	1	0	②
$A =$	-1	0	1	0	③
	-1	0	0	1	④
	0	0	-1	1	⑤

第一行代表边①，从结点 1 流出记为-1，从结点 2 流入记为 1。

边①、边②和边③构成了一个回路 称为环 (loop)。反映在矩阵上是这三个行向量线性相关。

源于现实问题的关联矩阵，通常描述了问题的结构。如果我们研究一个很大的图，则会构建一个很大的矩阵，但这个矩阵会是稀疏矩阵。

考察矩阵的零空间 即求 $Ax=0$ 的解。零空间告诉我们列向量线性组合的状态。

$$Ax = \begin{bmatrix} x_2 - x_1 \\ x_3 - x_2 \\ x_3 - x_1 \\ x_4 - x_1 \\ x_4 - x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

如果 x 为结点上的电势，则 Ax 给出了每个边上的电势差。求解可以得到零空

间为一维 $\dim N(A)=1$ ，它的基就是 $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ ，解集则是 $x=c \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ ，代表等电势，说明等

电势条件下不会有电流产生。常数 c 的确定需要边界条件 比如我们将结点 4 接地，

则 $x_4=0$ 。

若求 $Ax=b$ 的解，则相当于在给定了电压 b 的情况下，求各点的电势，但实际上我们得不到电势的准确值，因为零空间有常数解 c ，各点得到的电势需要加上常数 c ，这很类似于求积分要加上常函数，常数值需要边界条件来确定。

矩阵的列数为 4，而其零空间的维数为 1，则矩阵的秩为 3，矩阵第 1 列，第 2 列，第 4 列的列向量线性无关。

考察矩阵列空间，一个重要的问题就是对于什么样的 b ， $Ax=b$ 有解。边①,边②和边③构成了环，这三个行向量线性相关，同样的情况还有边④,边⑤和边③构成的环。于是 b 的分量需要满足 $b_1-b_2+b_3=0$ 以及 $b_3-b_4+b_5=0$ 。如果把边①,边②，边④,边⑤构成的大环也表示出来则还可以得到一个等式，但实际上这个等式就是之前这两个等式的组合。这两个等式就是基尔霍夫电压定律 (Kirchhoff's Voltage law)，即环路电势差之和为零。

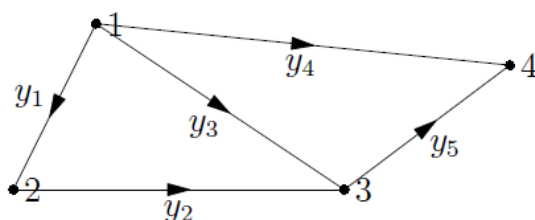
矩阵的左零空间是满足 $A^T y=0$ 的向量 y 的集合。因为矩阵 A^T 有 5 列，且矩阵的秩为 3，因此矩阵的左零空间维数为 2。这反应了行向量的线性关系，整个“图”中，环数为 2。

$$A^T y = \begin{bmatrix} -1 & 0 & -1 & -1 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}。y \text{ 的分量的值为“边”上的电流。}$$

在电势差和电流之间建立联系就是欧姆定律 (Ohm's Law)

$$\begin{array}{c} \text{电势} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \xrightarrow[\mathbf{e} = \mathbf{A}\mathbf{x}]{\text{矩阵 } \mathbf{A}} \text{电势差} \begin{bmatrix} x_2 - x_1 \\ x_3 - x_2 \\ x_3 - x_1 \\ x_4 - x_1 \\ x_4 - x_3 \end{bmatrix} \xrightarrow[\mathbf{y} = \mathbf{C}\mathbf{e}]{\text{矩阵 } \mathbf{C} \text{ 欧姆定律}} \text{电流} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \xrightarrow[\text{基尔霍夫定律}]{\text{矩阵 } \mathbf{A}^T} \mathbf{A}^T \mathbf{y} = 0 \end{array}$$

我们求解 $A^T y=0$ 就是在求 5 个满足基尔霍夫电流定律 (Kirchhoff's Law) 的电流值。



$$\begin{array}{rcl}
 -y_1 - y_3 - y_4 & = & 0 \\
 y_1 - y_2 & = & 0 \\
 y_2 + y_3 - y_5 & = & 0 \\
 y_4 + y_5 & = & 0
 \end{array}$$

$A^T \mathbf{y} = \mathbf{0}$ 的方程形式 , 每一个方程关于一个结点 , 方程表示

结点电流值为 0 , 即流入等于流出。

从图上解方程 , 而不是采用消元法解方程。如果我们设定 $y_1=1$, 并且让 y_1 ,

y_2 和 y_3 组成的回路的 “环流 ” 为 0 , 则有 $y_2=1$, $y_3=-1$ 。可解得 $\mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$ 。取另一

个回路的环流为 0 , 则有 $y_3=1$, $y_4=-1$, $y_5=1$ 。 $\mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}$ 。

如果设定 y_1 , y_2 , y_4 和 y_5 组成的大回路环流为 0 , 则可以得到另一个向量 \mathbf{y} , 而该向量在零空间内 , 是前两个向量的线性组合。

考察矩阵的行空间 , 因为矩阵 $r=3$, 所以存在 3 个线性无关的向量。第 1 行 , 第 2 行和第 4 行为线性无关 , 在 “图” 中 , 边①,边②和边④构成了一张小图 , 这三个边没有形成回路。线性相关问题等价于形成回路。没有回路的小图包含 4 个结点和 3 条边 , 再添加一条边就会产生回路 , 在矩阵里表现为在第 1 行 , 第 2 行和第 4 行之上再添加一个行向量就会变为线性相关。没有回路的图称为 “树”。

思考一下维数公式的在 “图” 中的意义 :

左零空间维数 $\dim N(A^T)=m-r$;

等价于 “环” 数量 = “边” 数量 - (“结点” 数量 - 1) ;

即 Euler 公式 : “结点” - “边” + “环” = 1。对所有图都成立。

矩阵的秩 $r = \text{“结点”} - 1$, 因为 r 表示了线性无关的边的数目 , 也就是 “树” 中 “边” 的数目。

$$\begin{array}{c}
 \text{电势} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \xrightarrow[\mathbf{e} = \mathbf{Ax}]{\text{矩阵 } \mathbf{A}} \text{电势差} \begin{bmatrix} x_2 - x_1 \\ x_3 - x_2 \\ x_3 - x_1 \\ x_4 - x_1 \\ x_4 - x_3 \end{bmatrix} \xrightarrow[\mathbf{y} = \mathbf{Ce}]{\text{矩阵 } \mathbf{C} \text{ 欧姆定律}} \text{电流} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \xrightarrow[\text{基尔霍夫定律}]{\text{矩阵 } \mathbf{A}^T} \mathbf{A}^T \mathbf{y} = \mathbf{f}
 \end{array}$$

之前的讨论都是针对于一个无源的电场 , 如果加入电源则情况又不同 , 例如加入电流源相当于将基尔霍夫定律的方程变为 $\mathbf{A}^T \mathbf{y} = \mathbf{f}$, \mathbf{f} 就是外部流入的电流。

将 $\mathbf{e} = \mathbf{Ax}$, $\mathbf{y} = \mathbf{Ce}$, $\mathbf{A}^T \mathbf{y} = \mathbf{f}$ 三个等式结合得到应用数学中的基本方程 $\mathbf{A}^T \mathbf{CAx} = \mathbf{f}$ 。

关于方程 $\mathbf{A}^T \mathbf{CAx} = \mathbf{f}$ 的更多内容可以阅读 GS 老先生 08 的书 “Computational science and engineering” 的第二章。

第 13 讲 复习一

Exam 1 review

本讲为考前复习课，考试范围就是 $Ax=b$ 这个单元，重点是长方形矩阵，与此相关的概念包括零空间、左零空间、秩、向量空间、子空间，特别是四个基本子空间。当矩阵为可逆的方阵时，很多性质是一目了然的，但是对于长方形矩阵则需要多加注意。

问题一

向量 u , v 和 w 是 \mathbf{R}^7 空间中的非零向量。它们张成了 \mathbf{R}^7 空间中的一个子空间，那么这个子空间的维数可能是多少？

解答：1, 2 或者 3。空间的基的个数不超过 3 个，所以其维数也不会超过 3。因为是非零向量，所以不能是 0。

问题二

给定矩阵 U 为 5×3 阶梯型矩阵，其秩 $r=3$ 。

GS 老师这里设定的矩阵 U 就是行最简阶梯型矩阵 R ，他在视频中说这是多年前卷子上的题目，那时候他都写 U 而不是 R 。而在 11 年的笔记中已经改为 R 了。此处我们与视频保持相同。

1. 求 U 的零空间 $N(U)$

解答：因为列数为 3 且秩为 3，其列向量线性无关，则 $Ux=0$ 只有零解，所以

其零空间 $N(U)=\{0\}=\left\{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right\}$

2. 令 $B=\begin{bmatrix} U \\ 2U \end{bmatrix}$ ，求 B 的行最简阶梯矩阵？

解答： $\begin{bmatrix} U \\ 0 \end{bmatrix}$

3. B 的秩？

解答：3。

4. 令 $C=\begin{bmatrix} U & U \\ U & 0 \end{bmatrix}$ ，求它的行最简阶梯矩阵？

解答：进行行消元， $\begin{bmatrix} U & U \\ U & 0 \end{bmatrix} \rightarrow \begin{bmatrix} U & U \\ 0 & -U \end{bmatrix} \rightarrow \begin{bmatrix} U & 0 \\ 0 & -U \end{bmatrix} \rightarrow \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix}$

最后还要做的就是将 R 中最后两行的零行向量下移到矩阵最下面。

5. C 的秩？

解答：6， B 的秩为3。

6.求 C 左零空间的维数 $\dim N(C^T)$?

解答： $m=10$ 而 $r=6$ ，所以 $\dim N(C^T)=4$ 。

问题三

$$Ax = \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix}, \text{ 其中 } x = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + c \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + d \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \text{ 为通解。}$$

1.矩阵 A 的形状？

解答：矩阵 A 为 3×3 矩阵，因为 x 的分量数即为 A 的列数，而 b 的分量数为 A 的行数。

2.矩阵 A 的行空间的维数？

解答：从通解的形式可以看出 A 零空间的维数为 2，则其行空间的维数为 $3-2=1$ 。

3.求矩阵 A ？

解答：代入特解 $A \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} = 2 \times (\text{列 } 1) = \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix}$ 可知， A 的第一列（列 1）= $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ ，代入零空间的特解 $A \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = (\text{列 } 1) + (\text{列 } 2) = 0$ ，可知（列 2）= $\begin{bmatrix} -1 \\ -2 \\ -1 \end{bmatrix}$ ，代入另一特解 $A \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = (\text{列 } 3) = 0$ ，可知 $A = \begin{bmatrix} 1 & -1 & 0 \\ 2 & -2 & 0 \\ 1 & -1 & 0 \end{bmatrix}$ 。

4.什么样的向量 b 使得方程 $Ax=b$ 有解 x ？

解答：当 b 在矩阵 A 的列空间内的时候，方程有解。本题中 b 需为 $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ 的倍数。

本题中矩阵的秩为 1，因此零空间很大。请注意系数矩阵满秩的情况，我们曾经花了不少时间进行讨论。

小问题

1. A 是方阵，零空间只有零向量 0 ， A^T 的零空间？

解答： A^T 的零空间也只有零向量。

2.所有的 5×5 可逆矩阵，是否构成 5×5 矩阵空间的子空间？

解答：不行，不包含零矩阵 0 ，不是子空间。顺便说一下奇异阵也不是一个子空间。

3.判断：如果 $B^2=O$ ，是否必有 $B=O$ ？

解答：否，反例 $B=\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ 。

4.判断：方程组 $Ax=b$ 具有 n 个方程 n 个未知数，若矩阵 A 的列向量线性无关，则 b 取任意向量，方程均有解。

解答：是， A 为可逆矩阵，所以 $x=A^{-1}b$ 是唯一解。

问题四

$$\text{矩阵 } B=CD=\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

1.给出矩阵 B 零空间的一组基？

解答：因为矩阵 C 为可逆矩阵，因此矩阵 B 零空间与矩阵 D 的零空间相同。

矩阵 D 已经是行最简阶梯矩阵了，可解得其零空间的一组基为 $\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}$ 和 $\begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$ 。

左乘可逆矩阵，相当于对行向量进行可逆的线性组合。既不改变行空间，也不改变零空间。

若 x 满足 $Bx=0$ ，则有 $C^{-1}Bx=0$ ，即 $Dx=0$ ， B 零空间中的向量均在 D 的零空间中；若 y 满足 $Dy=0$ ，则有 $CDy=0=By$ ，则 D 零空间中的向量均在 B 的零空间中，因此 B 零空间与矩阵 D 的零空间相同。

2.求 $Bx=\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ 的通解？

解答：向量 $b=\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ 与矩阵 C 第一列列向量完全相同，则问题变为 $Dx=\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ ，求

解方程得到一个特解为 $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ ，则其通解就是 $x=\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}+c\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}+d\begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$

小问题

1.判断： A 为方阵，则它的行空间和列空间相同？

解答：否，只是空间的维数相等，但空间不一定相同。反例 $B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ 。都是

\mathbb{R}^n 空间中的 r 维子空间，但是彼此不一定相同。对称矩阵的行空间和列空间相同。不是对称矩阵也可以两者相同，比如可逆矩阵，再比如可逆的 $(n-1) \times (n-1)$ 矩阵加上半圈 0 变成的 $n \times n$ 矩阵，加上一圈 0.....

2.判断：矩阵 A 和 $-A$ 的四个子空间相同？

解答：是，相同。

3.判断：如果 A 和 B 的四个子空间相同，则 A 一定是 B 的倍数？

解答：错误， A 和 B 的可以为同阶可逆方阵。

4.如果我们对矩阵 A 中的两行做行交换，四个子空间中不变的是哪些？

解答：是行空间和零空间。

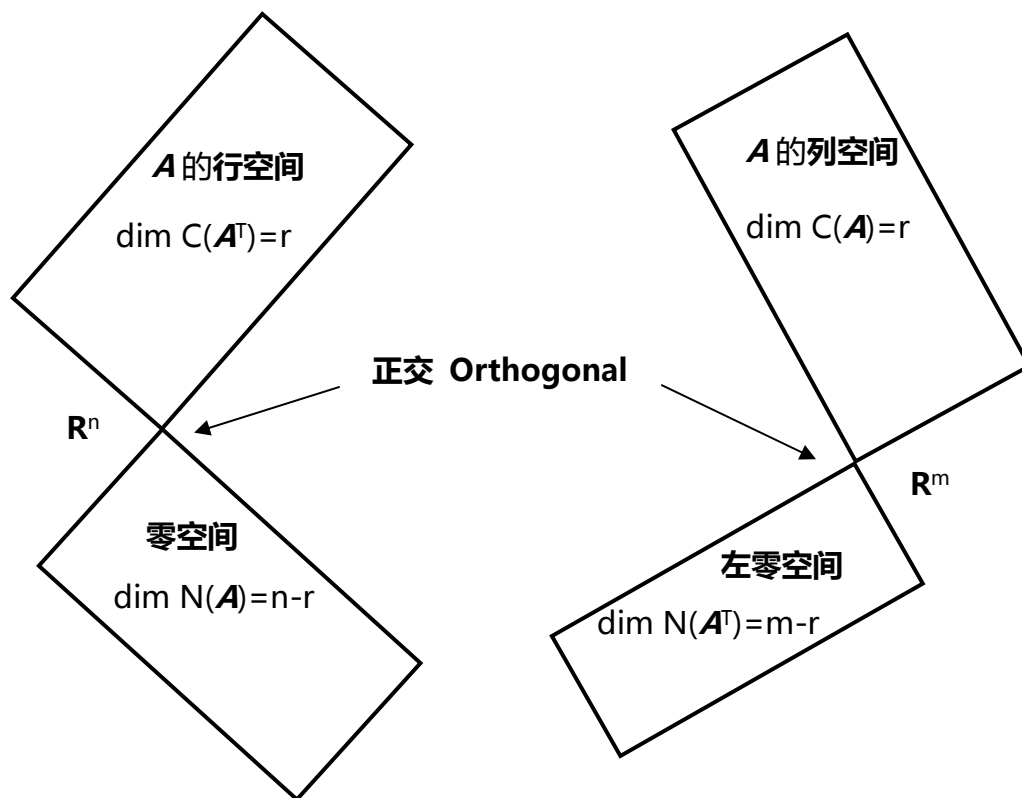
5.为什么向量 $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ 不可能既在矩阵 A 的零空间中，又成为 A 的一个行向量？

解答：反设成立。 \mathbf{v} 在零空间中，则有 $A\mathbf{v} = \mathbf{0}$ ，而它 \mathbf{v}^T 也是 A 的行向量，则 \mathbf{v} 和这个行向量 \mathbf{v}^T 做乘法得到数值为 14 而不是 0，与 $A\mathbf{v} = \mathbf{0}$ 矛盾。实际上 \mathbf{v} 若在零空间中，则根本不会再零空间中，在下一章中我们将看到矩阵的零空间和其行空间是正交的。

第 14 讲 正交向量与正交子空间

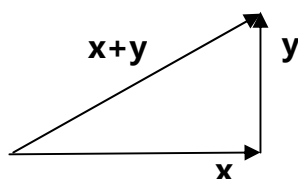
Orthogonal vectors & subspaces

本讲我们讨论正交 (orthogonal) 概念对于向量、基和子空间的意义。



图中绘制空间成 90 度角，这是表示这两个空间正交。这张图是 GS 最得意的作品之一，它反映了四个子空间的关系，在后面的课程中可以看到其两两形成正交补，在 \mathbb{R}^n 空间中的向量会向两个子空间射影，并向 \mathbb{R}^m 空间形成映射，反之亦然。

正交向量 Orthogonal vectors



正交就是垂直 (perpendicular) 的另一种说法。两向量正交的判据之一是其点积 $\mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x} = 0$ 。当两个向量的夹角为 90 度时，按照勾股定理 (毕达哥拉斯定理 Pythagorean theorem) \mathbf{x}, \mathbf{y} 满足：

$$\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 = \|\mathbf{x} + \mathbf{y}\|^2 \quad \text{其中 } \|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$$

$$\text{例如 } \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \text{ 则 } \mathbf{x} + \mathbf{y} = \begin{bmatrix} 3 \\ 1 \\ 3 \end{bmatrix}, \|\mathbf{x}\|^2 = 14, \|\mathbf{y}\|^2 = 5, \|\mathbf{x} + \mathbf{y}\|^2 = 19.$$

将勾股定理展开进行计算，则有 $\mathbf{x}^T\mathbf{x}+\mathbf{y}^T\mathbf{y}=(\mathbf{x}+\mathbf{y})^T(\mathbf{x}+\mathbf{y})=\mathbf{x}^T\mathbf{x}+\mathbf{y}^T\mathbf{y}+\mathbf{x}^T\mathbf{y}+\mathbf{y}^T\mathbf{x}$ 。得到 $2\mathbf{x}^T\mathbf{y}=0$ 。

零向量与所有向量都正交。

正交子空间 Orthogonal subspaces

子空间 S 与子空间 T 正交，则 S 中的任意一个向量都和 T 中的任意向量正交。黑板所在的平面和地板所在平面不是正交关系，沿两者的交线方向的向量同时属于两个平面，但并不与自己正交。

我们在平面内讨论正交子空间，平面的子空间包括只包含零向量的 $\mathbf{0}$ 空间、过原点的直线以及整个平面。经过原点的直线不会和整个空间正交； $\mathbf{0}$ 空间和过原点的直线正交；经过原点的两条直线若夹角为直角则互相正交。

零空间与行空间正交 Nullspace is perpendicular to row space

矩阵 A 的行空间和它的零空间正交。若 \mathbf{x} 在零空间内，则有 $A\mathbf{x}=\mathbf{0}$ ，将 A 表示为行向量的格式：

$$\begin{bmatrix} row_1 \\ row_2 \\ \vdots \\ row_m \end{bmatrix} \begin{bmatrix} \mathbf{x} \end{bmatrix} = \begin{bmatrix} row_1 \cdot \mathbf{x} \\ row_2 \cdot \mathbf{x} \\ \vdots \\ row_m \cdot \mathbf{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

\mathbf{x} 与矩阵 A 的行向量点积都等于 0，则它和矩阵 A 行向量的线性组合进行点积也为 0，所以 \mathbf{x} 与 A 的行空间正交。 \mathbf{x} 为零空间内的任意向量，所以零空间与行空间正交。同理可以证明列空间与左零空间正交。

行空间和零空间实际上把 \mathbf{R}^n 空间分割成了两个正交的子空间。例如对于矩阵：

$$A = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \end{bmatrix}$$

则其行空间是 1 维的，向量 $\begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$ 是它的基向量，而其零空间是垂直于 $\begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$ 并穿

过原点的 2 维平面。

行空间和零空间不仅仅是正交，并且其维数之和等于 n ，我们称行空间和零空间为 \mathbf{R}^n 空间内的正交补 (orthogonal complements)。这表示零空间包含所有和行空间正交的向量，反之亦然。想想我们之前提到的黑板和地板平面不是正交子空间的例子，二者都在 3 维空间中，分别为 2 维空间，因此不可能正交。一个空间中正交子空间的维数之和不可能超过原空间的维数。

我们可以称目前讨论的这部分内容是线性代数基本定理的第二部分。第一部分是给出四个子空间和它们的维数，第二部分说明它们是两两互为正交补，第三部分

讨论子空间的正交基。这些内容都反映在了本讲座开始的那幅图上。

矩阵 $A^T A$

下面讨论如何求解一个无解方程组 $Ax=b$ 的解。如果 A 是长方形矩阵， m 大于 n 。当左侧方程数特别多的时候，容易混入“坏”数据，方程变得无解。但是对于数据的可信度我们无从判断，线性代数要做的就是在这种条件下求一个方程的“最优解”。矩阵 $A^T A$ 会发挥重要作用，它是一个 $n \times n$ 方阵，并且是对称阵（ $A^T A^T = A^T A$ ）。

本章的核心内容就是当 $Ax=b$ 无解的时候，求解 $A^T A\hat{x} = A^T b$ 得到最优解。

$$\text{例：} A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix}, \text{ 则 } A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 8 & 30 \end{bmatrix} \text{ 是可逆矩阵。}$$

但是矩阵 $A^T A$ 并不总是可逆。

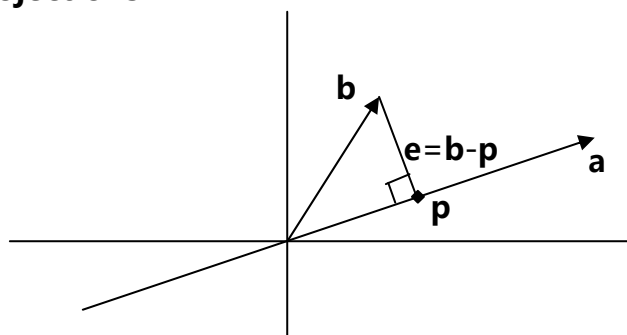
$$\text{例：} A = \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 3 \end{bmatrix}, \text{ 则 } A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 9 \\ 9 & 27 \end{bmatrix} \text{ 是不可逆矩阵。}$$

实际上 $N(A^T A) = N(A)$ ，并且矩阵 $A^T A$ 的秩等于 A 的秩。因此矩阵 $A^T A$ 可逆要求 A 的零空间只有零向量，即 A 的列向量线性无关。

第 15 讲 子空间投影

Projections onto subspaces

投影 (射影) Projections



投影问题的几何解释就是：如何在向量 \mathbf{a} 的方向上寻找与向量 \mathbf{b} 距离最近的一点。从图中可以看出，这个距离最近的点 \mathbf{p} 就位于穿过 \mathbf{b} 点并与向量 \mathbf{a} 正交的直线与向量 \mathbf{a} 所在直线的交点上。这就是 \mathbf{b} 在 \mathbf{a} 上的投影。如果我们将向量 \mathbf{p} 视为 \mathbf{b} 的一种近似，则长度 $\mathbf{e} = \mathbf{b} - \mathbf{p}$ 就是这一近似的误差。

因为 \mathbf{p} 在向量 \mathbf{a} 的方向上，因此可以令 $\mathbf{p} = x\mathbf{a}$ ，而因为它和 \mathbf{e} 正交，我们可以得到方程： $\mathbf{a}^T(\mathbf{b} - x\mathbf{a}) = 0$ 。

$$\text{解得：} x = \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}}, \quad \mathbf{p} = x\mathbf{a} = \mathbf{a} \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}}。$$

如果 \mathbf{b} 变为原来的 2 倍，则 \mathbf{p} 也变为原来的 2 倍。而如果 \mathbf{a} 变为原来的 2 倍， \mathbf{p} 不发生变化。从几何上和计算中都会得到验证。

本单元前半部分的核心内容就是射影。上一单元我们最核心的内容是认识消元法对于线性方程组的意义，并用矩阵的数学语言实现了消元过程，在那里最核心的策略就是利用矩阵乘法中的行操作来实现这一过程。这里面临类似的情况，我们有一个明确的几何目标，要将向量投影到已知子空间，而这里的策略就是误差向量和已知子空间正交，即两者求点积为 0。

投影矩阵 Projections matrix

我们将投影问题用投影矩阵的方式进行描述，即为 $\mathbf{p} = \mathbf{P}\mathbf{b}$ 其中 \mathbf{P} 为投影矩阵。

$$\mathbf{p} = x\mathbf{a} = \mathbf{a} \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}}。 \text{ 则有 } \mathbf{P} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T \mathbf{a}}, \text{ 其分子 } \mathbf{a}\mathbf{a}^T \text{ 是一个矩阵，而分母是一个数。}$$

观察这个矩阵可知，矩阵 \mathbf{P} 的列空间就是向量 \mathbf{a} 所在的直线，矩阵的秩是 1。投影矩阵 \mathbf{P} 是一个对称矩阵。另一方面，如果做两次投影则有 $\mathbf{P}^2 \mathbf{b} = \mathbf{P}\mathbf{b}$ ，这是因为第二次投影还在原来的位置。因此矩阵 \mathbf{P} 有如下性质： $\mathbf{P}^T = \mathbf{P}$ ， $\mathbf{P}^2 = \mathbf{P}$ 。

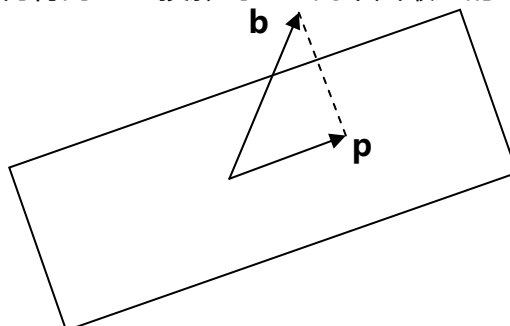
为什么要投影 Why Project

如前所述，方程 $\mathbf{Ax} = \mathbf{b}$ 有可能无解，我们需要得到方程的“最优解”。这里的问题

题在于向量 \mathbf{Ax} 一定在矩阵 \mathbf{A} 的列空间之内，但是 \mathbf{b} 不一定，因此我们希望将 \mathbf{b} 投影到 \mathbf{A} 的列空间得到 \mathbf{p} ，将问题转化为求解 $\mathbf{Ax}=\mathbf{p}$ 。

在高维投影 Projection in higher dimensions

在 \mathbf{R}^3 空间内，如何将向量 \mathbf{b} 投影到它距离平面最近的一点 \mathbf{p} ？



如果 \mathbf{a}_1 和 \mathbf{a}_2 构成了平面的一组基，则平面就是矩阵 $\mathbf{A}=[\mathbf{a}_1\mathbf{a}_2]$ 的列空间。

已知向量 \mathbf{p} 在平面内，则有 $\mathbf{p}=\hat{x}_1\mathbf{a}_1+\hat{x}_2\mathbf{a}_2=\mathbf{Ax}$ 。而 $\mathbf{e}=\mathbf{b}-\mathbf{p}=\mathbf{b}-\mathbf{Ax}$ 与投影平面正交(重点)，因此 \mathbf{e} 与 \mathbf{a}_1 和 \mathbf{a}_2 均正交，因此可以得到： $\mathbf{a}_1^T(\mathbf{b}-\mathbf{Ax})=0$ 并且 $\mathbf{a}_2^T(\mathbf{b}-\mathbf{Ax})=0$ 。因为 \mathbf{a}_1 和 \mathbf{a}_2 分别为矩阵 \mathbf{A} 的列向量，即 \mathbf{a}_1^T 和 \mathbf{a}_2^T 为矩阵 \mathbf{A}^T 的行向量，所以将两个方程式写成矩阵形式即为 $\mathbf{A}^T(\mathbf{b}-\mathbf{Ax})=0$ 。这与一维投影的方程形式相同。

向量 $\mathbf{e}=\mathbf{b}-\mathbf{Ax}$ 存在于矩阵 \mathbf{A}^T 的零空间 $N(\mathbf{A}^T)$ 里，从上一讲讨论子空间的正交性可知，向量 \mathbf{e} 与矩阵 \mathbf{A} 的列空间正交，这也正是方程的意义。

将方程 $\mathbf{A}^T(\mathbf{b}-\mathbf{Ax})=0$ 改写，可得 $\mathbf{A}^T\mathbf{Ax}=\mathbf{A}^T\mathbf{b}$ 。两侧左乘 $(\mathbf{A}^T\mathbf{A})^{-1}$ ，得到：

$$\hat{\mathbf{x}}=(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

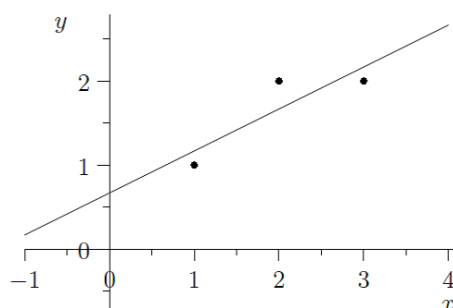
$$\mathbf{p}=\mathbf{Ax}=\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

$$\mathbf{P}=\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$$

因为矩阵 \mathbf{A} 不是方阵，无法简单的用 $(\mathbf{A}^T\mathbf{A})^{-1}=\mathbf{A}^{-1}(\mathbf{A}^T)^{-1}$ 对投影矩阵公式进行化简。若 \mathbf{A} 是可逆方阵，则化简得到 $\mathbf{P}=\mathbf{I}$ 。此时 \mathbf{A} 的列空间就是整个 \mathbf{R}^n 空间， \mathbf{b} 到这个空间的投影就是其本身，投影矩阵等于单位阵。

对 $\mathbf{P}=\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ 用矩阵乘法的结合律和矩阵乘积的转置公式，可以证明投影矩阵的性质： $\mathbf{P}^T=\mathbf{P}$ ， $\mathbf{P}^2=\mathbf{P}$ 。

最小二乘 Least Squares



应用投影矩阵求方程组最优解的方法，最常用于“最小二乘法”拟合曲线。

有三个数据点 $\{(1,1), (2,2), (3,2)\}$ ，求直线方程 $b=C+Dt$ ，要求直线尽量接近于三个点。把三个点的数据代入方程则有：

$$C+D=1$$

$$C+2D=2$$

$$C+3D=2$$

$$\text{矩阵形式为} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

这个的方程 $Ax=b$ 是无解的，解决办法就是求其最优解，即方程 $A^T A \hat{x} = A^T b$ 的解。

第 16 讲 投影矩阵和最小二乘法

Projection matrices and least squares

投影 Projections

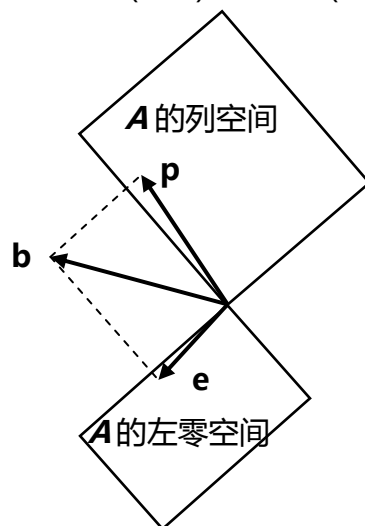
上一讲介绍了投影矩阵 $P=A(A^T A)^{-1}A^T$ ，当它作用于向量 b ，相当于把 b 投影到矩阵 A 的列空间。

如果向量 b 本身就在 A 列空间之内，即存在 x 使得 $Ax=b$ ，则有：

$$\begin{aligned} Pb &= A(A^T A)^{-1}A^T b \\ &= A(A^T A)^{-1}A^T Ax \\ &= A(A^T A)^{-1}A^T A x \\ &= Ax = b \end{aligned}$$

如果向量 b 与 A 的列空间正交，即向量 b 在矩阵 A 的左零空间 $N(A)$ 中，则有

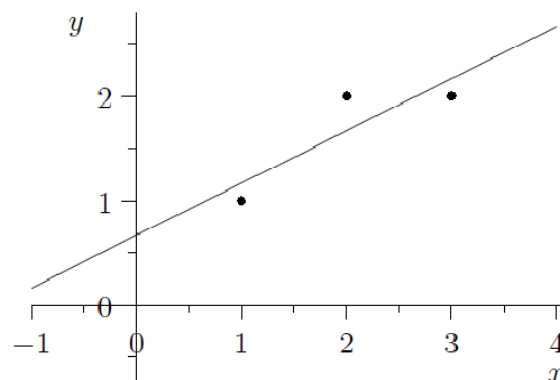
$$Pb = A(A^T A)^{-1}A^T b = A(A^T A)^{-1}(A^T b) = A(A^T A)^{-1}0 = 0。$$



$p+e=b$ ，说明 b 由两部分组成：
 $p=Pb$ 为 A 的列空间中的部分；
 $e=(I-P)b$ 为 A 的左零空间中的部分。
 $I-P$ 为左零空间的投影矩阵，可以验证 $(I-P)^T=(I-P)$ ，并且 $(I-P)^2=(I-P)$ 。

最小二乘法 Least Squares

应用投影矩阵求方程组最优解的方法，最常用于“最小二乘法”拟合曲线。

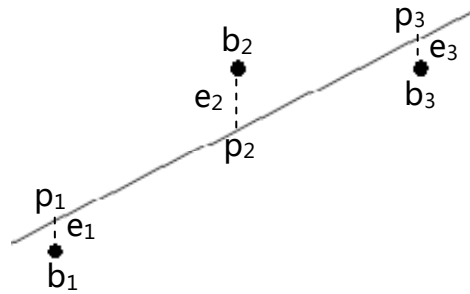


三个点 $\{(1,1), (2,2), (3,2)\}$ ，求直线方程 $b=C+Dt$ ，要求直线尽量接近于三个点。

$$\begin{aligned} C + D &= 1 \\ C + 2D &= 2 \\ C + 3D &= 2 \end{aligned}$$

$$\text{矩阵形式为 } \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

这个的方程 $\mathbf{Ax}=\mathbf{b}$ 是无解的，解决办法就是求其最优解，最优解的含义即为误差最小，这里误差就是每个方程误差值的平方和 $\|\mathbf{e}\|^2 = \|\mathbf{Ax}-\mathbf{b}\|^2$ ，因此就是寻找具有最小误差平方和的解 \mathbf{x} ，这就是所谓的“最小二乘”问题。



$$\|\mathbf{e}\|^2 = \|\mathbf{Ax}-\mathbf{b}\|^2 = e_1^2 + e_2^2 + e_3^2$$

误差即为数据点到直线距离的平方和。

这部分工作可称为线性回归，在数据点中没有“离群值”时，这是非常有用的方法。

从几何上讨论求解过程，就是试图寻找数据点到直线距离的平方和 $e_1^2 + e_2^2 + e_3^2$ 最小的情况，此时得到的 $C+Dt$ 分别为 p_1, p_2 和 p_3 ，它们是满足方程并最接近于 \mathbf{b} 的结果。另一种看法是，对于 \mathbf{R}^3 空间上的向量 \mathbf{b} ，它投影到矩阵 \mathbf{A} 的列空间中会得到向量 $\mathbf{p}=[p_1 \ p_2 \ p_3]^T$ ，投影到矩阵 \mathbf{A} 的零空间中则为 \mathbf{e} 。

$$\text{现在求解 } \hat{\mathbf{x}} = \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} \text{ 和 } \mathbf{p}.$$

$$\mathbf{A}^T \mathbf{A} \hat{\mathbf{x}} = \mathbf{A}^T \mathbf{b}$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}, \text{ 则有 } \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

得到解为 $\hat{C} = 2/3, \hat{D} = 1/2$ 。

还可以从误差最小的角度出发求解：

$$e_1^2 + e_2^2 + e_3^2 = (C+D-1)^2 + (C+2D-2)^2 + (C+3D-2)^2$$

对等号右边的表达式求偏导数，极值出现在偏导数为 0 的位置。求偏导最终会得到相同的线性方程组和相同的解。

展开结果为 $\|\mathbf{e}\|^2 = 3C^2 + 14D^2 + 9 - 10C - 22D + 12CD$ ，求偏导为 $12C - 10 + 24D = 0$ ； $28D - 22 + 12C = 0$ 。与 $\mathbf{A}^T \mathbf{A} \hat{\mathbf{x}} = \mathbf{A}^T \mathbf{b}$ 相同。

得到直线表达式 $y = 2/3 + t/2$ 。将 $t = 1, 2, 3$ 分别代入可得：

i	p_i	e_i
1	7/6	-1/6
2	5/3	1/3
3	13/6	-1/6

可以验证，向量 \mathbf{p} 与 \mathbf{e} 正交，并且 \mathbf{e} 与矩阵 \mathbf{A} 的列空间正交。

矩阵 $\mathbf{A}^T \mathbf{A}$

证明：若 \mathbf{A} 的列向量线性无关时，矩阵 $\mathbf{A}^T \mathbf{A}$ 为可逆矩阵。

假设存在 \mathbf{x} 使得 $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{0}$ 。则有 $\mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{0} = (\mathbf{A} \mathbf{x})^T (\mathbf{A} \mathbf{x})$ ，因此 $\mathbf{A} \mathbf{x} = \mathbf{0}$ 。因为 \mathbf{A} 的列向量线性无关，所以只有当 $\mathbf{x} = \mathbf{0}$ 时有 $\mathbf{A} \mathbf{x} = \mathbf{0}$ 。因此只有当 $\mathbf{x} = \mathbf{0}$ 时有 $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{0}$ 。即矩阵 $\mathbf{A}^T \mathbf{A}$ 为可逆矩阵。

如果矩阵的列向量是互相垂直的单位向量，则它们一定是线性无关的。我们将这种向量称之为标准正交 (orthonormal)。

例如： $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ ， $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ 和 $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ 。还有 $\begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}$ 和 $\begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix}$ 。

第 17 讲 正交矩阵和施密特正交化

Orthogonal matrices & Gram-Schmidt

本讲我们完成对“正交”的介绍。Gram-Schmidt 过程可以将原空间的一组基转变为标准正交基。

正交向量 Orthonormal vectors

满足如下条件的向量 $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$ 为标准正交：

$$\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 0 & \text{若 } i \neq j \\ 1 & \text{若 } i = j \end{cases}$$

换言之，它们都具有单位长度 1，并且彼此正交。标准正交向量是线性无关的。很多线性代数的计算都建立在标准正交基础上，它让一切变得简单可控。

标准正交矩阵 Orthonormal matrix

如果矩阵 \mathbf{Q} 的列向量为标准正交向量，则 $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$ 为单位阵。

$$\mathbf{Q} = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} \quad \mathbf{Q}^T \mathbf{Q} = \begin{bmatrix} \text{---} \mathbf{q}_1^T \text{---} \\ \vdots \\ \text{---} \mathbf{q}_n^T \text{---} \end{bmatrix} \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} = \mathbf{I}$$

注意这里的矩阵 \mathbf{Q} 可以不是方阵。我们已经学过了一系列矩阵，包括三角阵、对角阵、置换矩阵、对称矩阵、行最简梯形矩阵、投影矩阵等等，现在有了“标准正交”矩阵。

一个标准正交的方阵我们称之为“正交矩阵” (orthogonal matrix)。如果 \mathbf{Q} 为方阵，因为 $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$ ，所以 $\mathbf{Q}^T = \mathbf{Q}^{-1}$ 。注意必须是方阵，必须是标准正交，而不只是正交。

$$\text{例如，置换矩阵 } \mathbf{Q} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \text{ 则有 } \mathbf{Q}^T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \text{ 两者皆为正交矩阵，并}$$

且两者乘积为单位阵。

$$\text{再例如，} \mathbf{Q} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \text{ 为正交矩阵。而矩阵 } \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \text{ 并不是正交矩阵，而}$$

通过调整得到的矩阵 $\mathbf{Q} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ 为正交矩阵，在矩阵外面要除以向量的长度。

再比如 $Q = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$ 也是由-1 和+1 组成的正交矩阵，这种类型的矩

阵称之为阿达玛 Hadamard 矩阵，不同阶数矩阵性质不同并且没有规律，无从判断几阶的阿达玛矩阵为正交阵。

再给一个长方形矩阵的例子，其列向量为标准正交：

$$Q = \frac{1}{3} \begin{bmatrix} 1 & -2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix}, \text{ 我们可以拓展其成为正交矩阵 } \frac{1}{3} \begin{bmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{bmatrix}.$$

标准正交列向量的优势 Orthonormal columns are good

若 Q 的列向量为标准正交向量，则投影到 Q 的列空间的投影矩阵为：

$$P = Q(Q^T Q)^{-1} Q^T$$

因为 $Q^T Q = I$ ，所以 $P = QQ^T$ 。这种情况会降低很多运算量。如果 Q 为方阵，则 $P = I$ ，因为 Q 的列向量张成了整个空间，投影过程不会对向量有任何改变。

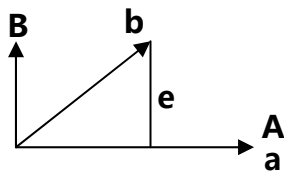
在很多复杂问题中使用标准正交向量之后都变得简单。如果基为标准正交，则方程 $A^T A \hat{x} = A^T b$ 的解变为 $\hat{x} = Q^T b$ ， \hat{x} 的分量 \hat{x}_i 就等于 $q_i^T b$ 。

施密特正交化 Gram-Schmidt

从两个线性无关的向量 a 和 b 开始，它们张成了一个空间，我们的目标是希望找到两个标准正交的向量 q_1, q_2 能张成同样的空间。Schmidt 给出的结论是如果我们有一组正交基 A 和 B ，那么我们令它们除以自己的长度就得到标准正交基：

$$q_1 = \frac{A}{\|A\|}, q_2 = \frac{B}{\|B\|}$$

Gram 做了重要的工作，令 $A=a$ ，我们在 a 和 b 张成的空间中，取与 A 正交向量做成标准正交基，方法就是将 b 投影到 a 的方向，然后取 $B=b-p$ (B 就是之前谈论过的误差 e 的方向)。



$$\text{则有 } B = b - \frac{A^T b}{A^T A} A$$

注意这个小节中 A, B, C 均为向量。

如果从等式两端左乘 A^T ，可以得到 $A^T B = 0$ 。

如果从三个线性无关的向量 a, b 和 c 出发，则可以通过从 c 中减去其在 A 和 B 两个方向的投影来得到 C 。

$$C = c - \frac{A^T c}{A^T A} A - \frac{B^T c}{B^T B} B$$

例如： $\mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$, 则有 $\mathbf{A} = \mathbf{a}$, $\mathbf{B} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - \frac{3}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$, 验证计算得 $\mathbf{A}^T \mathbf{B} = 0$ 。

写出 \mathbf{q}_1 , \mathbf{q}_2 所组成的矩阵为：

$$\mathbf{Q} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & -1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{2} \end{bmatrix}$$

\mathbf{Q} 列向量的空间就是 \mathbf{a} 和 \mathbf{b} 张成的空间。因此矩阵 \mathbf{Q} 和矩阵 $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 2 \end{bmatrix}$ 有相同

的列空间。

在消元过程中，我们可以对矩阵进行分解得到 $\mathbf{A} = \mathbf{L}\mathbf{U}$ ，而在对 \mathbf{A} 做施密特正交化的过程也可以用矩阵运算的方式表示为 $\mathbf{A} = \mathbf{Q}\mathbf{R}$ 。此处 \mathbf{R} 为上三角阵。

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1^T \mathbf{q}_1 & \mathbf{a}_2^T \mathbf{q}_1 \\ \mathbf{a}_1^T \mathbf{q}_2 & \mathbf{a}_2^T \mathbf{q}_2 \end{bmatrix}$$

$\mathbf{A} \qquad \qquad \mathbf{Q} \qquad \qquad \mathbf{R}$

\mathbf{R} 为上三角阵，则 $\mathbf{a}_1^T \mathbf{q}_2 = 0$ 。这是因为 \mathbf{a}_1 就是 \mathbf{q}_1 的方向，而 \mathbf{q}_1 和 \mathbf{q}_2 为标准正交向量，因此 \mathbf{q}_2 的方向与 \mathbf{a}_1 垂直，因此内积为 0。 \mathbf{R} 在 \mathbf{Q} 右侧相当于对 \mathbf{Q} 做列操作，即 \mathbf{A} 的列向量是 \mathbf{Q} 列向量的线性组合，而 \mathbf{Q} 为 \mathbf{A} 列空间的一组标准正交基，则 \mathbf{R} 的元素实际上是 \mathbf{A} 的列向量基于 \mathbf{Q} 这组标准正交基的权。

采用矩阵的 \mathbf{QR} 分解来帮助求解 $\mathbf{Ax} = \mathbf{b}$ 的问题，最大的优势是提高了数值的稳定性。

第 18 讲 行列式及其性质

Properties of determinants

课程进入第二大部分，之前学习了大量长方形矩阵的性质，现在我们集中讨论方阵的性质，行列式和特征值将我们的又一个重点，求行列式则与特征值息息相关。

行列式 Determinants

行列式是一个每个方阵都具有的数值，我们将矩阵 A 的行列式记作 $\det(A) = |A|$ 。它将尽可能多的矩阵信息压缩在这一个数里。例如矩阵不可逆或称奇异与矩阵的行列式等于 0 等价，因此可以用行列式来判定矩阵是否可逆。

性质 Properties

直接给出 n 阶行列式的公式，则一下子代入了大量信息，并不利于接受这个概念，我们从行列式的三个性质开始讲起，这三个性质定义了行列式。

这里可以给出二阶行列式的表达式 $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$ 。我们可以用它来验证行列

式的性质，也可以看到它本身是可以从行列式的性质推导出来的。

1. $\det(I) = 1$ 。 $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = +1$
2. 如果交换行列式的两行，则行列式的数值会反号。从前两条可以推知置换矩阵的行列式是 +1 或者 -1。 $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$
3. (a) 如果在矩阵的一行乘上 t ，则行列式的值就要乘上 t 。 $\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}$
(b) 行列式是“矩阵的行”的线性函数。 $\begin{vmatrix} a + a' & b + b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$

行列式本身是有显式的，但是直接给出显式真的无益于理解行列式，我觉得 GS 这里做了一个类似公理化的办法来给出行列式，是一个比较高明的办法。如他所言，很少有人用公式进行行列式计算，计算机也是用消元的办法来求解的。窃以为把大量时间花在“逆序”这一概念上对学习线代帮助并不大。

更多的性质可以从以上的三条性质中推导出来。

4. 如果矩阵的两行是完全相同的，则它的行列式为 0。这可以从第二条性质推导出来，因为交换这个相同的两行，行列式应该变号；但是新生成的矩阵跟原矩阵没有区别，因此行列式应该不变，所以有 $\det = -\det$ ，所以 \det 等于 0。
5. 从矩阵的某行 k 减去另一行 i 的倍数，并不改变行列式的数值，我们以二阶为例：

$$\begin{aligned}
 \begin{vmatrix} a & b \\ c - ta & d - tb \end{vmatrix} &= \begin{vmatrix} a & b \\ c & d \end{vmatrix} - \begin{vmatrix} a & b \\ ta & tb \end{vmatrix} \text{性质 3(b)} \\
 &= \begin{vmatrix} a & b \\ c & d \end{vmatrix} - t \begin{vmatrix} a & b \\ a & b \end{vmatrix} \text{性质 3(a)} \\
 &= \begin{vmatrix} a & b \\ c & d \end{vmatrix} \text{性质 4}
 \end{aligned}$$

6. 如矩阵 A 的某一行都是 0 ,则其行列式为 0。可以应用性质 3(a) ,取 $t=0$ 证明。

7. 三角阵的行列式的值等于其对角线上数值 (主元) 的乘积。

$$\begin{vmatrix} d_1 & * & * & \cdots & * \\ 0 & d_2 & * & \cdots & * \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & d_n \end{vmatrix} = \begin{vmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & d_n \end{vmatrix} = d_1 d_2 \cdots d_n \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 \end{vmatrix} = d_1 d_2 \cdots d_n$$

性质 5 告诉我们三角阵通过行消元法得到对角阵的过程中,行列式的数值没有发生变化。性质 3(a)告诉我们对角阵的行列式等于其主元的乘积再乘以单位阵的行列式。而性质 1 表明单位阵行列式为 1。

8. 当且仅当矩阵 A 为奇异矩阵时,其行列式为 0。

如果矩阵 A 为奇异阵,则必可通过消元法使得矩阵的某行全等于零,则按照性质 6, A 的行列式为 0。

如果其不是奇异阵,则通过消元可以得到一个上三角矩阵,且其主元均不为 0,则按照性质 7,行列式的数值等于主元的乘积也不等于 0。

计算非奇异矩阵的行列式有确切的公式,但通常计算机是靠消元的方法来转化为三角阵,然后将主元相乘来进行计算的。

$$\begin{aligned}
 \text{例如: 二阶矩阵 } \begin{bmatrix} a & b \\ c & d \end{bmatrix} &\rightarrow \begin{bmatrix} a & b \\ 0 & d - \frac{c}{a}b \end{bmatrix} \text{ 若 } a \neq 0, \text{ 则有} \\
 \begin{vmatrix} a & b \\ c & d \end{vmatrix} &= a(d - \frac{c}{a}b) = ad - bc
 \end{aligned}$$

9. $\det(AB) = \det(A)\det(B)$

尽管矩阵的和的行列式不等于行列式的和,但矩阵乘积的行列式等于矩阵行列式的乘积。

在课本中, GS 给了一个证明,其主要思想就是证明 $\det(AB)/\det(B)$ 与矩阵 A 的关系完全符合行列式的前三个性质,而因为前三个性质定义了行列式,所以 $\det(AB)/\det(B)$ 这就等于矩阵的行列式值,即 $\det(AB)/\det(B) = \det(A)$ 。

$$\text{如果 } A \text{ 为可逆矩阵,则 } A^{-1}A = I, \text{ 所以有 } \det(A^{-1}) = \frac{1}{\det(A)}$$

此外, $\det(A^2) = \det(A)^2$ 并且有 $\det(2A) = 2^n \det(A)$ 。后一个公式让我们容易联

想到体积，当长宽高都倍增之后，体积变成了原来的 $2^3=8$ 倍。

10. $\det(\mathbf{A}^T) = \det(\mathbf{A})$

对于二阶矩阵这显然成立: $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc$

证明：矩阵消元可得 $\mathbf{A} = \mathbf{L}\mathbf{U}$ 则 $\mathbf{A}^T = \mathbf{U}^T\mathbf{L}^T$,由性质 9 可知 $\det(\mathbf{A}) = \det(\mathbf{L})\det(\mathbf{U})$,
 $\det(\mathbf{A}^T) = \det(\mathbf{L}^T)\det(\mathbf{U}^T)$ 根据性质 7 可知 $\det(\mathbf{L}^T) = \det(\mathbf{L})$, $\det(\mathbf{U}^T) = \det(\mathbf{U})$,
则二者乘积相等。

因为性质 10 成立，所以性质 2,3,4,5,6 可以用在行列式的列性质上。

行列式的性质 2 中隐藏着一个内容，这就是置换隐藏着奇偶性，一个矩阵不可能经过奇数次置换得到和偶数次置换相同的方阵。

第 19 讲 行列式公式和代数余子式

Determinant formulas and cofactors

我们已经认识到了行列式的性质，应该推导出其公式了。

行列式公式 Formula for the determinant

行列式有如下三个性质：

1. $\det(I)=1$ 。
2. 如果交换行列式的两行，则行列式的数值会反号。
3. 行列式是“矩阵的行”的线性函数。

从这三条性质可以推导出后续的七条性质，从这十个性质出发可以得到二阶方阵的行列式公式：

$$\begin{aligned}\begin{vmatrix} a & b \\ c & d \end{vmatrix} &= \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} \\ &= \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix} \\ &= 0 + ad - cb + 0 \\ &= ad - bc\end{aligned}$$

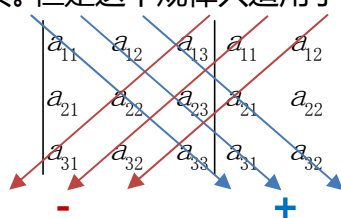
通过性质 3 对 n 阶矩阵的行列式进行拆分，我们可以得到所有只包含 n 个非零元素的行列式，对于二阶行列式我们从 1 个拆分为 2 个，然后拆分成 4 个。而对于三阶矩阵我们从 1 个拆分成 3 个，然后拆分成 9 个，最后要拆分成 27 个。但最终这些行列式中有很大部分等于 0。

$$\begin{aligned}\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & 0 \\ 0 & 0 & a_{33} \end{vmatrix} \\ &+ \begin{vmatrix} 0 & a_{12} & 0 \\ 0 & 0 & a_{23} \\ a_{31} & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ 0 & a_{22} & 0 \\ a_{31} & 0 & 0 \end{vmatrix} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} \\ &+ a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}\end{aligned}$$

每一个拆分出来的非 0 行列式都是在每行每列都有且只有一个元素，就如同置换矩阵的元素分布。应用性质 3 可以将元素从行列式中提出来，而置换矩阵的行列式值为+1 或者-1，因此可以给出行列式的公式。 n 阶拆分矩阵非 0 行列式的个数的计算方法就如同计算置换矩阵的个数一样，第一行放置一个非 0 元素的位置有 n 个选择，第二行为 $n-1$ 个……。最后得到共 $n!$ 个矩阵。

对于拆分得到的三阶矩阵，元素从上至下朝向右侧方向的，其行列式的数值为

正,朝向左侧方向的则为负。但是这个规律只适用于三阶矩阵,不适用于高阶矩阵。



行列式的公式：

$$\det \mathbf{A} = \sum_n \pm a_{1\alpha} a_{2\beta} a_{3\gamma} \cdots a_{n\omega}$$

其中列标号 ($\alpha, \beta, \gamma, \dots, \omega$) 是列标号 (1, 2, 3, ..., n) 的某个排列。比如说对于单位阵而言,只有 $\alpha=1, \beta=2, \dots, \omega=n$ 所得到的行列式为+1,其它都为零,所以单位阵的行列式为 1。

例如：

$$\begin{vmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

列标号取 (4,3,2,1) 得到第一个拆分行列式,符号为正,因为只要经过两次交换就能变为 (1,2,3,4)。第二个为 (3,2,1,4),因为只需交换一次就可变为正序,所以符号为负。因此本行列式为 0。

代数余子式 Cofactor formula

代数余子式是用较小的矩阵的行列式来写出 n 阶行列式的公式。

$$\det (\mathbf{A}) = a_{11}(a_{22}a_{33}-a_{23}a_{32})+a_{12}(-a_{21}a_{33}+a_{23}a_{31})+a_{13}(a_{21}a_{32}-a_{22}a_{31})$$

$$= \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{vmatrix}$$

将原公式中属于矩阵第一行的 a_{1j} 提出来,其系数即为代数余子式,是一个低阶行列式的值。这个低阶行列式是由原矩阵去掉 a_{1j} 所在的行和列组成的。

对矩阵中任意元素 a_{ij} 而言,其代数余子式 C_{ij} 就是矩阵的行列式的公式中 a_{ij} 的系数。 C_{ij} 等于原矩阵移除第 i 行和第 j 列后剩余元素组成的 n-1 阶矩阵的行列式数值乘以 $(-1)^{i+j}$ 。(C_{ij} 在 i+j 为偶数时为正,奇数时为负数。)

对于 n 阶方阵,其行列式的代数余子式公式为：

$$\det (\mathbf{A}) = a_{11}C_{11}+a_{12}C_{12}+\cdots+a_{1n}C_{1n}$$

对于二阶矩阵,按照代数余子式公式则有 $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad+b(-c)$

对于矩阵行列式的计算,消元的得到主元是一个很好的方法,与之相比行列式

的展开公式较为复杂，而代数余子式的方法介于两者之间，它的核心想法是通过降阶来将原来的行列式展开成更简单的行列式。

举三对角阵 (tridiagonal matrix) 为例，它除了对角线和对角线两侧相邻的元素之外，其它元素均为 0。例如由 1 组成的 4 阶三对角阵为

$$\mathbf{A}_4 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

由 1 组成的 n 阶三对角阵的行列式等于多少？

我们可以从 1 阶开始算起：

$$|\mathbf{A}_1| = 1, \quad |\mathbf{A}_2| = \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = 0, \quad |\mathbf{A}_3| = \begin{vmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} = -1,$$

$$\text{则有 } |\mathbf{A}_4| = 1 \begin{vmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} - 1 \begin{vmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} = |\mathbf{A}_3| - 1|\mathbf{A}_2| = -1,$$

从矩阵的特殊结构我们可以得到： $|\mathbf{A}_n| = |\mathbf{A}_{n-1}| - 1|\mathbf{A}_{n-2}|$

由 1 组成的 n 阶三对角阵的行列式从 1 阶开始按照 1, 0, -1, -1, 0, 1 进行循环。

第 20 讲 克莱姆法则、逆矩阵、体积

Cramer's rule, inverse matrix, and volume

我们已经了解了行列式的公式和性质，下面讨论它的应用。

逆矩阵的公式 Formula for A^{-1}

我们已经知道二阶矩阵的逆矩阵公式：

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

那么我们能写出三阶甚至高阶的公式么？通过观察二阶矩阵逆矩阵的公式，我们可以用同样的策略来构造高阶矩阵的求逆公式，为：

$$A^{-1} = \frac{1}{\det(A)} C^T$$

矩阵外因子的分母是矩阵的行列式的值，而矩阵是“代数余子式矩阵”(cofactor matrix) C 的转置矩阵 C^T 。即伴随矩阵 (adjoint matrix)。

矩阵 A 的行列式的计算中包含的都是 n 个元素的乘积：

$$\det A = \sum_{n!} \pm a_{1\alpha} a_{2\beta} a_{3\gamma} \cdots a_{n\omega}$$

而伴随矩阵中的元素都是 $n-1$ 阶行列式，它的运算中出现的是 $n-1$ 个矩阵 A 中元素的乘积。所以矩阵 A 与两者相乘才能完全消去，而得到单位矩阵。下面我们就用矩阵 A 与矩阵 C^T 相乘来验证 $AC^T = \det(A)I$ ，并且理解逆矩阵的构造策略。

$$AC^T = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1n} & \cdots & C_{nn} \end{bmatrix}$$

矩阵 AC^T 第一行第一列的元素等于矩阵 A 第一行和矩阵 C^T 第一列进行点积，计算可得：

$$\sum_{j=1}^n a_{1j} C_{1j} = \det(A)$$

点积的算式本身就是矩阵 A 的计算公式，因此结果为 A 行列式的值。而矩阵 AC^T 对角线上所有的元素都是如此，因此其对角线上的元素都等于 $\det A$ 。

而对于非对角线元素，我们以第二行第一列的元素为例，其计算公式为：

$$\sum_{j=1}^n a_{2j} C_{1j} = \det(A_s)$$

这可以视为矩阵 A_s 的行列式数值，各个代数余子式的形式不变，但是与代数余子式相乘的变为了矩阵 A 第二行第 j 列元素。因此 A_s 的形式相当于用矩阵 A 第二行的元素替代第一行的元素得到的矩阵。因为该矩阵中前两行完全相同，因此按照行

列式性质 4, $\det(\mathbf{A}_s)=0$ (GS 在课堂上是用二阶矩阵为例):

$$\sum_{j=1}^n a_{2j} C_{1j} = \begin{vmatrix} a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ a_{31} & a_{32} & \ddots & & a_{3n} \\ \vdots & & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} \end{vmatrix} = 0$$

而 $\mathbf{A}\mathbf{C}^T$ 乘积矩阵中非对角线元素的计算均服从此规律, 因此均为 0。则:

$$\mathbf{A}\mathbf{C}^T = \begin{bmatrix} \det \mathbf{A} & 0 & 0 & \cdots & 0 \\ 0 & \det \mathbf{A} & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \det \mathbf{A} \end{bmatrix} = \det(\mathbf{A})\mathbf{I}$$

即 $\mathbf{A} \frac{1}{\det(\mathbf{A})} \mathbf{C}^T = \mathbf{I}$, 因此有 $\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \mathbf{C}^T$ 。

逆矩阵公式的一个好处就是, 我们从中可以看到, 当改变原矩阵中的一个元素时, 给逆矩阵带来了怎样的变化。

克莱姆法则 Cramer's Rule for $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$

对于可逆矩阵 \mathbf{A} , 方程 $\mathbf{A}\mathbf{x}=\mathbf{b}$ 必然有解 $\mathbf{x}=\mathbf{A}^{-1}\mathbf{b}$, 将逆矩阵的公式带入其中, 则有:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \frac{1}{\det(\mathbf{A})} \mathbf{C}^T \mathbf{b}$$

克莱姆法则从另一个角度来看待这个公式。实际上 \mathbf{x} 的分量

$$x_j = \frac{\det(\mathbf{B}_j)}{\det(\mathbf{A})}$$

其中, 矩阵 \mathbf{B}_j 为用向量 \mathbf{b} 替换矩阵 \mathbf{A} 的第 j 列所得到的新矩阵。例如:

$$\mathbf{B}_1 = \begin{bmatrix} b_1 & a_{12} & \cdots & \cdots & a_{1n} \\ b_2 & a_{22} & \cdots & \cdots & a_{2n} \\ b_3 & a_{32} & \ddots & & a_{3n} \\ \vdots & & & \ddots & \vdots \\ b_n & a_{n2} & \cdots & \cdots & a_{nn} \end{bmatrix}, \quad \mathbf{B}_n = \begin{bmatrix} a_{11} & \cdots & \cdots & a_{1n-1} & b_1 \\ a_{21} & \cdots & \cdots & a_{2n-1} & b_2 \\ \vdots & & \ddots & \vdots & \vdots \\ \vdots & & & a_{n-1n-1} & b_{n-1} \\ a_{n1} & a_{n2} & \cdots & a_{nn-1} & b_n \end{bmatrix}$$

矩阵 \mathbf{B}_j 的行列式的数值从第 j 列用代数余子式进行展开计算, 正好是伴随矩阵 \mathbf{C}^T 的第 j 行与向量 \mathbf{b} 点积的结果。此处我们用到了行列式的性质 10。

相比于消元法, 采用克莱姆法则计算方程的解效率较低。

体积 $|\det(\mathbf{A})| = \text{volume of box}$

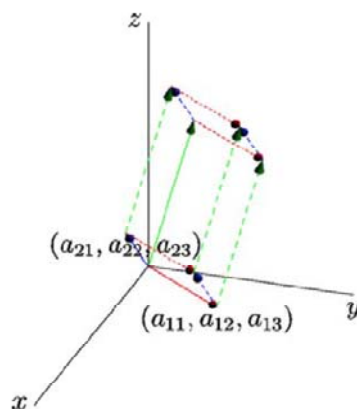
矩阵 \mathbf{A} 行列式的绝对值等于以矩阵 \mathbf{A} 行 (列) 向量为边所构成的平行六面体的体积。行列式的正负对应左手系和右手系。之前提到过行列式是将矩阵的信息压缩

成一个数，可以将“体积”视为它压缩后给出的信息。

如果矩阵 A 是单位矩阵，则其构成的是三个边长均为 1 且互相垂直的立方体，其体积为 1，这与上面的结论相符。这也是行列式的性质 1。

而如果矩阵 A 为正交矩阵 Q ，则其构成的也是三个边长为 1 且三边互相垂直的立方体，其体积也为 1 只是取向与单位阵不同。这也与上面的结论相符，因为 $Q^T Q = I$ ，且 $\det Q = \det Q^T$ ，所以 $\det Q = \pm 1$ 。

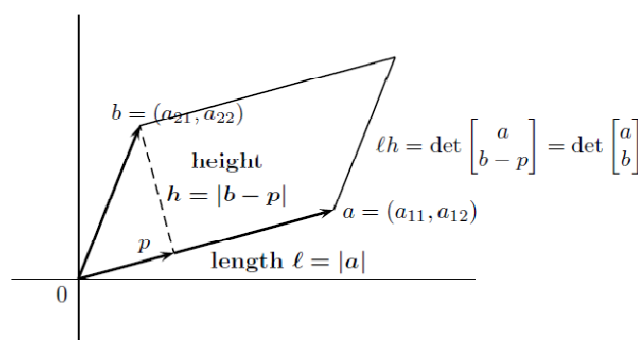
交换矩阵 A 中的行并不会改变其行列式的绝对值，显然也不会改变向量围成的体积，因此这也和体积理论相符。这是行列式的性质 2。



对于长方体，也非常直观，当你将其中一条边的边长增加 2 倍时，正方体的体积也会增加 2 倍成为长方体，这相当于性质 3a。

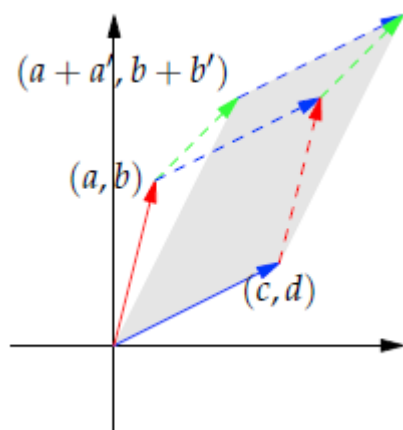
对于性质 3b，我们可以在二维条件下简单证明。

证明二阶行列式是平行四边形的面积



如图所示， a 和 b 围成的平行四边形的面积，就等于 a 和 h 的长度之积，就等于两个互相垂直的向量 a 和 h 所围成的长方形的面积，按照前面所述即为 $\begin{vmatrix} a_1 & a_2 \\ h_1 & h_2 \end{vmatrix}$ 。

根据行列式运算法则有 $\begin{vmatrix} a_1 & a_2 \\ h_1 & h_2 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} - \begin{vmatrix} a_1 & a_2 \\ p_1 & p_2 \end{vmatrix}$ ，而 p 与 a 同方向因此第二项为 0，因此平行四边形的面积等于 $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$ 。结果可以推广到高维。

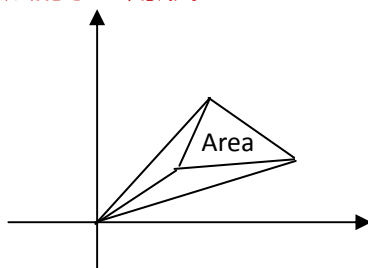


$$\begin{vmatrix} a + a' & b + b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$$

二阶行列式的值等于平行四边形的面积，同时其一半也是 $\{(a,b), (c,d), (0,0)\}$ 围成的三角形的面积。

若三角形不过原点，为 $\{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$ 围成，则其面积等于 $\frac{1}{2} \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$ 。

这个公式可以从第三列用代数余子式展开，所得结果可以看做是从一个过原点的大三角形中减去两个过原点的小三角形。



除以上各条之外，性质 4 也比较直观，当有两条边重合时，平行六面体或者平行四边形被压扁，体积或者面积为 0。网上有一个《线性代数的几何意义》的文稿，其中详细讨论了与行列式各条性质相对应的体积内含。

第 21 讲 特征值和特征向量

Eigenvalues and eigenvectors

本单元后面的课程主要围绕特征值和特征向量。在这个议题下讨论得都是方阵。

特征向量和特征值 Eigenvectors and eigenvalues

将矩阵 A 与向量 x 相乘当做是对向量的一种操作或者函数 输入 x 而输出 Ax 。

特征向量即在特定的向量 x 方向上输出的 Ax 平行于 x ，即为：

$$Ax = \lambda x$$

其中 x 为矩阵 A 的特征向量，而 λ 为 A 的特征值。

如果 0 是矩阵的特征值，则有 $Ax = 0x = 0$ 。特征值 0 所对应的向量生成了矩阵的零空间。如果矩阵 A 为不可逆矩阵，则 0 是其特征值之一。

例 1 矩阵 P 是朝向某平面的投影矩阵。对于这个平面之内的 x 均有 $Px = x$ ，因此 x 是特征向量而 1 为特征值。垂直于该平面的向量 x 经投影得到 $Px = 0$ ，这个 x 也是矩阵的特征向量而 0 为特征值。矩阵 P 的所有特征向量张成了整个空间。

例 2：交换矩阵 $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ ，具有特征向量 $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ，对应的特征值为 1；另一个特征值为 $x = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ ，对应的特征值为 -1。这些特征向量张成了整个空间。因为是对称矩阵，其特征向量互相垂直。

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

任意 $n \times n$ 矩阵 A 具有 n 个特征值，并且它们的和等于矩阵对角线上的元素之和，这个数值为矩阵的迹 (trace)。对于二阶矩阵，在已知一个特征值的条件下，可以据此得到另一个特征值。

方程 $Ax = \lambda x$ 中特征值和特征向量均未知，没法直接求解。因此我们做如下数学处理： $Ax = \lambda x$ ，因此有 $(A - \lambda I)x = 0$ 。则 $A - \lambda I$ 为奇异阵，因此 $\det(A - \lambda I) = 0$ 。在这个没有 x 的“特征方程”中，可以解得 n 个特征值，但是有可能方程有重根，则会得到重复的特征值。

得到特征值之后，可以用消元法解 $(A - \lambda I)$ ，这一矩阵零空间中的向量为矩阵 A 的特征向量。

例如：矩阵 $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ ，则 $\det(A - \lambda I) = \begin{vmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} = (3-\lambda)^2 - 1 = \lambda^2 - 6\lambda + 8$ 。可

以看到其中的参数 6 是矩阵 A 的迹而 8 是行列式的值。通常一个 2 阶矩阵的特征值

是如下方程的解： $\lambda^2 - \text{trace}(\mathbf{A})\lambda + \det \mathbf{A} = 0$ 。

对于上述矩阵 \mathbf{A} ，可以求得特征值 $\lambda_1 = 4$ 和 $\lambda_2 = 2$ 。

$$\mathbf{A} - 4\mathbf{I} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}, (\mathbf{A} - 4\mathbf{I})\mathbf{x}_1 = 0, \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix};$$

$$\mathbf{A} - 2\mathbf{I} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, (\mathbf{A} - 2\mathbf{I})\mathbf{x}_2 = 0, \mathbf{x}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix};$$

与前例中矩阵 $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ 的特征值和特征向量相对比，可知两者为一组平移矩阵。在对角元素上分别加 3，改变了特征值但是没有改变特征向量。

$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ ，则有 $(\mathbf{A} + 3\mathbf{I})\mathbf{x} = \lambda\mathbf{x} + 3\mathbf{x} = (\lambda + 3)\mathbf{x}$ 。

需要注意的是，两个矩阵的和的特征值不是两特征值直接相加之和，因为特征向量并不相同。

矩阵的迹等于特征值之和：如上所述，将 $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ 展开会得到 λ 的 n 阶多项式，多项式的解就是矩阵 \mathbf{A} 的特征值，根据多项式根与系数的关系，解之和即特征值之和等于 λ^{n-1} 的系数。而行列式展开式中只有对角线的积这一项包含的 λ^{n-1} （其它项最高是 $n-2$ 次方），而其系数为矩阵 \mathbf{A} 对角线元素之和即矩阵 \mathbf{A} 的迹，因此特征值之和与矩阵的迹相等。

对称矩阵的特征向量正交： λ_1 和 λ_2 是对称矩阵的两个不同的特征值，对应的特征向量分别为 \mathbf{x}_1 和 \mathbf{x}_2 。则有 $\mathbf{A}\mathbf{x}_1 = \lambda_1\mathbf{x}_1$ ，左乘 \mathbf{x}_2^T 得 $\mathbf{x}_2^T \mathbf{A}\mathbf{x}_1 = \lambda_1 \mathbf{x}_2^T \mathbf{x}_1$ 。而又有 $\mathbf{x}_2^T \mathbf{A}\mathbf{x}_1 = (\mathbf{A}^T \mathbf{x}_2)^T \mathbf{x}_1 = \lambda_2 \mathbf{x}_2^T \mathbf{x}_1$ 。因此有 $(\lambda_1 - \lambda_2) \mathbf{x}_2^T \mathbf{x}_1 = 0$ ，而两特征值不等，所以两特征向量正交。

复数特征值 Complex eigenvalues

矩阵 $\mathbf{Q} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \cos 90^\circ & -\sin 90^\circ \\ \sin 90^\circ & \cos 90^\circ \end{bmatrix}$ 是一个 90 度旋转矩阵。从矩阵的迹和行列式的值可以得到 $\lambda_1 + \lambda_2 = 0$ ， $\lambda_1 \lambda_2 = 1$ 。从矩阵的性质可知它的实数特征向量只有零向量，因为其他任何向量乘以旋转矩阵，向量的方向都会发生改变。计算可得：

$$\det(\mathbf{Q} - \lambda\mathbf{I}) = \begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 + 1 = 0$$

可以解得 $\lambda_1 = i$ 和 $\lambda_2 = -i$ 。如果一个矩阵具有复数特征值 $a + bi$ 则，它的共轭复数 $a - bi$ 也是矩阵的特征值。**实数特征值让特征向量伸缩而虚数让其旋转。**

对称矩阵永远具有实数的特征值，而反对称矩阵 (antisymmetric matrices)，即满足 $\mathbf{A}^T = -\mathbf{A}$ 的矩阵，具有纯虚数的特征值。

三角阵和重特征值 Triangular matrices and repeated eigenvalues

对于如 $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$ 的三角矩阵，特征值就是矩阵对角线上的元素。

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \begin{vmatrix} 3-\lambda & 1 \\ 0 & 3-\lambda \end{vmatrix} = (3-\lambda)(3-\lambda) = 0$$

可以解得 $\lambda_1 = \lambda_2 = 3$ 。

$(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x} = 0$ 。得到 $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ，而没有线性无关的 \mathbf{x}_2 。说明 \mathbf{A} 是一个退化矩阵，对应相同的特征值，而特征向量短缺。

第 22 讲 对角化和矩阵的幂

Diagonalization and powers of A

本讲中将学习如何对角化含有 n 个线性无关特征向量的矩阵，以及对角化是怎样简化计算的。

对角化矩阵 Diagonalizing a matrix $S^{-1}AS = \Lambda$

如果矩阵 A 具有 n 个线性无关的特征向量，将它们作为列向量可以组成一个可逆方阵 S ，并且有：

$$\begin{aligned} AS &= A[\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n] \\ &= [\lambda_1 \mathbf{x}_1 \quad \lambda_2 \mathbf{x}_2 \quad \cdots \quad \lambda_n \mathbf{x}_n] \\ &= S \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} = S\Lambda \end{aligned}$$

这里的矩阵 Λ 为对角阵，它的非零元素就是矩阵 A 的特征值。因为矩阵 S 中的列向量线性无关，因此逆矩阵 S^{-1} 存在。在等式两侧左乘逆矩阵，得到 $S^{-1}AS = \Lambda$ 。同样地， $A = S\Lambda S^{-1}$ 。

对于消元法而言，矩阵有 LU 分解，对于施密特正交法，矩阵有 QR 分解，而上面的推导是一种新的矩阵分解。

之前曾经提到过消元进行行操作和列操作最后会得到“相抵标准型”。现在我们得到的是矩阵的“相似标准形”。以前听过一个比喻，说矩阵分解得到标准型，就好像带着功能眼镜看人，红外眼镜看到的是相似标准型，还保有矩阵操作的基本性质——特征值，而相抵标准型就是X射线看出来的，只剩下最内核的秩信息还保留着。

矩阵的幂 Powers of A

特征值给矩阵的幂计算提供了方法。

如果 $A\mathbf{x} = \lambda\mathbf{x}$ ，则有 $A^2\mathbf{x} = \lambda A\mathbf{x} = \lambda^2\mathbf{x}$ 。说明矩阵 A^2 有着和 A 一样的特征向量，而特征值为 λ^2 。我们将写成对角化形式则有： $A^2 = S\Lambda S^{-1}S\Lambda S^{-1} = S\Lambda^2 S^{-1}$ 。做相同的处理还可以得到： $A^k = S\Lambda^k S^{-1}$ 。这说明 A^k 有着和 A 一样的特征向量，而特征值为 λ^k 。

如果矩阵 A 具有 n 个线性无关的特征向量，如果所有的特征值均满足 $|\lambda_i| < 1$ 。则 $k \rightarrow \infty$ 时， $A^k \rightarrow 0$ 。

重特征值 Repeated eigenvalues

如果矩阵 A 没有重特征值，则其一定具有 n 个线性无关的特征向量。

如果矩阵 A 有重特征值,它有可能具有 n 个线性无关的特征向量,也可能没有。

比如单位阵的特征值为重特征值 1,但是其具有 n 个线性无关的特征向量。

对于如 $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ 的三角矩阵,特征值就是矩阵对角线上的元素 2。其特征

向量在 $A - \lambda I$ 的零空间中,满足 $(A - \lambda I)x = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x = 0$ 。求解可得 $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, 而没有第二个特征向量。

差分方程 Difference equations $u_{k+1} = Au_k$

从给定的一个向量 u_0 出发,我们可以通过对前一项乘以矩阵 A 得到下一项的方式,得到一个向量序列: $u_{k+1} = Au_k$ 。

这里的 $u_{k+1} = Au_k$ 可以是一个一阶差分方程,而 $u_k = A^k u_0$ 就是方程的解。但这种简洁形式并没有给出足够的信息,我们需要通过特征向量和矩阵的幂运算给出真实解的结构。

将 u_0 写成特征向量的线性组合:

$$u_0 = C_1 x_1 + C_2 x_2 + \dots + C_n x_n = Sc,$$

$$Au_0 = C_1 \lambda_1 x_1 + C_2 \lambda_2 x_2 + \dots + C_n \lambda_n x_n,$$

$$u_k = A^k u_0 = C_1 \lambda_1^k x_1 + C_2 \lambda_2^k x_2 + \dots + C_n \lambda_n^k x_n = A^k Sc.$$

下面用斐波那契数列进行举例。

斐波那契数列 Fibonacci sequence

斐波那契数列为 0,1,1,2,3,4,8,13.....其通项公式为 $F_{k+2} = F_{k+1} + F_k$ 。求 F_{100} ?

如果我们以矩阵的方式来理解数列,则矩阵的特征值可以告诉我们数列中数值的增长速度。

为了凑成矩阵形式,需要用一个比较巧妙的技巧。令 $u_k = \begin{bmatrix} F_{k+2} \\ F_{k+1} \end{bmatrix}$, 则有:

$$\begin{array}{l} F_{k+2} = F_{k+1} + F_k \\ F_{k+1} = F_{k+1} \end{array} \quad \text{写成矩阵形式为 } u_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} u_k$$

观察矩阵 $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ 的特征值和特征向量,因为其为对称矩阵,特征值为实数,

且特征向量正交。

$$\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 - \lambda - 1 = 0$$

解得 $\lambda_1 = \frac{1+\sqrt{5}}{2}$, $\lambda_2 = \frac{1-\sqrt{5}}{2}$ 绝对值大于 1 的 λ_1 控制着在斐波那契数列的增长。

$u_k = A^k u_0 = C_1 \lambda_1^k x_1 + C_2 \lambda_2^k x_2$, 因为 $|\lambda_2| < 1$, 所以当乘方次数较高时有 $k \rightarrow \infty, \lambda_2^k \rightarrow 0$ 。

从特征值可以求得对应的特征向量 $\mathbf{x}_1 = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}$ 的和 $\mathbf{x}_2 = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$ 。在这里 GS 用了个小技巧，因为是二阶方程，而且矩阵 $\mathbf{A} - \lambda \mathbf{I}$ 是奇异矩阵，所以只要符合其中一个方程即可，立刻可以看出 $[\lambda \ 1]^T$ 是解。

$$\text{从 } \mathbf{u}_0 = \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2, \text{ 可以求得 } c_1 = -c_2 = \frac{1}{\sqrt{5}}。$$

$$\begin{bmatrix} F_{100} \\ F_{99} \end{bmatrix} = \mathbf{A}^{99} \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1^{99} & \lambda_2^{99} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} c_1 \lambda_1^{100} + c_2 \lambda_2^{100} \end{bmatrix}$$

可知 $F_{100} \approx c_1 \lambda_1^{100}$

很多人会问矩阵的特征值特征向量为什么这么神奇，可以把矩阵的操作变成一个简单的参数 λ 。还有人会问道为什么特征值在物理中出现非常频繁。对此我只能简单解释一下，物理中常见的被研究物体都有一个自身的内禀结构，这个内在结构的方向往往和观察者也就是外场的坐标有区别。当我们给物体施加一个外场刺激的时候，比如说外力或者电场极化等等，物体沿着其内在结构的取向来响应外场，但是观察者从外场坐标下采集反馈。实际上矩阵在不同坐标之间实现变换，特征向量显示了物体内结构的方向，特征向量则是在这个主方向上物体对外场的响应参数。在有的领域直接将特征值称为伸缩系数，实际上它反应了在其所对应的特征向量方向上，内结构与外场之间的相互关系。

特征值还有一个应用是作为降维的判据，比如在图像压缩过程中，极小的特征值会被赋值为 0，以此节省存储空间，也便于其它操作。反应在图像上，降维后的图像基本轮廓依旧清晰，图像细节有所牺牲。

第 23 讲 微分方程和 e^{At}

Differential equations and e^{At}

$$\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}$$

本讲中我们将面对微分方程，将一阶常系数微分方程转化为线性代数问题进行处理。主要思路基于常系数线性方程的解是指数形式，而寻找其指数和系数就是线代主要研究的问题。这里会涉及到矩阵型指数的运算 e^{At} 。

微分方程 Differential equations

$$\begin{aligned} \text{例: } \frac{du_1}{dt} &= -u_1 + 2u_2 \\ \frac{du_2}{dt} &= u_1 - 2u_2 \end{aligned} \quad \text{初值条件 } u_1=1, u_2=0 \rightarrow \frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}, \mathbf{A} = \begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix}, \mathbf{u}(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

分析矩阵 \mathbf{A} 的目的是要追踪 \mathbf{u} 随时间的变化，而首先要做的是找到矩阵的特征值和特征变量。矩阵 \mathbf{A} 为奇异矩阵，因此存在一个特征值 $\lambda_1=1$ ，而矩阵的迹为-3，因此还有一个特征值为 $\lambda_2=-3$ 。

一阶线性微分方程的解的形式是 $e^{\lambda t}$ 。两个特征值中，1 会使结果达到稳态，而-3 所对应的 e^{-3t} 会随时间增大而变小

方程的通解为 $\mathbf{u}(t) = c_1 e^{\lambda_1 t} \mathbf{x}_1 + c_2 e^{\lambda_2 t} \mathbf{x}_2$ 。

尽管直接代入可以验证 $e^{\lambda_1 t} \mathbf{x}_1$ 是方程的解，但类似我这种比较笨的，微分方程的解 $e^{\lambda t}$ 是如何跟矩阵的特征值联系起来的根本理解不了，只好用差商方程帮助理解。方程所求的是 t 时刻 \mathbf{u} 的状态 $\mathbf{u}(t)$ ，设需要 N 步从 $\mathbf{u}(0)$ 达到 $\mathbf{u}(t)$ ，且每一步都满足 $\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}$ ，即 $\frac{\mathbf{u}_{\text{step}(n+1)} - \mathbf{u}_{\text{step}(n)}}{\Delta t} = \frac{\mathbf{u}_{\text{step}(n+1)} - \mathbf{u}_{\text{step}(n)}}{t/N} = \mathbf{A}\mathbf{u}_{\text{step}(n)}$ ，从 $\mathbf{u}(0) = \mathbf{u}_{\text{step}(0)}$ 开始进行计算：

$$\frac{\mathbf{u}_{\text{step}(1)} - \mathbf{u}_{\text{step}(0)}}{t/N} = \mathbf{A}\mathbf{u}_{\text{step}(0)}$$

$$\text{则有 } \mathbf{u}_{\text{step}(1)} = \left(\frac{t}{N}\mathbf{A} + 1\right)\mathbf{u}_{\text{step}(0)} = \left(\frac{t}{N}\mathbf{A} + 1\right)\mathbf{u}(0)$$

将 $\mathbf{u}(0)$ 表示成为矩阵特征向量的线性组合，则有

$$\mathbf{u}(0) = c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2, \mathbf{A}\mathbf{u}(0) = c_1 \lambda_1 \mathbf{x}_1 + c_2 \lambda_2 \mathbf{x}_2,$$

$$\text{代入前式得到 } \mathbf{u}_{\text{step}(1)} = c_1 \left(\frac{t}{N}\lambda_1 + 1\right) \mathbf{x}_1 + c_2 \left(\frac{t}{N}\lambda_2 + 1\right) \mathbf{x}_2,$$

$$\text{将之代入 } \mathbf{u}_{\text{step}(2)} = \left(\frac{t}{N}\mathbf{A} + 1\right)\mathbf{u}_{\text{step}(1)},$$

可得 $\mathbf{u}_{\text{step}(2)} = c_1 \left(\frac{t}{N}\lambda_1 + 1\right)^2 \mathbf{x}_1 + c_2 \left(\frac{t}{N}\lambda_2 + 1\right)^2 \mathbf{x}_2$ ，重复代入过程最后得到：

$$\mathbf{u}(t) = \mathbf{u}_{step(N)} = c_1 \left(\frac{t}{N} \lambda_1 + 1\right)^N \mathbf{x}_1 + c_2 \left(\frac{t}{N} \lambda_2 + 1\right)^N \mathbf{x}_2$$

$$\text{令 } \frac{1}{k_1} = \frac{t}{N} \lambda_1, \frac{1}{k_2} = \frac{t}{N} \lambda_2, \text{ 则有 } \mathbf{u}(t) = c_1 \left(\frac{1}{k_1} + 1\right)^{k_1 t \lambda_1} \mathbf{x}_1 + c_2 \left(\frac{1}{k_2} + 1\right)^{k_2 t \lambda_2} \mathbf{x}_2$$

$$\text{因为 } \lim_{N \rightarrow \infty} \left(\frac{1}{k} + 1\right)^k \rightarrow e, \text{ 所以 } \mathbf{u}(t) = c_1 e^{\lambda_1 t} \mathbf{x}_1 + c_2 e^{\lambda_2 t} \mathbf{x}_2$$

$$\text{将 } \lambda_1=0, \lambda_2=-3 \text{ 代入 } (\mathbf{A}-\lambda \mathbf{I})\mathbf{x}=0, \text{ 分别求得对应的特征向量 } \mathbf{x}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

$$\mathbf{u}(t) = c_1 e^{\lambda_1 t} \mathbf{x}_1 + c_2 e^{\lambda_2 t} \mathbf{x}_2 = c_1 e^0 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 e^{-3t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\mathbf{u}(0) = c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \text{ 可解得 } c_1 = c_2 = 1/3.$$

$$\text{因此 } \mathbf{u}(t) = \frac{1}{3} \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{1}{3} e^{-3t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \text{ 前一项为稳态状态, 后一项随着时间衰减.}$$

特征向量给出的是方向，c 给出的是在不同方向上所占权重，而特征值给出的是矩阵操作在特征向量方向的操作效果，在上一讲差分方程里它就是一个倍数即线性增长能力，而在微分方程中它代表在该方向上的指数增长能力。因此两者稳定性的评价并不相同。

稳定性：并不是所有的系统都会达到稳态，矩阵的特征值会告诉我们 $\mathbf{u}(t)$ 的发展趋势。

- 1) $\text{Re}(\lambda) < 0$ ，则有 $\mathbf{u}(t) \rightarrow 0$ 。（支配稳定性的是实部，虚部的效果是在单位圆上转圈。）
- 2) 稳态：有一个特征值为 0，并且其它所有的特征值实部为负数。
- 3) 至少有一个特征值满足 $\text{Re}(\lambda) > 0$ ，发散。

如果二阶矩阵 $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 的两个特征值实部为负数，则矩阵的迹 $a+d$ 也是负数。反之并不一定成立，例如矩阵 $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ 的迹为 -1，但是一个特征值为 1。如果二阶矩阵的行列式为正则迹为负，则解为收敛的。

在方程 $\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}$ 中，矩阵 \mathbf{A} 的状态表明不同分量之间相互耦合，而用特征值和特征向量处理进行对角化是为了解耦。令 $\mathbf{u} = \mathbf{S}\mathbf{v}$ ，其中 \mathbf{S} 是由矩阵 \mathbf{A} 的特征向量组成。则有：

$$\begin{aligned} \mathbf{S} \frac{d\mathbf{v}}{dt} &= \mathbf{A}\mathbf{S}\mathbf{v} \\ \Rightarrow \frac{d\mathbf{v}}{dt} &= \mathbf{S}^{-1} \mathbf{A}\mathbf{S}\mathbf{v} = \mathbf{\Lambda}\mathbf{v} \quad \text{新的方程不再耦合} \end{aligned}$$

则方程组的对角线为： $\frac{dv_i}{dt} = \lambda_i v_i$ ，方程组的通解为：

$$\mathbf{v}(t) = e^{\mathbf{A}t} \mathbf{v}(0)$$

$$\mathbf{u}(t) = \mathbf{S} e^{\mathbf{A}t} \mathbf{S}^{-1} \mathbf{u}(0) = e^{\mathbf{A}t} \mathbf{u}(0)$$

矩阵指数函数 Matrix exponential $e^{\mathbf{A}t}$

我们可以用幂级数的公式：

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \dots$$

来定义矩阵型指数运算 $e^{\mathbf{A}t}$ ：

$$e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{(\mathbf{A}t)^2}{2} + \frac{(\mathbf{A}t)^3}{6} + \dots$$

如果 $\mathbf{A}t$ 的特征值很小，满足收敛条件 $|\lambda(\mathbf{A}t)| < 1$ ，则可以用几何级数来定义矩阵型指数：

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \rightarrow (\mathbf{I} + \mathbf{A}t)^{-1} = \mathbf{I} + \mathbf{A}t + (\mathbf{A}t)^2 + (\mathbf{A}t)^3 + \dots$$

前文中我们已经写出了矩阵指数函数的公式 $e^{\mathbf{A}t} = \mathbf{S} e^{\mathbf{A}t} \mathbf{S}^{-1}$ 。如果矩阵 A 具有 n 个线性无关的特征向量，我们可以从幂级数定义的矩阵指数公式来再次验证：

$$\begin{aligned} e^{\mathbf{A}t} &= \mathbf{I} + \mathbf{A}t + \frac{(\mathbf{A}t)^2}{2} + \frac{(\mathbf{A}t)^3}{6} + \dots \\ &= \mathbf{S} \mathbf{S}^{-1} + \mathbf{S} \mathbf{A} \mathbf{S}^{-1} t + \frac{\mathbf{S} \mathbf{A}^2 \mathbf{S}^{-1}}{2} t^2 + \frac{\mathbf{S} \mathbf{A}^3 \mathbf{S}^{-1}}{6} t^3 + \dots \\ &= \mathbf{S} (\mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2}{2} t^2 + \frac{\mathbf{A}^3}{6} t^3 + \dots) \mathbf{S}^{-1} \\ &= \mathbf{S} e^{\mathbf{A}t} \mathbf{S}^{-1} \end{aligned}$$

能够对角化的矩阵都可以表示为上式。

$$e^{\mathbf{A}t} = \begin{bmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & e^{\lambda_n t} \end{bmatrix}$$

二阶微分方程 Second order differential equations

我们可以将二阶微分方程 $y'' + by' + ky = 0$ 转化为 2×2 的一阶问题进行处理，构造方法类似于我们对斐波那契数列的处理方法。

$$\text{令 } \mathbf{u} = \begin{bmatrix} y' \\ y \end{bmatrix}, \text{ 则有 } \mathbf{u}' = \begin{bmatrix} y'' \\ y' \end{bmatrix} = \begin{bmatrix} -b & -k \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y' \\ y \end{bmatrix}$$

如果是 k 阶微分方程，那么需要一个 $k \times k$ 矩阵，除了第一行和对角线下面一排斜线上的元素之外，这个系数矩阵其它元素均为 0。

第 24 讲 马尔可夫矩阵；傅里叶级数

Markov matrices; Fourier series

本讲学习马尔可夫矩阵和傅里叶级数，两者是关于特征值和投影矩阵的应用。

马尔可夫矩阵 Markov matrices

$$\mathbf{A} = \begin{bmatrix} 0.1 & 0.01 & 0.3 \\ 0.2 & 0.99 & 0.3 \\ 0.7 & 0 & 0.4 \end{bmatrix}$$

形如矩阵 \mathbf{A} ，任何元素非负，且每列的元素加和为 1 的矩阵被称为马尔可夫矩阵。马尔可夫矩阵主要应用在概率领域。将一个马尔可夫矩阵进行方幂运算得到的仍旧是马尔可夫矩阵。

当处理一个微分方程问题时，特征值 0 意味着得到一个稳态。当进行矩阵的方幂运算时，特征值给出稳态的条件包括：

1. $\lambda_1=1$ 是特征值之一。
2. 其它特征值的绝对值都比 1 小， $|\lambda_i| < 1$ 。

如我们所知，如果矩阵具有 n 个线性无关的特征向量，则有：

$$\mathbf{u}_k = \mathbf{A}^k \mathbf{u}_0 = c_1 \lambda_1^k \mathbf{x}_1 + c_2 \lambda_2^k \mathbf{x}_2 + \dots + c_n \lambda_n^k \mathbf{x}_n.$$

如果 $\lambda_1=1$ 并且其他的特征值都小于 1，则系统在 k 增大过程中趋近于 \mathbf{u}_0 的分量 $c_1 \mathbf{x}_1$ ，即给出了一个稳态状况。这里特征向量 \mathbf{x}_1 的每一分量都是正的，因此若初始值为正，则最终的稳态也是正的。

Markov 矩阵每一列的元素加和为 1 这个条件，保证了矩阵具有 1 这个特征值。

$$\mathbf{A} - \mathbf{I} = \begin{bmatrix} -.9 & .01 & .3 \\ .2 & -.01 & .3 \\ .7 & 0 & -.6 \end{bmatrix}$$

从每一列减去 1，则每列的加和都从 1 变为 0。这时候行向量相加的结构就是 $\mathbf{0}$ 向量，因此行向量线性相关，矩阵为奇异矩阵。矩阵 \mathbf{A} 有特征向量在 $\mathbf{A} - \mathbf{I}$ 的零空间中，其对应的特征值为 1。回代计算可得 $\mathbf{x}_1 = \begin{bmatrix} 0.6 \\ 33 \\ 0.7 \end{bmatrix}$

转置矩阵的特征值：转置矩阵 \mathbf{A}^T 具有和原矩阵 \mathbf{A} 相同的特征值。

$$(\mathbf{A} - \lambda \mathbf{I})^T = \mathbf{A}^T - \lambda \mathbf{I}$$

则由行列式的性质 10 可知 $\det(\mathbf{A} - \lambda \mathbf{I}) = \det(\mathbf{A}^T - \lambda \mathbf{I})$ 。如果 λ 是矩阵 \mathbf{A} 的特征值，则其满足 $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$ ，因此它也是转置矩阵 \mathbf{A}^T 的特征值。

但是特征向量有所区别，零空间不等同于左零空间。

我们用马尔可夫矩阵来研究人口流动问题。

$$\begin{bmatrix} u_{Cal} \\ u_{Mass} \end{bmatrix}_{t=k+1} = \begin{bmatrix} .9 & .2 \\ .1 & .8 \end{bmatrix} \begin{bmatrix} u_{Cal} \\ u_{Mass} \end{bmatrix}_{t=k}$$

方程中 \mathbf{u} 的分量分别代表加利福尼亚州和马萨诸塞州的人口，矩阵中的每一列中元素代表着人口去留比例，比如第一列 0.9 表示留在加州的人口占加州人口的 90%，而 10% 进入麻省，第二列中由麻省进入加州的人口占麻省的 20%，而 80% 选择留在麻省。可以看到列向量分量的加和为 1 保证了整体人数不会变化，而在这种问题中矩阵也不会出现负的元素。

$$\text{如果取初值} \begin{bmatrix} u_{Cal} \\ u_{Mass} \end{bmatrix}_0 = \begin{bmatrix} 0 \\ 1000 \end{bmatrix}$$

$$\text{则经过一次迁徙} \begin{bmatrix} u_{Cal} \\ u_{Mass} \end{bmatrix}_1 = \begin{bmatrix} .9 & .2 \\ .1 & .8 \end{bmatrix} \begin{bmatrix} 0 \\ 1000 \end{bmatrix} = \begin{bmatrix} 200 \\ 800 \end{bmatrix}$$

为了获取长时间后的人口分布，我们需要了解矩阵的特征值和特征向量。因为这是马尔可夫矩阵，所以有一个特征值 1，则另一个特征值为 $0.9+0.8-1=0.7$ 。可以求得 $\mathbf{x}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ ， $\mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ 。从 $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ 可知最后的稳态为加州人口 2/3，麻省人口 1/3。

通解为： $\mathbf{u}_k = c_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2 (0.7)^k \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ ，可以从 \mathbf{u}_0 解得 $c_1 = 1000/3, c_2 = 2000/3$ 。

傅里叶级数 Fourier series

如果有一组标准正交基 $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$ 为则任意向量 \mathbf{v} 可以写成：

$$\mathbf{v} = x_1 \mathbf{q}_1 + x_2 \mathbf{q}_2 + \dots + x_n \mathbf{q}_n,$$

因为当 i, j 不相等时有 $\mathbf{q}_i^T \mathbf{q}_j = 0$ 。因此有 $\mathbf{q}_i^T \mathbf{v} = x_1 \mathbf{q}_i^T \mathbf{q}_1 + x_2 \mathbf{q}_i^T \mathbf{q}_2 + \dots + x_n \mathbf{q}_i^T \mathbf{q}_n = x_i$ 。

我们得到了分量 x_i 的公式： $x_i = \mathbf{q}_i^T \mathbf{v}$ 。

因为 $\mathbf{v} = [\mathbf{q}_1 \ \dots \ \mathbf{q}_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ ，即 $\mathbf{v} = \mathbf{Q}\mathbf{x}$ ，所以 $\mathbf{x} = \mathbf{Q}^{-1}\mathbf{v}$ 。因为 \mathbf{Q} 为正交矩阵，所以有 $\mathbf{Q}^{-1} = \mathbf{Q}^T$ 。 $\mathbf{x} = \mathbf{Q}^T \mathbf{v}$ ，这与我们之前得到的 $x_i = \mathbf{q}_i^T \mathbf{v}$ 完全相同。这里给出了求分量的思路就是用空间的一组标准正交基去点乘目标向量，利用其标准正交的性质得到所求。

标准正交是此处的核心概念。而傅里叶级数也是在这个概念上构建的。我们可以对任意函数做傅里叶展开，得到表达式：

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \dots$$

与之前的有限个标准正交向量组成的正交矩阵不同，这个空间是无限维，它的一组基是 $1, \cos x, \sin x, \cos 2x, \sin 2x, \dots$ 。

此处的正交概念与 \mathbf{R}^n 空间不同，点积的概念也不同。

$$\text{向量 } \mathbf{v}^T \mathbf{w} = v_1 w_1 + v_2 w_2 + \dots + v_n w_n$$

$$\text{函数 } f^T g = \int_0^{2\pi} f(x) g(x) dx$$

计算基 $\sin x$ 和 $\cos x$ 的点积可以验证其正交性。

$$\int_0^{2\pi} \sin x \cos x dx = \frac{1}{2} (\sin 2x)^2 \Big|_0^{2\pi} = 0$$

采用和标准正交基相同的策略可以得到傅里叶变换的参数。

$$\begin{aligned} \int_0^{2\pi} f(x) \cos x dx &= \int_0^{2\pi} (a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + \dots) \cos x dx \\ &= 0 + \int_0^{2\pi} a_1 \cos^2 x dx + 0 + 0 \dots \\ &= a_1 \pi \end{aligned}$$

可以得到 $a_1 = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos x dx$ ，同理可以求得其它参数。

之前我们已经提到了 GS 所讲的线代和其他人的差别，学习越深入则越能体会到这一点。网上有一篇《理解矩阵》的博客介绍了很多大家对线代的认识误区，在开篇处作者就点出了初学者面对线代为什么觉得困难，因为我们目前学习的线代是通过公理化来表述的第二代数学模型，而此前大家都是在以实用为导向的第一代数学模型中，困难来自于全面提升的抽象性。GS 做的事情大体上可以理解为在第一代数学模型中讨论线性计算的核心内容，因此在学习中可以保留直觉性，非常程度地降低了抽象性带来的困难。这种讲授思路的优劣，只有在有明确的目标情况下才好判定。换句话说，看你学线代的目的是什么，如果是数值计算或者是在一些实验科学中使用矩阵工具，那么 GS 讲授的内容非常恰如其分。至于比较抽象性的部分，我推荐《线性代数应该这样学》这本书，以线性算子作为核心，可以从另一种框架下领略线代的魅力，并且深入地理解线性代数的内涵。

复习二

Exam 2 Review

第二单元主要内容

- 正交矩阵 Q ，用矩阵形式描述正交性质。

$$Q = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} \quad Q^T Q = \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} = I$$

投影矩阵 P ，最小二乘法，在方程无解时求“最优解”。

Gram-Schmidt 正交化——从任意一组基得到标准正交基，策略是从向量中减去投影到其它向量方向的分量。

- 行列式 $\det(\mathbf{A})$

三个性质定义了行列式，可以推导出之后的性质 4~10。

行列式展开公式包含 $n!$ 个非零项，一半取+，一半取-。

代数余子式公式，以及推导出逆矩阵公式 $\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \mathbf{C}^T$ 。

- 特征值 $\mathbf{Ax} = \lambda \mathbf{x}$

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0$$

对角化：如果矩阵 \mathbf{A} 包含 n 个线性无关的特征向量，可以对角化得到

$$\mathbf{S}^{-1} \mathbf{AS} = \mathbf{\Lambda}$$

矩阵 \mathbf{A} 的幂： $\mathbf{A}^k = (\mathbf{S} \mathbf{\Lambda} \mathbf{S}^{-1})^k = \mathbf{S} \mathbf{\Lambda}^k \mathbf{S}^{-1}$

例题

1. $\mathbf{a} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$

a) 求投影到向量 \mathbf{a} 方向的投影矩阵 P 。

解： $\mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ ，这里 \mathbf{a} 是向量，因此有 $\mathbf{P} = \frac{\mathbf{a} \mathbf{a}^T}{\mathbf{a}^T \mathbf{a}} = \frac{1}{9} \begin{bmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{bmatrix}$

b) 求矩阵 P 的秩

解：矩阵的秩为 1，因为所有的列都是第二列的倍数。或者从投影矩阵投影的空间是 1 维可以判断出来。

c) 矩阵 P 的列空间？

解：向量 a 所在的直线。

d) 矩阵 P 的特征值？

解：矩阵的秩为 1，因此存在重特征值 0，再从矩阵的迹可以得到另一个特征值为 1。即特征值为 1,0,0。

e) 求矩阵 P 对应特征值 1 的特征向量？

解：因为矩阵 P 为投影矩阵，在投影空间中的向量就是对应特征值 1 的特征向量，因此 a 就是对应的特征向量。

f) 若有 $u_{k+1} = Pu_k$ ，且有初值 $u_0 = \begin{bmatrix} 9 \\ 9 \\ 0 \end{bmatrix}$ 。求 u_k ？

解：重复将一个向量投影到一条直线，从第二次开始投影结果即不发生变化。

$$u_k = P^k u_0 = P u_0 = a \frac{a^T u_0}{a^T a} = 3a = \begin{bmatrix} 6 \\ 3 \\ 6 \end{bmatrix}$$

g) 测验中可能出现 $u_{k+1} = Au_k$ ，其中的矩阵 A 不是投影矩阵，没有投影矩阵的性质 $P^k u_0 = Pu_0$ ，此时需要求矩阵的特征值和特征向量。

$$u_0 = C_1 x_1 + C_2 x_2 + C_3 x_3$$

$$\text{则 } u_k = C_1 \lambda_1^k x_1 + C_2 \lambda_2^k x_2 + C_3 \lambda_3^k x_3$$

对于投影矩阵而言， $\lambda_1 = 1$ ， $\lambda_2 = \lambda_3 = 0$ 。因此有 $u_1 = u_2 = u_3 = \dots$

2. 已知以下数据点

t	y
1	4
2	5
3	8

a) 求利用三个数据点拟合过原点的一条直线 $y = Dt$ ？

解： $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} D = \begin{bmatrix} 4 \\ 5 \\ 8 \end{bmatrix}$ ，我们求解的问题就是求方程的最优解 \hat{D} 。 $A^T A \hat{D} = A^T b$ 。

解得 $\hat{D} = 19/7$ 。因此直线解析式为 $y = (19/7)t$ 。

b) 怎样从投影来理解这个问题？

解：对于最小二乘问题有两种理解方法。其中一种是找到平面内最优的一条直

线。另一种是将 $\mathbf{b} = \begin{bmatrix} 4 \\ 5 \\ 8 \end{bmatrix}$ 投影到 \mathbf{A} 的列空间，来得到最接近于 $\mathbf{Ax}=\mathbf{b}$ 的解。

3. 向量 $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ 和 $\mathbf{a}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ 向量确定了一个平面，找到该平面的一组正交基。

解：应用 Gram-Schmidt 正交化方法。取 \mathbf{a}_1 为第一个方向，找出垂直于该方向的向量 \mathbf{B} 。策略就是在 \mathbf{a}_2 中减掉它在 \mathbf{a}_1 方向上的分量，得到完全垂直于 \mathbf{a}_1 的部分：

$$\mathbf{B} = \mathbf{a}_2 - \frac{\mathbf{a}_1^T \mathbf{a}_2}{\mathbf{a}_1^T \mathbf{a}_1} \mathbf{a}_1 = \begin{bmatrix} 4/7 \\ 1/7 \\ -2/7 \end{bmatrix}$$

4. 已知一个 4 阶方阵 \mathbf{A} 具有特征值 $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 。

a) 特征值需要满足什么条件才能保证 \mathbf{A} 为可逆矩阵。

解：当且仅当矩阵的特征值均不为 0 的时候， \mathbf{A} 为可逆矩阵。若有特征值 0 存在，则矩阵零空间有非零向量，矩阵不可逆。

b) 求逆矩阵行列式的值？

解： \mathbf{A}^{-1} 的特征值为 \mathbf{A} 特征值的倒数。因此 $\det(\mathbf{A}^{-1}) = \frac{1}{\lambda_1} \frac{1}{\lambda_2} \frac{1}{\lambda_3} \frac{1}{\lambda_4}$

c) 求 $(\mathbf{A} + \mathbf{I})$ 的迹？

解：矩阵的迹等于其特征值的和， $(\mathbf{A} + \mathbf{I})$ 的特征值为 $\lambda_1 + 1, \lambda_2 + 1, \lambda_3 + 1, \lambda_4 + 1$ ，则该矩阵的迹为 $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + 4$ 。

5. 已知三对角矩阵：

$$\mathbf{A}_4 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

令 $D_n = \det(\mathbf{A}_n)$

a) 用代数余子式的方法求出 $D_n = aD_{n-1} + bD_{n-2}$ 中的 a, b ?

解：用代数余子式可得 $D_4 = (1)D_3 + (-1)D_2$ 。

b) 利用找到的递归方程 $D_n = aD_{n-1} + bD_{n-2}$ ，求 D_n ?

解：把它当作方程组来求解，我们首先解得初值 $D_1 = 1, D_2 = 0$ 。

按照递归方程构造 2 阶线性方程： $\begin{bmatrix} D_n \\ D_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} D_{n-1} \\ D_{n-2} \end{bmatrix}$ 。求解矩阵的特征

值得： $\lambda = \frac{1 \pm \sqrt{3}i}{2} = e^{\pm i\pi/3}$ ，均为模为 1 的复数，在单位圆上旋转。可以看到

$\lambda_1^6 = \lambda_2^6 = 1$ ，矩阵经六次变换变为单位阵。该序列既不发散也不收敛，数列以 6 次为重复周期不停循环，1,0,-1,0,1,1。

6. 有一组对称矩阵：

$$A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad A_3 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix} \quad A_4 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 3 & 0 \end{bmatrix}$$

a) 找到投影到矩阵 A_3 列空间的投影矩阵 P 。

解：矩阵 A_3 为奇异阵，其第 3 列列向量为第 1 列的 3 倍，而第 1,2 列线性无关，因此其列空间为一个平面，取其前两列列向量组成矩阵 A ，则有

$$P = A(A^T A)^{-1} A^T = \begin{bmatrix} 1/5 & 0 & 2/5 \\ 0 & 1 & 0 \\ 2/5 & 0 & 4/5 \end{bmatrix}$$

可以通过将矩阵 A_3 的列向量乘以投影矩阵来验证这一结果，在平面内的向量投影后应该不发生变化。

b) 求 A_3 的特征值和特征向量？

解：

$$\det(A_3 - \lambda I) = \begin{vmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 2 \\ 0 & 2 & -\lambda \end{vmatrix} = -\lambda^3 + 5\lambda = 0$$

解得三个特征值， $\lambda_1 = 0$ ， $\lambda_2 = \sqrt{5}$ ， $\lambda_3 = -\sqrt{5}$ 。可用矩阵的迹做检查。

$$\text{求解 } (A_3 - \lambda I) \mathbf{x} = \mathbf{0}, \text{ 可得特征向量 } \mathbf{x}_1 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ -\sqrt{5} \\ 2 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 1 \\ \sqrt{5} \\ 2 \end{bmatrix}.$$

c) 找到投影到矩阵 A_4 列空间的投影矩阵 P ？

解：因为 A_4 为可逆矩阵，所以它的列空间就是整个 \mathbf{R}^4 空间，所以 $P = I$ 。而可逆矩阵的证明可以采用求行列式的办法，用代数余子式展开可得 $\det(A_4) = 9$ 。

第 25 讲 对称矩阵和正定性

Symmetric matrices and positive definiteness

对称矩阵是最重要的矩阵之一，其特征值为实数并且拥有一套正交特征向量。正定矩阵的性质则更好。

对称矩阵 Symmetric matrices

包含特殊性质的矩阵，例如 Markov 矩阵，其特征值和特征向量往往拥有一定特性。对称矩阵 $A=A^T$ 的特征值为实数，具有完全正交的特征向量。这里的“有”，是指可以选出一套完全正交的特征向量（例如在重特征值条件下，可能存在一个平面内向量都可以作为特征向量）。

如果 A 具有 n 个线性无关的特征向量，可以对角化得到 $A=SMS^{-1}$ 。而对于对称矩阵， $A=Q\Lambda Q^{-1}=Q\Lambda Q^T$ ，其中 Q 为正交矩阵，列向量为标准正交基，这个公式本身还显示了矩阵的对称性。矩阵能够进行这种分解，在数学上称为“谱定理”（spectral theorem），将特征值视为“谱”，在物理上称之为“主轴定理”。

实特征值 Real eigenvalues

矩阵 A 具有特征值 λ 和特征向量 x ，则有 $Ax=\lambda x$ 。若为 λ 复数，则其共轭复数满足 $A\bar{x}=\bar{\lambda}\bar{x}$ 。两侧取转置，则有：

$$\bar{x}^T A^T = \bar{x}^T \bar{\lambda},$$

两侧同乘 x ，则：

$$\bar{x}^T A^T x = \bar{x}^T \bar{\lambda} x,$$

因为矩阵为对称矩阵，所以有：

$$\bar{x}^T \lambda x = \bar{x}^T \bar{\lambda} x。$$

因此若有 $\bar{x}^T x \neq 0$ 则 $\lambda = \bar{\lambda}$ ，即 λ 为实数。

因为 \bar{x} 为 x 的共轭，因此两者点积 $\bar{x}^T x$ 得到的是 x 中的每一分量与其共轭复数的乘积之和，即复向量 x 的模，当且仅当 $x=0$ 时，其模为 0。

$$\bar{x}^T x = [\bar{x}_1 \quad \bar{x}_2 \quad \bar{x}_3 \quad \bar{x}_4] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = |x_1|^2 + |x_2|^2 + |x_3|^2 + |x_4|^2$$

因此得证 λ 为实数。

对于复矩阵 A ，若有 $\bar{A}^T = A$ 则上面的推导全部成立，因此共轭对称复矩阵 A 的特征值为实数。

对于对称矩阵， $A=Q\Lambda Q^{-1}=Q\Lambda Q^T$ ，可以写作：

$$A = Q\Lambda Q^T = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_n] \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix}$$

$$= \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^T + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T$$

矩阵是 $\mathbf{q}_k \mathbf{q}_k^T$ ($= \mathbf{q}_k \mathbf{q}_k^T / (\mathbf{q}_k^T \mathbf{q}_k)$) 是朝向向量 \mathbf{q}_k 的投影矩阵, 所以每一个对称矩阵都是正交投影矩阵的线性组合。这是理解谱定理的另一种方法。

当确认矩阵特征值为实数后, 下一个要考虑的问题就是它是正还是负数, 因为这影响着微分方程中体系的稳定与否。但是对于大型矩阵, 通过计算 $|\mathbf{A} - \lambda \mathbf{I}| = 0$ 得到特征值进行判定难以实现, 即使用 MATLAB 求解, 结果也不一定可靠, 但 MATLAB 可以得到矩阵的主元, 而对称阵的主元中正负数的个数与特征值相同, 即正主元的数目等于正特征值的数目。

对于对称阵主元与特征值符号相匹配这件事情, 通常是用合同矩阵的惯性定理加以证明。若存在可逆矩阵 \mathbf{C} 满足 $\mathbf{A} = \mathbf{C} \mathbf{B} \mathbf{C}^T$, 就称 \mathbf{A} 与 \mathbf{B} 为合同矩阵, 惯性定理是指矩阵的特征值符号经过合同变换后不发生变化。对称阵 \mathbf{A} 经过消元操作可得 $\mathbf{A} = \mathbf{L} \mathbf{D} \mathbf{L}^T$, 而经过对角化处理得到 $\mathbf{A} = \mathbf{Q} \Lambda \mathbf{Q}^T$, 两者比较可知 $\mathbf{D} = (\mathbf{L}^{-1} \mathbf{Q}) \Lambda (\mathbf{L}^{-1} \mathbf{Q})^T$, 因此两个对角阵为合同矩阵, 主元和特征值的符号匹配。

GS 在书里给了一个接近同伦理论的证明:

对于矩阵 $\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & \\ & -8 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$, 在下三角阵转变为 $\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 的过程中, 矩阵的变化为 $\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 \\ 0 & -8 \end{bmatrix}$, 矩阵的特征值也从 4, -2 变为 1, -8

(即原矩阵的主元)。这个过程中如果发生符号变化, 则总有一个瞬间特征值穿过 0 点, 那个瞬间矩阵应该是奇异阵, 但是相乘的三个矩阵一直是非奇异阵, 因此不可能有这样一个时刻存在, 所以特征值不会发生符号变化, 即和主元符号保持一致。

矩阵 $\mathbf{A} + b\mathbf{I}$ 的特征值比矩阵 \mathbf{A} 的特征值大 b , 可以通过 $\mathbf{A} + b\mathbf{I}$ 的主元来了解矩阵 \mathbf{A} 的特征值与 b 的大小关系, 因此利用这个性质可以估计特征值的状态。

正定矩阵 Positive definite matrices

正定矩阵 \mathbf{A} 是特征值都为正数的对称矩阵。它的主元也均为正数。

例如矩阵 $\mathbf{A} = \begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}$ 。主元为 5 和 $(\det \mathbf{A})/5 = 11/5$ 。主元都为正数, 且本身为对称矩阵, 因此 \mathbf{A} 是正定矩阵。计算其特征值为: $|\mathbf{A} - \lambda \mathbf{I}| = \lambda^2 - 8\lambda + 11 = 0$, $\lambda = 4 \pm \sqrt{5}$ 。

因为主元和特征值都是正的, 可以知道正定矩阵的行列式也是正的。但是反之, 并不成立。矩阵 $\begin{bmatrix} -1 & 0 \\ 0 & -3 \end{bmatrix}$ 的行列式也是正的, 但不是正定矩阵。

若将行列式作为正定的判据，则要求 n 阶矩阵左上角的所有 $k \times k$ ($1 \leq k \leq n$) 子行列式 (subdeterminant) 数值均为正，矩阵才能确定为正定矩阵。

本讲的内容将之前教授的主元、行列式和特征值的概念结合在了一起，对于正定矩阵这些都是正的，当完全掌握了它们的性质后会推广到非对称矩阵，甚至非方阵。

第 26 讲 复矩阵；快速傅里叶变换

Complex matrices; fast Fourier transform

实矩阵也可能有复特征值，因此无法避免在矩阵运算中碰到复数，本讲学习处理复数矩阵和复向量。

最重要的复矩阵是傅里叶矩阵，它用于傅里叶变换。而对于大数据处理快速傅里叶变换（FFT）显得更为重要，它将傅立叶变换的矩阵乘法中运算的次数从 n^2 次降至 $n \log_2 n$ 次。

复向量 Complex vectors

对于给定的复向量 $\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \in \mathbb{C}^n$ ，其元素中有复数，因此 $\mathbf{z}^T \mathbf{z}$ 无法给出向量的长

度。例如： $\begin{bmatrix} 1 & i \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = 0$ 。则定义 $|\mathbf{z}|^2 = \bar{\mathbf{z}}^T \mathbf{z} = |z_1|^2 + |z_2|^2 + \cdots + |z_n|^2$ 为向量长度。因此

向量 $\begin{bmatrix} 1 \\ i \end{bmatrix}$ 的长度就是 $\begin{bmatrix} 1 & -i \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = 2$ 。记 $|\mathbf{z}|^2 = \bar{\mathbf{z}}^T \mathbf{z} = \mathbf{z}^H \mathbf{z}$ 。H 来自于 “Hermite”。

与之相似，内积的定义也变为 $\mathbf{y}^H \mathbf{x} = \bar{\mathbf{y}}^T \mathbf{x} = \bar{y}_1 x_1 + \bar{y}_2 x_2 + \cdots + \bar{y}_n x_n$ 。

复矩阵 Complex matrices

上一讲中讲到了对于复矩阵 \mathbf{A} ，若有 $\bar{\mathbf{A}}^T = \mathbf{A}$ 则复矩阵 \mathbf{A} 的特征值为实数。这种复矩阵被称为埃尔米特矩阵（Hermitian matrixes，又译作“厄米特矩阵”或者“厄米矩阵”）。转置共轭记作 $\mathbf{A}^H = \bar{\mathbf{A}}^T$ 。

例如矩阵 $\begin{bmatrix} 2 & 3+i \\ 3-i & 5 \end{bmatrix}$ 为埃尔米特矩阵。它具有实数特征值和正交的特征向量。

由性质可知埃尔米特矩阵对角线均为实数。

此处向量标准正交的意思是 $\bar{\mathbf{q}}_j^T \mathbf{q}_k = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases}$ 。用 n 个标准正交的复向量作为

列向量可以构造一个矩阵 \mathbf{Q} ，则有 $\mathbf{Q}^T \mathbf{Q} = \mathbf{I} = \mathbf{Q}^H \mathbf{Q}$ 。这个复空间的正交矩阵称为酉矩阵（unitary matrix）。

傅里叶变换 Fourier transform

傅里叶级数是将周期函数或者信号变换为不同频率的三角函数的和函数。

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots$$

在电子工程或者计算机科学中 $n \times n$ 矩阵的行和列从第 0 行和第 0 列开始计数，最后到第 $n-1$ 行和第 $n-1$ 列。我们在讨论傅里叶矩阵的时候遵从这种习惯。

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi nk/N} = \sum_{n=0}^{N-1} x_n w^{nk}, k = 0, 1, 2, \dots, N-1 \quad (1)$$

其中 $w = e^{-j2\pi/N}$ ，公式 (1) 对一个序列 $[x_0, x_1, \dots, x_{N-1}]$ 做 N 点傅里叶变换。我们将这个公式写成矩阵乘以向量的形式，令 $\hat{x} = [X_0, X_1, \dots, X_{N-1}]^T$ ， $x = [x_0, x_1, \dots, x_{N-1}]^T$ ，即：

$$\begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ \vdots \\ X_{N-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & w & w^2 & \dots & w^{N-1} \\ 1 & w^2 & w^4 & \dots & w^{2(N-1)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & w^{N-1} & w^{2(N-1)} & \dots & w^{(N-1)(N-1)} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{bmatrix} \quad (2)$$

公式 (2) 用矩阵变量表示为：

$$\hat{x} = Fx$$

$$F_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} \quad (3)$$

从矩阵可以得到一个四点（离散的）傅里叶变换，它的逆矩阵就是反傅里叶变换。因为傅里叶矩阵列向量正交，所以其逆矩阵很容易计算。实际上这个矩阵可以分解成一系列稀疏矩阵，并且它们的逆矩阵都很容易得到。

计算可知列向量的模不是 1，矩阵除以 2 之后，向量标准正交： $\frac{1}{4} F_4^H F_4 = I$ 。它的逆矩阵就是共轭转置。

快速傅里叶变换 Fast Fourier transform

64 阶傅里叶矩阵 F_{64} 中的 ω_{64} 与 32 阶傅里叶矩阵 F_{32} 的元素 ω_{32} 相比，幅角是其一半， $(\omega_{64})^2 = \omega_{32}$ 。可以从分块矩阵运算找到两者的联系：

$$[F_{64}] = \begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} F_{32} & 0 \\ 0 & F_{32} \end{bmatrix} [P]$$

其中 P 是置换矩阵，而 D 为对角矩阵

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & & & 0 & 0 \\ & & \vdots & & & & & \\ 0 & 0 & 0 & 0 & \dots & \dots & 1 & 0 \\ 0 & 1 & 0 & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & \dots & 0 & 0 \\ & & \vdots & & & & & \\ 0 & 0 & 0 & 0 & \dots & \dots & 0 & 1 \end{bmatrix}, D = \begin{bmatrix} 1 & & & & & & & \\ & \omega & & & & & & \\ & & \omega^2 & & & & & \\ & & & \ddots & & & & \\ & & & & \omega^{31} & & & \end{bmatrix}$$

P 的效果是使得所乘的向量 x 中序数为奇数的分量如 x_1, x_3, x_5 等提到前面，而偶数分量 x_2, x_4 等放到后面。

计算 64 阶傅里叶变换（傅里叶矩阵乘以向量）的计算量是 64^2 ，而等式右侧的计算量是 2×32^2 （两个 32 阶的傅里叶矩阵）再加上一些修正项，修正项主要来自

于与对角矩阵 \mathbf{D} 的乘法，大约为 32 次。继续对 \mathbf{F}_{32} 进行分解，计算的运算量再一次下降变为 $2 \times (2 \times 16^2 + 16) + 32$ 。连续进行拆分，傅里叶矩阵的尺寸变化依次为 64、32、16、8、4、2、1，经过 $\log_2 64$ 次分解，最后仅剩修正项的运算， $\log_2 64 \times 32$ 次。对于 n 阶矩阵，可将 n^2 次计算降至 $(n/2) \log_2 n$ 。例如对于 1024 阶矩阵，运算量从 1024×1024 降至 5×1024 。

第 27 讲 正定矩阵和最小值

Positive definite matrices and minima

本讲学习正定矩阵，这部分内容将本课程之前的知识点：主元、行列式、特征值以及方程的稳定性融为一体。本讲介绍如何判定一个矩阵是否正定矩阵，以及当一个矩阵是正定矩阵时，其内涵和矩阵操作的效果有何特别之处。此外还有正定矩阵与几何的关系：椭圆和正定有关，双曲线与正定无关。

正定矩阵 Positive definite matrices

给定一个 2×2 矩阵，有四个途径判定矩阵是否正定矩阵：

- 1) 特征值： $\lambda_1 > 0, \lambda_2 > 0$
- 2) 行列式（所有子行列式）： $a > 0, ac - b^2 > 0$
- 3) 主元： $a > 0, (ac - b^2)/a > 0$
- 4) 表达式 $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ ($\mathbf{x} = 0$ 除外)。通常这就是正定的定义，而前三条是用来验证正定性的条件。

给定矩阵 $\mathbf{A} = \begin{bmatrix} 2 & 6 \\ 6 & y \end{bmatrix}$ ，从判据可知矩阵为正定阵的条件是 $2y - 36 > 0$ ，即 $y > 18$ 。

矩阵 $\begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix}$ 正好处在判定为正定矩阵的临界点上，称之为半正定 (positive semidefinite) 矩阵，它具有一个特征值 0，是奇异矩阵，只有一个主元，行列式为 0。半正定矩阵特征值大于等于 0。

再观察 $\mathbf{x}^T \mathbf{A} \mathbf{x}$ 判据：

$$\begin{aligned} \mathbf{x}^T \mathbf{A} \mathbf{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= 2x_1^2 + 12x_1x_2 + 18x_2^2 \end{aligned}$$

之前讨论得都是线性方程 $\mathbf{A} \mathbf{x}$ ，现在引入 \mathbf{x}^T ，变成二次，如果对于任意 x, y ，这种二次型 (quadratic form) $ax^2 + 2bxy + cy^2$ 均大于零，则矩阵为正定矩阵。

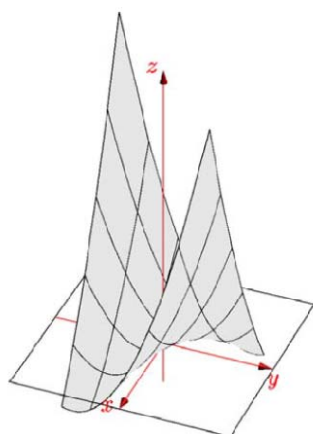
在本例的半正定矩阵中，当 $x_1 = 3, x_2 = -1$ 时 $2x_1^2 + 12x_1x_2 + 18x_2^2 = 2(x_1 + 3x_2)^2 = 0$ 。

如果将矩阵变为 $\begin{bmatrix} 2 & 6 \\ 6 & 7 \end{bmatrix}$ ，二次型为 $f(x, y) = 2x^2 + 12xy + 7y^2$ ，从几何图像上看没有最小值点，在 origin 处有一鞍点。鞍点在某个方向上看是极大值点，在另一方向上是极小值点，实际上最佳观测角度是特征向量的方向。

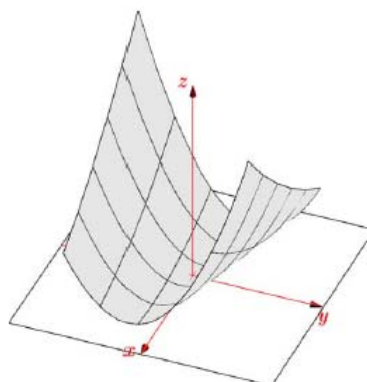
如果将矩阵变为 $\begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix}$ ，主元为正；特征值之积为行列式的值 4，和为矩阵的迹 22，因此特征值为正；子行列式均为正。矩阵为正定矩阵。

二次型 $f(x, y) = 2x^2 + 12xy + 20y^2$ ，其图像最小值点为原点，一阶偏导数为 0，

二阶偏导数为正。

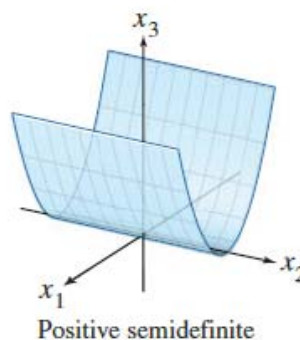
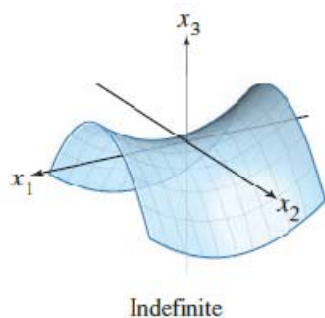
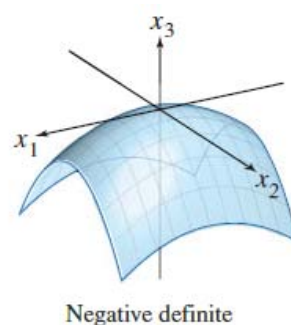
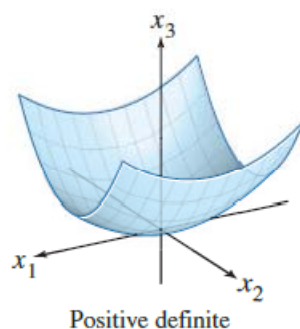


$f(x, y) = 2x^2 + 12xy + 7y^2$ 的图像



$f(x, y) = 2x^2 + 12xy + 20y^2$ 的图像

当年上课时候老师讲双曲面的鞍点,画的图不好,有个篮球队的哥们理解不了,于是下了课跑去问老师,老师指着他的热身篮球裤说,就你这种裤腿两侧能撕开的篮球裤,全部打开,拎起来,裤裆地方就是鞍点.....算了,反正 GS 画图真心不行,看看 Lay 的吧:



微积分中判定最小值点的判据:一阶导数等于零 $\frac{du}{dx} = 0$, 二阶导数为正 $\frac{d^2u}{dx^2} > 0$ 。

线性代数中判据为二阶导数矩阵正定。

对于二次型我们可以用配方的办法来验证其是否具有最小值:

$$f(x, y) = 2x^2 + 12xy + 20y^2 = 2(x + 3y)^2 + 2y^2$$

配方使得 x^2 的系数和交叉项 xy 的系数配合形成完全平方的形式,这个时候用到的 y^2 的系数正好是 18, 即判定正定的临界点。如果实际的系数 d 大于 18, 则还

剩余 $(d-18)y^2$ 二次型在原点之外一定大于零 若小于 18 则二次型可以小于等于 0。

对于 $f(x, y) = 2x^2 + 12xy + 20y^2 = 2(x+3y)^2 + 2y^2$ ，其几何图像为碗型的曲面，如果我们用 $f=1$ 的截面横截曲面，得到的就是 $2(x+3y)^2 + 2y^2 = 1$ 的椭圆曲线。而对于双曲面进行切割就得到双曲线。

配方法其实就是消元：

$$\begin{bmatrix} 2x^2 & 6xy \\ 6xy & 20y^2 \end{bmatrix}, \begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 6 \\ 0 & 2 \end{bmatrix}. \quad f(x, y) = 2(x+3y)^2 + 2y^2$$

主元就是平方项系数， L 矩阵中的行操作数 l_{21} 就是配方项内 y 的系数。因此这就是为什么主元为正则矩阵为正定矩阵，因为主元是每一个完全平方项的系数。本例中二次型表达式的配方说明了二维的情形，而线代的理论可以将之推广到 n 维。

二阶导数的矩阵记为 $\begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$ ，矩阵对称代表交叉二阶偏导数与求导顺序无关 $f_{xy} = f_{yx}$ 。在微积分中我们学到的判据 $f_{xx}f_{yy} > f_{xy}^2$ ，和二阶矩阵判定正定是等价的，并且线代可以推广到 n 维。

3 阶矩阵 $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ ，它是正定矩阵。计算子行列式得到 $|2| = 2$ ， $\begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3$ ， $\begin{vmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{vmatrix} = 4$ 。主元是 2，3/2，4/3。特征值是 $2 - \sqrt{2}$ ，2， $2 + \sqrt{2}$ 。

这是 GS 最爱的矩阵之一，可以用来把二阶微分方程变成离散问题，因为它每一行都是差分方程 $f_{n+1} - 2f_n + f_{n-1}$ 。

其二次型为 $\mathbf{x}^T A \mathbf{x} = 2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3$ 。

这是一个四维的图像，三个维度 x_1, x_2, x_3 ，还有函数 f ，如果用 $f=1$ 切割图像，则得到 $2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3 = 1$ 。这是一个椭球体，三个特征值不同，因此椭球的三个长轴长度不同。三个轴的方向就是特征向量的方向，轴长度就是特征值，矩阵的分解 $A = Q\Lambda Q^T$ 很好的说明了这件事，这就是所谓的“主轴定理”。

对于三条判据可以判定正定： $\mathbf{x}^T A \mathbf{x} > 0$ ($\mathbf{x} = 0$ 除外)。已经分析了主元要大于零的原因，因为它是配方后的参数，只有都大于零才能保证正定。以下对于判据 1 和 2 做简要说明。

对称矩阵 A ，其正交的特征向量可以张成整个空间，因此任意向量 \mathbf{x} 均可表示成特征向量的线性组合 $\mathbf{x} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \dots + c_n\mathbf{x}_n$ ，

代入得 $\mathbf{x}^T A \mathbf{x} = c_1^2\lambda_1 + c_2^2\lambda_2 + \dots + c_n^2\lambda_n$ ，当特征值都大于零且 $\mathbf{x} \neq 0$ 时，才能保证 $\mathbf{x}^T A \mathbf{x} > 0$ 。因此条件 1 与正定性等价。

记 \mathbf{A}_k 为矩阵 \mathbf{A} 左上角 k 阶方块，取特殊向量 $\mathbf{x} = \begin{bmatrix} x_k \\ 0 \end{bmatrix}$ ，即后 $n-k$ 个元素为 0，则有 $\mathbf{x}^T \mathbf{A} \mathbf{x} = \begin{bmatrix} x_k & 0 \end{bmatrix} \begin{bmatrix} \mathbf{A}_k & * \\ * & * \end{bmatrix} \begin{bmatrix} x_k \\ 0 \end{bmatrix} = \mathbf{x}_k^T \mathbf{A}_k \mathbf{x}_k$ 。若矩阵 \mathbf{A} 满足正定性，则所有 \mathbf{A}_k 均满足正定性。已证明正定性等价于特征值均为正，而矩阵行列式等于特征值之积，因此可知子行列式均大于零。反之亦成立，两命题等价。

第 28 讲 相似矩阵和若尔当标准型

Similar matrices and Jordan form

本讲介绍相似矩阵，这些内容以及奇异值分解是线性代数最核心的概念。

正定矩阵 $A^T A$

若矩阵 A 满足对任意向量 $x \neq 0$ 均有 $x^T A x > 0$ ，则称矩阵为正定矩阵，可以通过特征值、主元和行列式的办法来判断矩阵的正定性。

正定矩阵来自于最小二乘问题。有大量的实际问题用到了长方形矩阵，而最小二乘问题中用到了长方形矩阵的积 $A^T A$ ，它是正定矩阵。

正定矩阵 A 是对称矩阵，它的逆矩阵 A^{-1} 也是正定矩阵，逆矩阵的特征值是原矩阵的倒数，因此也都是正数。若矩阵 A 和 B 都是正定矩阵，则 $A+B$ 也是正定矩阵： $x^T A x > 0$ ， $x^T B x > 0$ ，则有 $x^T (A+B) x > 0$ 。

如果 A 是一个 $m \times n$ 长方形矩阵，则 $A^T A$ 是对称方阵。通过讨论 $x^T (A^T A) x$ 的正负可以确认它是正定矩阵： $x^T (A^T A) x = (Ax)^T (Ax) = |Ax|^2 \geq 0$ 。当且仅当 $Ax=0$ 时，表达式为 0。当矩阵 A 的各列线性无关时，即矩阵为列满秩 $r=n$ ， A 的零空间只有零向量，即此条件下仅有零向量，满足 $x^T (A^T A) x = 0$ 。因此矩阵列满秩时， $A^T A$ 是正定矩阵。正定矩阵将之前的知识点串联起来。

相似矩阵 Similar matrices

A 和 B 均是 $n \times n$ 方阵，若存在可逆矩阵 M ，使得 $B = M^{-1} A M$ ，则 A 和 B 为相似矩阵。

特征值互不相同 Distinct eigenvalues

若矩阵 A 具有 n 个线性无关的特征向量，可以对角化得到 $S^{-1} A S = \Lambda$ ，则 A 相似于 Λ ，这里的 M 是特征向量矩阵 S 。如果将 M 取其它可逆矩阵，可以得到和 A 相似的另一矩阵 B ，实际上这样可以定义一类矩阵， Λ 是其中最简洁的一个。

$$\text{例：} A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \text{ 则 } \Lambda = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}, \text{ 而取另一 } M, \text{ 则有}$$
$$B = \begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & -15 \\ 1 & 6 \end{bmatrix}.$$

相似矩阵最重要的特性是：**相似矩阵具有相同的特征值**。事实上，所有特征值为 3 和 1 的二阶矩阵都是 A 的相似矩阵。

证明矩阵 A 的相似矩阵 $B = M^{-1} A M$ ，具有和矩阵 A 相同的特征值 λ ：矩阵 A 具有的特征值 λ ，即存在特征向量 x 满足 $Ax = \lambda x$ 。则有：

$$\begin{aligned} A M M^{-1} x &= \lambda x \\ M^{-1} A M M^{-1} x &= \lambda M^{-1} x \end{aligned}$$

$$BM^{-1}x = \lambda M^{-1}x$$

即矩阵具有特征值 λ ，且特征向量为 $M^{-1}x$ 。

因此，相似矩阵具有相同的特征值，并且线性无关的特征向量的个数相同，但是特征向量往往不同。如果矩阵 A 的特征值互不相等 $\lambda_1 \neq \lambda_2 \neq \dots \neq \lambda_n$ ，而与另一个矩阵 B 的特征值完全相同 $\lambda_1 = \lambda_1', \lambda_2 = \lambda_2', \dots, \lambda_n = \lambda_n'$ ，则它们与相同的对角矩阵 Λ 相似。

重特征值 Repeated eigenvalues

如果矩阵有重特征值，则可能无法进行对角化。

例：二阶矩阵有重特征值 $\lambda_1 = \lambda_2 = 4$ 。

第一类： $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ 只与自己相似， $M^{-1} \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} M = 4M^{-1}IM = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ 。这个系列的相似矩阵仅包含其自身。

第二类包含其它所有的重特征值为4的矩阵：其中最简洁的是 $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ ，元素1的位置换上其它数值仍然是相似矩阵。这个最优形式称为若尔当（Jordan form）标准型。有了这个理论，就可以处理不可对角化的矩阵，完成近似的“对角化”转化为若尔当标准型进行处理。

与 $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ 相似的矩阵，迹为8，行列式为16，因此我们可以构造出很多相似矩阵： $\begin{bmatrix} 5 & 1 \\ -1 & 3 \end{bmatrix}$ ， $\begin{bmatrix} 4 & 0 \\ 17 & 4 \end{bmatrix}$ ， $\begin{bmatrix} a & * \\ * & 8-a \end{bmatrix}$它们都不能对角化（因为若可以对角化则按照特征值可知结果为 $4I$ ，而 $4I$ 只与自己相似）。

若尔当标准型 Jordan form

更复杂的情况，一个四阶矩阵具有重特征值0， $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0$ 。

$A = \left[\begin{array}{cccc|c} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]$ 它的秩为2，因此其零空间的维数为 $4-2=2$ ，而零空间的向

量就是矩阵的特征向量，满足 $Ax=0x$ ，所以矩阵 A 只有两个特征向量。若尔当指出上对角线每增加一个1，矩阵就减掉一个特征向量，本例中特征向量数为 $4-2=2$ 。

矩阵 $B = \left[\begin{array}{cccc|c} 0 & 1 & 7 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]$ 与矩阵 $A = \left[\begin{array}{cccc|c} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]$ 为相似矩阵。

但矩阵 $C = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array} \right]$ 与 $A = \left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array} \right]$ 并不是相似矩阵，两者具有不同

的若尔当块。若尔当块形如 $J_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & 0 & \lambda_i \end{bmatrix}$ ，对角线上为重特征值 λ_i ，上

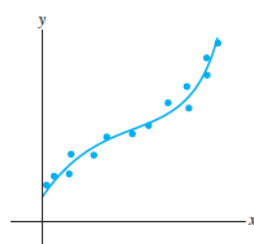
对角线为 1，其它位置的元素均为 0，每个若尔当块只有 1 个特征向量。若干个若尔当块可以拼成一个若尔当矩阵。

若尔当矩阵： $J = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & J_d \end{bmatrix}$

两个矩阵具有相同的特征值和特征向量个数，但是其若尔当块的尺寸不同，两者也并不是相似矩阵。如前述矩阵 A 与 B 并不相似。

若尔当理论：任意 n 阶矩阵 A 都与一个若尔当矩阵 J 相似。若尔当矩阵中的每一个若尔当块对应一个特征向量。若矩阵具有 n 个不同的特征向量 则可以对角化，此时其若尔当标准型 J 就是对角矩阵 Λ 。若出现重特征值，则特征向量个数变少。

说到了 $A^T A$ 和最小二乘问题就要解释一下，GS 举得曲线拟合的例子，都是线性公式 $y=ax+b$ ，但实际上最小二乘法也处理非线性方程，因为这里所谓的非线性是对 x 而言，而只要对于所求的参数是线性方程就可以。比如下面的例子中 x 的方幂组成的矩阵 X 只是一个系数矩阵，对于所求的参数 β 这仍是个线性方程组。



Observation vector: $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, Design matrix: $X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}$, Parameter vector: $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$, Residual vector: $\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$

第 29 讲 奇异值分解

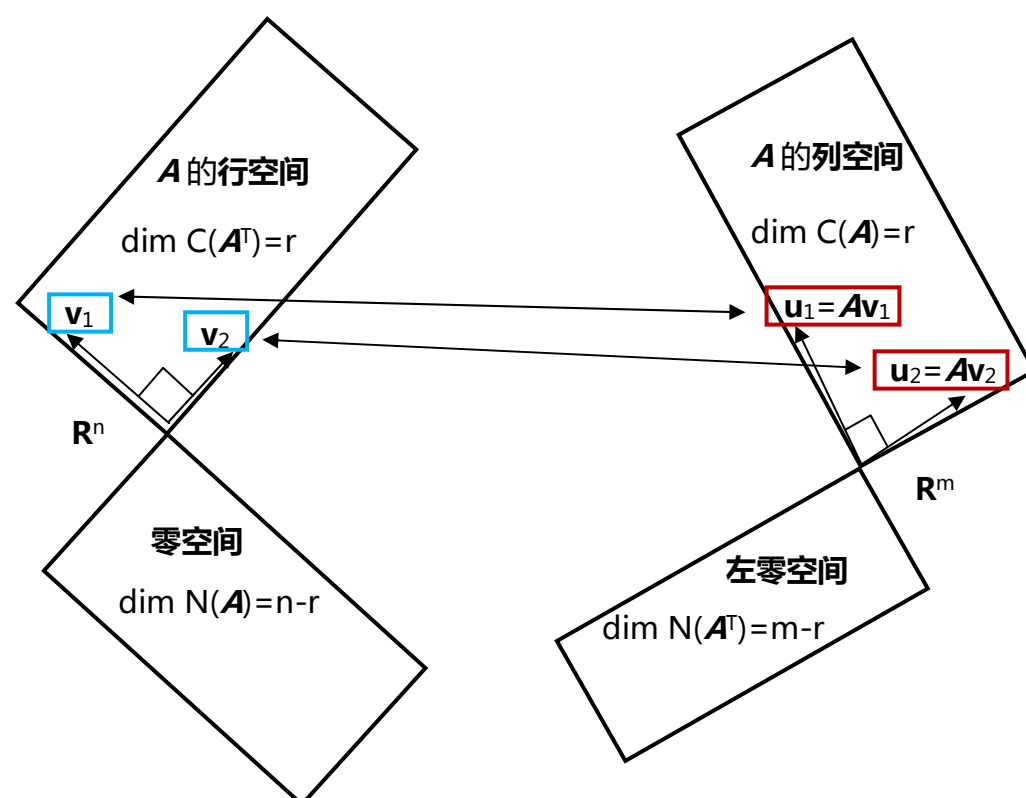
Singular value decomposition

本讲介绍奇异值分解 (Singular value decomposition), 简称 SVD。这是矩阵最终也是最好的分解, 任意矩阵可分解为 $A = U\Sigma V^T$, 分解结果为正交矩阵 U , 对角阵 Σ 和正交矩阵 V 。

如果矩阵 A 是正定矩阵, 它的奇异值分解就是 $A = Q\Lambda Q^T$, 一个正交矩阵 Q 可以满足分解, 而不需要两个。而对于可对角化的矩阵有 $A = SAS^{-1}$, 但特征向量矩阵 S 并不是正交矩阵, 而 SVD 中的 U 和 V 都是正交矩阵。

如何实现

可以将矩阵 A 视为一种线性变换操作, 将其行空间中的一个向量 v_1 变为其列空间中的向量 $u_1 = Av_1$ 。奇异值分解就是要在行空间中寻找一组正交基, 将其通过矩阵 A 线性变换生成列空间中的一组正交基 $Av_i = \sigma_i u_i$ 。



找出矩阵 A 行空间中的正交基很容易, Gram-Schmidt 正交化过程就可以做到, 但是随便的一组正交基经过矩阵 A 变换得到的向量并不一定正交, 因此满足此要求的行空间的正交基非常特殊。而矩阵 A 零空间的向量所对应的是矩阵 Σ 对角线上的 0 元素, 因此很容易处理。

用矩阵数学语言描述这一过程

问题的核心就是找到行空间中一组特殊的正交基：

$$\begin{aligned} A[\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_r] &= [\sigma_1 \mathbf{u}_1 \quad \sigma_2 \mathbf{u}_2 \quad \cdots \quad \sigma_r \mathbf{u}_r] \\ &= [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_r] \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} \end{aligned}$$

如果加入零空间的部分 等式就变为 $A\mathbf{V} = \mathbf{U}\Sigma$ 。其中零空间对应的正交基 $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$, 经过线性变换得到 $A\mathbf{v} = 0$, 对应 Σ 矩阵中对角线最后的特征值 $\sigma_{r+1} = \sigma_{r+2} = \dots = 0$ 。在等式 $A\mathbf{V} = \mathbf{U}\Sigma$ 两侧右乘 \mathbf{V}^{-1} 得到 $A = \mathbf{U}\Sigma\mathbf{V}^{-1} = \mathbf{U}\Sigma\mathbf{V}^T$ 。

现在的问题就是怎么找到符合要求的向量 \mathbf{v}_i 和 \mathbf{u}_i 。

为了得到这两个正交矩阵，考虑首先解决其中的一个，在等式 $A = \mathbf{U}\Sigma\mathbf{V}^T$ 两侧分别乘上等式 $A^T = \mathbf{V}\Sigma^T\mathbf{U}^T$ 两侧的项：

$$\begin{aligned} A^T A &= \mathbf{V}\Sigma^T\mathbf{U}^T\mathbf{U}\Sigma\mathbf{V}^T \\ &= \mathbf{V}\Sigma^T\Sigma\mathbf{V}^T \\ &= \mathbf{V} \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_r^2 \end{bmatrix} \mathbf{V}^T \end{aligned}$$

上式其实是正定矩阵 $A^T A$ 的正交分解， \mathbf{v}_i 就是矩阵 $A^T A$ 的特征向量， σ_i^2 就是矩阵 $A^T A$ 的特征值，奇异值 σ_i 要取正平方根。用同样的办法也可以求得 \mathbf{U} ，它的列向量就是矩阵 AA^T 的特征向量。

例 1：矩阵 $A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$ ，求其 SVD 分解。

矩阵为可逆矩阵，秩为 2，则需要在行空间中求得 $\mathbf{v}_1, \mathbf{v}_2$ ，列空间中求得 $\mathbf{u}_1, \mathbf{u}_2$ ，以及伸缩因子 σ_1, σ_2 。

计算得 $A^T A = \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} = \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}$ ，它的特征向量为 $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ 和 $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ 。标准化得到 $\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ ， $\mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$ ，求得 $\sigma_1^2 = 32$ ， $\sigma_2^2 = 18$ 。

求 \mathbf{U} 的过程可以利用矩阵 AA^T 。

$AA^T = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}$ ，求得 $\sigma_1^2 = 32$ ， $\sigma_2^2 = 18$ ，它的特征向量为 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 和 $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 。

但是我们直接将 $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 作为矩阵 U 代入，会产生错误 $U\Sigma V^T = \begin{bmatrix} 4 & 4 \\ 3 & -3 \end{bmatrix}$ 。这是因为确定特征向量的过程中，特征向量反向仍然符合要求，通过现在的方法无法确认向量的符号，但是一旦我们确认 v 的方向之后， u 的方向也就随之确定，将 v 代入 $AV = U\Sigma$ 计算 u 可以避免这种问题。 u 和 v 之间的符号联系在进行 AA^T 的计算时被切断了，而用 $AV = U\Sigma$ 可以避免此问题。

例 2：奇异阵 $A = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}$ ，求其 SVD 分解。

矩阵的秩为 1，行空间和列空间都是 1 维的。行空间和零空间可以找到一组正交基转换得到列空间和左零空间的一组正交基。

很容易确定 $v_1 = \begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix}$ ， $v_2 = \begin{bmatrix} 0.6 \\ -0.8 \end{bmatrix}$ ， $u_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ ， $u_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ 。

$A^T A = \begin{bmatrix} 4 & 8 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} = \begin{bmatrix} 80 & 60 \\ 60 & 45 \end{bmatrix}$ ，秩 1 矩阵，很容易求得 $\sigma_1^2 = 125$ ， $\sigma_2^2 = 0$ 。

矩阵的 SVD 分解为： $A = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{125} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0.8 & 0.6 \\ 0.6 & -0.8 \end{bmatrix}$

做奇异值分解就是在矩阵的四个子空间中找到合适的基：

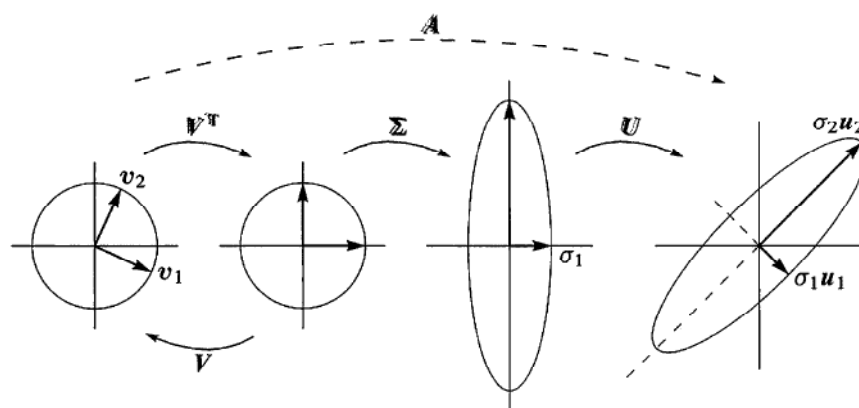
v_1, v_2, \dots, v_r 为行空间的标准正交基。

u_1, u_2, \dots, u_r 为列空间的标准正交基。

$v_{r+1}, v_{r+2}, \dots, v_n$ 为零空间的标准正交基。

$u_{r+1}, u_{r+2}, \dots, u_m$ 为左零空间的标准正交基。

奇异值分解在最小二乘法问题中有重要应用，因为在实际问题中常碰到矩阵 A 不是列满秩的状态，因此 $A^T A$ 不可逆，无法用之前的方法求最优解。即使是列满秩的情况当矩阵是超大型矩阵时， $A^T A$ 的计算量太大，用奇异值分解的办法会降低计算量。



图为 GS 给出的二阶方阵 SVD 的几何意义。

第 30 讲 线性变换及对应矩阵

Linear transformations and their matrices

线性变换 Linear transformations

本讲介绍线性变换。在传统的线代课程中，线性变换会出现在的比矩阵还要早的位置，实际上可以在脱离坐标和具体数值的情况下讨论线性变换。但是面对科学计算的课题，还是要回到坐标上来。每个线性变换都对应于一个矩阵，矩阵概念的背后就是线性变换的概念。

无论是“线性变换”还是“矩阵”，对于初学者来说都是陌生而且突兀的概念。矩阵看起来直观，但是其运算规则总让人摸不着头脑；线性变换的概念显得抽象一些，但其实也可以描述得很朴素。GS 是从矩阵的概念出发，往求解线性方程组的方向引导大家，他最成功的地方是引入了列图像以及矩阵乘法的行操作、列操作，将矩阵运算变得不那么“没头没脑”。而之前推荐过的《线性代数应该这样学》是从线性变换的角度出发，用朴素的描述在更高的抽象层面讨论，很多原理与法则就变得比较简洁。个人认为两者都优于从行列式出发进入线性代数的路线。在《理解矩阵》中，作者写到矩阵的背后就是线性变换，相似矩阵描述的实际上是同一种线性变换。GS 说理解线性变换的本质就是找到它背后的矩阵。这实际上是一个意思，只是出发点和看问题的角度不同。当具体的矩阵和抽象的线性变换在大脑中合二为一的时候，才是对线代最本质的理解。

例：投影 Projection

抛开矩阵，从线性变换的概念来讨论“投影”。通过线性变换使得平面内的一个向量变为平面内的另一个向量，这种变换关系通常称之为“映射”（mapping）。

$$T: \mathbf{R}^2 \rightarrow \mathbf{R}^2$$

$T(\mathbf{v})$ 就像一个函数，给一个输入，进行线性变换后得到一个输出。比如将二维平面的向量投影到一条直线上。

线性代数只讨论线性变换，而线性变换符合如下规则：对于任意向量 \mathbf{v} ， \mathbf{w} 和标量 c ，

$$T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w}) ; T(c\mathbf{v}) = c T(\mathbf{v})$$

将两者结合就得到： $T(c\mathbf{v} + d\mathbf{w}) = c T(\mathbf{v}) + d T(\mathbf{w})$ 。

反例：平面平移 Shift whole plane

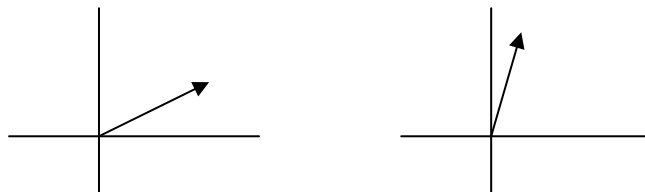
沿着某方向 \mathbf{v}_0 平移一个平面。这就不是一个线性变换。两条法则都违反。更简单的验证方法是线性运算规则的特例 $T(\mathbf{0}) = \mathbf{0}$ ，“平移”不符合这个规则特例。

反例：求长度 $T(\mathbf{v}) = \|\mathbf{v}\|$

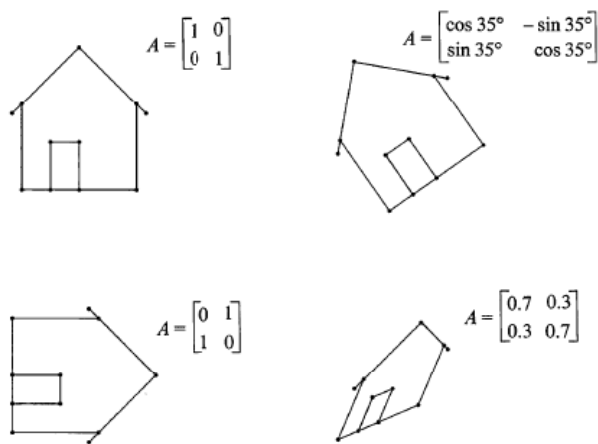
这个线性变换，输入一个三维向量，得到一个数值，或者说一维向量 $T: \mathbf{R}^3 \rightarrow \mathbf{R}^1$ 。这个变换可以满足 $T(\mathbf{0}) = \mathbf{0}$ ，但是数乘负数就不符合规则。

例：旋转 45 度 Rotation by 45°

这个变换 $T: \mathbf{R}^2 \rightarrow \mathbf{R}^2$ ，是一个线性变换，数乘和加法均符合。



对二维平面的图像做线性变换的操作示意图。



例：矩阵

$T(\mathbf{v}) = \mathbf{A}\mathbf{v}$ ，矩阵乘法是一个线性变换。 $\mathbf{A}(\mathbf{v} + \mathbf{w}) = \mathbf{A}(\mathbf{v}) + \mathbf{A}(\mathbf{w})$ ； $\mathbf{A}(c\mathbf{v}) = c\mathbf{A}(\mathbf{v})$ 。整个平面可以通过矩阵的乘法完成变换。

例如，对上图的房子图像施加矩阵 $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ ，则输出的是上下颠倒的房子。

理解线性变换的本质就是确定它背后的矩阵。

例：

对某一线性变换 $T: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ ，输入一个三维向量而输出是一个二维向量。变成矩阵的形式 $T(\mathbf{v}) = \mathbf{A}\mathbf{v}$ ，则矩阵 \mathbf{A} 是一个 2×3 矩阵。

描述线性变换 Describing $T(\mathbf{v})$

在平面内，如果我们已经了解两个线性无关的向量 \mathbf{v}_1 和 \mathbf{v}_2 经过线性变换的结果 $T(\mathbf{v}_1)$ 和 $T(\mathbf{v}_2)$ ，我们实际上可以通过其线性组合，了解平面内所有的向量线性变换的结果。因此如果我们想了解线性变换对整个输入空间的影响，只需要确定它的一组基 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 线性变换的结果。

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n$$

$$T(\mathbf{v}) = c_1 T(\mathbf{v}_1) + c_2 T(\mathbf{v}_2) + \dots + c_n T(\mathbf{v}_n)$$

线性变换与坐标无关，而矩阵是与坐标有关的。选定一组基，则对于一个向量而言 c_1, c_2 等等就是一组坐标值，给定了将向量表示为基向量线性组合的唯一的表达式。因此可以说坐标源自于一组基， c_1, c_2, \dots, c_n 就是向量的一组坐标值。通常给出空间的坐标是标准坐标，即一组标准基。例如，

$$\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 4 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

如果用矩阵 \mathbf{A} 来表示线性变换 $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ 。则需要两组基，即输入空间的一组基和输出空间的一组基，来分别确认输入向量与输出向量的坐标。设 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 为输入空间的基，而 $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$ 为输出空间的基，用矩阵来表示线性变换就是将向量的坐标乘以矩阵得到它在输出空间的坐标。

例：投影

将二维平面内的向量投影到一条直线，选择输入空间的基向量为 $\mathbf{v}_1, \mathbf{v}_2$ ，其中 \mathbf{v}_1 是沿着投影方向的向量， \mathbf{v}_2 是垂直于投影方向的向量，而输出空间的基选择为 $\mathbf{w}_1 = \mathbf{v}_1, \mathbf{w}_2 = \mathbf{v}_2$ 。对于输入空间中任意向量 \mathbf{v} 有 $\mathbf{v} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2$ ，输出为 $T(\mathbf{v}) = c_1 \mathbf{v}_1$ 。

因此这个线性变换的矩阵就是 $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ ，输入 $\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ 得到 $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ 0 \end{bmatrix}$ 。

这组例子中实际上我们以投影矩阵的特征向量为基，因此得到的矩阵 \mathbf{A} 是对角阵。对角线上就是投影矩阵的特征值 1 和 0。

而如果我们以标准坐标为基，即 $\mathbf{w}_1 = \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ， $\mathbf{w}_2 = \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 。对应的投影矩阵

就是之前学到的投影矩阵 $\mathbf{P} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}$ 对于投影到斜率 45 度的直线， $\mathbf{P} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$ 。

如何确定矩阵 \mathbf{A} Rule to find \mathbf{A}

矩阵 \mathbf{A} 的列实际上是描述输入原空间的基向量得到的列空间线性组合的系数：

$$T(\mathbf{v}_1) = a_{11}\mathbf{w}_1 + a_{21}\mathbf{w}_2 + \dots + a_{m1}\mathbf{w}_m,$$

$$T(\mathbf{v}_2) = a_{12}\mathbf{w}_1 + a_{22}\mathbf{w}_2 + \dots + a_{m2}\mathbf{w}_m,$$

.....

这样矩阵 \mathbf{A} 就满足 “ $\mathbf{A}[\text{输入向量的坐标}] = [\text{输出空间的坐标}]$ ”。

例如 $\mathbf{A} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$ ，所得结果就是输出空间的坐标，也就是输出空间的基进行

线性组合所需要的系数。

介绍一个特别的线性变换——求导， $T = \frac{d}{dx}$ 。

输入： $c_1 + c_2x + c_3x^2$ 基： $1, x, x^2$

输出： $c_2 + 2c_3x$ 。 基： $1, x$

这是一个 $T: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 的线性变换。

矩阵 A 满足 $A \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} c_2 \\ 2c_3 \end{bmatrix}$ ，可求得矩阵 $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ 。

更普遍的来讲，矩阵的逆矩阵就是线性变换的逆变换，矩阵的乘积就是线性变换的乘积，矩阵乘法源自于线性变换。

因为 GS 举了个 $T(v) = Av$ 的例子，所以有人可能晕了，不理解矩阵到底乘以坐标值还是向量本身。简单说明一下，始终记住“ A [输入向量的坐标]=[输出空间的坐标]”，是坐标值！坐标值！坐标值！在 $T(v) = Av$ 的例子中使用的是标准基，因此对 v 做线性变换 Av 所对应的矩阵就是 A 本身。

而如果定义另一个线性变换 $T(v)$ ，对 $v = c_1v_1 + c_2v_2 + \dots + c_nv_n$ 进行线性变换，

用矩阵 A 描述线性变换，则矩阵 A 要满足乘以坐标值 $\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$ 得到输出向量的坐标值

$\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{bmatrix}$ ，用这个坐标值对输出空间的基 w_1, w_2, \dots, w_m 进行线性组合，从而得到输出

向量 $w = d_1w_1 + d_2w_2 + \dots + d_mw_m$ 。

如果我们对输入空间的一个基向量 v_1 做线性变换 $T(v_1)$ ，则矩阵 A 乘以的就是

坐标值 $\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ 。得到的输出坐标是 $A \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$ ，当我们知道线

性变换结果 w 的时候，就可以用它的坐标值确认矩阵 A 第一列的分量。

第 31 讲 基变换和图像压缩

Change of basis; image compression

本讲介绍基变换。选择合适的基向量会给计算制造便利。基变换的一个重要应用就是压缩，图像、影像、音频和其它一些数据都会因为基变换而得到更高效的压缩储存。本讲的主题仍旧是线性变换和矩阵的关联。

图像压缩 Compression of images

本讲涉及的压缩过程是有损压缩。例如一幅像素是 512×512 的静态黑白图像，图像用一个向量来表示，向量的分量 x_i 表示像素的灰度，变化范围 $0 \leq x_i \leq 255$ ，占 8bits。该向量属于 \mathbf{R}^n 空间， $n=(512)^2$ 。彩色图像描述每个点的像素需要三个数据，向量长度是黑白图像的 3 倍。

图像的标准压缩方式为 JPEG（联合图像专家组 Joint Photographic Experts Group）。图像压缩的本质就是基变换。

压缩前图像采用的基向量是标准基。但是在图像中离得很近的区域，颜色是非常接近的，比如教学视频中黑板的一个区域，这些区域像素的灰度值很接近，但是用标准基来存储并没有利用上这一特点，这就给了我们压缩的空间。

标准基就是 $\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$ ，而显然对于灰度很接近的情况， $\begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ 是一个很好的

基（代表低频信号，频率为 0，平滑。在图像压缩后经常会存在，因为其系数通常很大。括号里的话是 GS 后来又提到的），在一组基中有一个这样的向量能解决很大的问题，可以处理像素灰度接近一致的情况。但图像不是灰度完全一致的，因此接

下来的问题是跟它相配合的基要选择哪些。极端的情况包括选择 $\begin{bmatrix} +1 \\ -1 \\ +1 \\ -1 \\ \vdots \end{bmatrix}$ ，它可以给出

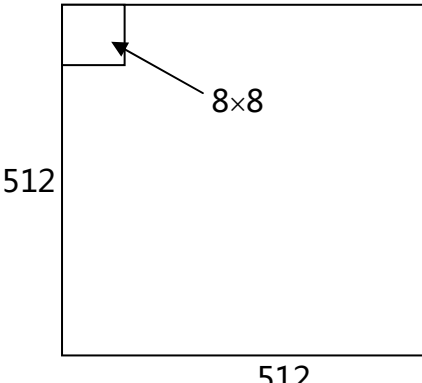
类似国际象棋盘那种黑白相间的状态（最高频信号，噪音、扰动.....在图像压缩后

很少存在），还有图像一半图像暗一半图像亮，可以选择 $\begin{bmatrix} \vdots \\ +1 \\ -1 \\ \vdots \\ -1 \end{bmatrix}$ 。

傅里叶基 Fourier basis

在 JPEG 中,将 512×512 的区域划分为 8×8 的区块进行处理,所用到的基是傅里叶基向量。

以 \mathbf{R}^8 中的基举例

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ \omega \\ \omega^2 \\ \omega^3 \\ \omega^4 \\ \omega^5 \\ \omega^6 \\ \omega^7 \end{bmatrix}, \begin{bmatrix} 1 \\ \omega^2 \\ \omega^4 \\ \omega^6 \\ \omega^8 \\ \omega^{10} \\ \omega^{12} \\ \omega^{14} \end{bmatrix}, \dots, \begin{bmatrix} 1 \\ \omega^7 \\ \omega^{14} \\ \omega^{21} \\ \omega^{28} \\ \omega^{35} \\ \omega^{42} \\ \omega^{49} \end{bmatrix}$$


傅里叶基 (Fourier basis) 就是之前讲过的傅里叶矩阵的列向量,每个元素为复数的幂。在 8×8 区块中有 64 个系数,64 个基向量,在这个 64 维空间中做基变换。

$$\text{信号 } \mathbf{x} \xrightarrow{\text{无损}} \text{系数 } \mathbf{c} \xrightarrow{\text{有损}} \text{系数 } \hat{\mathbf{c}} \text{ (很多的 0)} \longrightarrow \hat{\mathbf{x}} = \sum \hat{c}_i \mathbf{v}_i$$

首先对输入的信号 \mathbf{x} ,从标准基变换为傅里叶基,得到系数 \mathbf{c} ,这一步是无损的过程。之后设置阈值进行压缩,超过阈值的认定为肉眼无法分辨的信号。在数学上表现为某些基向量的系数很小,这部分可以丢弃,随之得到新的系数 $\hat{\mathbf{c}}$ 。将新的系数赋值在傅里叶基上求和得到 $\hat{\mathbf{x}} = \sum \hat{c}_i \mathbf{v}_i$ 。此时求和项已经不是 64 项,可能只剩下两三项,这就是压缩。

视频文件可以视为图像的序列,一幅一幅进行图像压缩即可。但这样做没有利用好视频的性质,因为视频是连续的,一幅图像和下一幅图像非常接近,因此可以存储一幅基础图像,随后只存储下一幅图像对它的修正部分。

小波 Wavelets

下面介绍另一组和傅里叶竞争的基向量——小波。

以 \mathbf{R}^8 中的基举例

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}.$$

这个只是一个小波选择,还有很多种更多精细的选择,这一组基中有太多从 +1

跳转到-1 的变化。线性代数要做的基变换，就是将标准基下的向量 $\mathbf{p}=[p_1, p_2, p_3, \dots, p_8]^T$ 表示为小波基的线性组合，求出线性组合的参数 \mathbf{c} 满足 $\mathbf{p}=\mathbf{c}_1\mathbf{w}_1+\mathbf{c}_2\mathbf{w}_2+\dots+\mathbf{c}_8\mathbf{w}_8$ ，即 $\mathbf{p}=\mathbf{W}\mathbf{c}$ 。 \mathbf{W} 即以小波基向量为列向量的小波矩阵。因此有 $\mathbf{c}=\mathbf{W}^{-1}\mathbf{p}$ 。

好的基向量组要求：第一，可以快速求逆矩阵，例如快速傅里叶变换，这里也存在快速小波变换，因为小波矩阵列向量正交，因此可以转置得到逆矩阵；第二，要少量基向量就可以近似信号，可压缩的比例就比较高。

基变换 Change of basis

\mathbf{W} 的列向量是一组新的基向量。在旧基向量体系下的向量 \mathbf{x} 和新基向量下的 \mathbf{c} 的关系是： $\mathbf{x}=\mathbf{W}\mathbf{c}$ 。

变换矩阵

已知一个线性变换 $T: \mathbf{R}^8 \rightarrow \mathbf{R}^8$ 。当使用空间的一组基 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_8$ 时，线性变换对应的矩阵为 \mathbf{A} ；当使用一组新的基 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_8$ 时，线性变换对应的矩阵为矩阵 \mathbf{B} 。两个矩阵对应的是同一个线性变换，只是使用了不同的基向量。那么 \mathbf{A} 和 \mathbf{B} 为相似矩阵 $\mathbf{B}=\mathbf{M}^{-1}\mathbf{A}\mathbf{M}$ ， \mathbf{M} 是基变换矩阵。

复习一下线性变换的内容：

$$\begin{aligned}\mathbf{x} &= \mathbf{c}_1\mathbf{v}_1 + \mathbf{c}_2\mathbf{v}_2 + \dots + \mathbf{c}_8\mathbf{v}_8 \\ T(\mathbf{x}) &= \mathbf{c}_1 T(\mathbf{v}_1) + \mathbf{c}_2 T(\mathbf{v}_2) + \dots + \mathbf{c}_8 T(\mathbf{v}_8) \\ T(\mathbf{v}_1) &= a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + \dots + a_{81}\mathbf{v}_8 \\ T(\mathbf{v}_2) &= a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + \dots + a_{82}\mathbf{v}_8 \\ &\dots\end{aligned}$$

矩阵 \mathbf{A} 就是 a_{ij} 组成的矩阵。

如果我们使用的基向量就是特征向量 $T(\mathbf{v}_i)=\lambda_i\mathbf{v}_i$ ，矩阵 \mathbf{A} 就变成对角阵 $\mathbf{\Lambda}$ 。找出特征向量是压缩最理想的结果，但是找出图像的特征向量代价太大，因此我们找到代价小但是接近理想状态的基向量（例如小波基）进行基变换，完成压缩过程。

再强调一次，与矩阵 \mathbf{A} 做运算的是坐标值，运算结果也是坐标值。此处线性变换中输入和输出空间都是同一个空间，矩阵 \mathbf{A} 对应的是线性变换前后都使用一组基 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_8$ 的情形，而矩阵 \mathbf{B} 对应的是线性变换前后都用基 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_8$ 的情形。跟上一讲变换前用 \mathbf{v} 变换后用 \mathbf{w} 是不同的。

第 32 讲 复习三

Exam 3 review

主要内容

- 特征值和特征向量 $A\mathbf{x}=\lambda\mathbf{x}$
- 微分方程 $d\mathbf{u}/dt=A\mathbf{u}$ 和矩阵指数形式 e^{At}
- 对称矩阵 $A=A^T$ 的特征值永远为实数，总有足够的特征向量实现矩阵的对角化 $A=Q\Lambda Q^T$ 。
- 正定矩阵
- 相似矩阵 $B=M^{-1}AM$ ，两矩阵具有相同的特征值，但是特征向量不同，矩阵的乘方形式很接近 $B^k=M^{-1}A^kM$ 。
- 奇异值分解 SVD， $A=U\Sigma V^T$ 。

例题：

1) 微分方程：

$$\frac{d\mathbf{u}}{dx} = A\mathbf{u} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{u}$$

a) 求通解为 $\mathbf{u}(t) = c_1 e^{\lambda_1 t} x_1 + c_2 e^{\lambda_2 t} x_2 + c_3 e^{\lambda_3 t} x_3$

答：首先求处矩阵的特征值，矩阵为奇异矩阵，因此具有一个特征值 0。矩阵为反对称矩阵，因此特征值为纯虚数。

$$|A - \lambda I| = \begin{vmatrix} -\lambda & -1 & 0 \\ 1 & -\lambda & -1 \\ 0 & 1 & -\lambda \end{vmatrix} = -\lambda^3 - 2\lambda = 0$$

解得 $\lambda_1=0$ ， $\lambda_2=\sqrt{2}i$ ， $\lambda_3=-\sqrt{2}i$ 。

则通解形式为 $\mathbf{u}(t) = c_1 x_1 + c_2 e^{\sqrt{2}it} x_2 + c_3 e^{-\sqrt{2}it} x_3$ ，解既不发散也不收敛，纯虚数特征值使得解的分量始终在单位圆上循环。

b) 解为周期性函数，何时返回初值？（周期是多少）

解： $\sqrt{2}Ti = 2\pi i$ ，所以有 $T = \sqrt{2}\pi$ 。

c) 求出两个特征向量，证明其正交？

满足 $A^T A = A A^T$ 的矩阵具有正交的特征向量。对称和反对称矩阵都具有正交的特征向量。正交矩阵也符合此要求。

答：将特征值代入可以求得 $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} -1 \\ \sqrt{2}i \\ 1 \end{bmatrix}$, $\mathbf{x}_3 = \begin{bmatrix} 1 \\ \sqrt{2}i \\ -1 \end{bmatrix}$ 。可以用内积验证其正交性，要注意对复向量取内积要做共轭。

d) 微分方程的解写成矩阵指数形式为 $\mathbf{u}(t) = e^{At}\mathbf{u}(0)$ ，如何计算 e^{At} ？

答：若有 $\mathbf{A} = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1}$ 。则有 $e^{At} = \mathbf{S}e^{\mathbf{\Lambda}t}\mathbf{S}^{-1}$ 。

$$\text{其中 } e^{\mathbf{\Lambda}t} = \begin{bmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & e^{\lambda_n t} \end{bmatrix}。$$

2) 一个 3×3 矩阵 \mathbf{A} 具有特征值 $\lambda_1 = 0$ $\lambda_2 = c$ $\lambda_3 = 2$ 。对应的特征向量为 $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ ，

$$\mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}。$$

a) 求 c 使得矩阵可对角化。

答：矩阵具有足够多的线性无关的特征向量时，可以完成对角化。此处的特征向量不仅线性无关而且正交，因此任意 c 都可以满足矩阵可对角化。

b) 求 c 使得矩阵对称？

答：如果矩阵对称 $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ ，则其特征值为实数，因此 c 为实数时，此矩阵为对称矩阵。

c) 求 c 使得矩阵正定？

答：有一个特征值为 0，矩阵不可能正定。 c 为实数且 $c \geq 0$ 时，矩阵半正定。

d) 此矩阵可能是 Markov 矩阵么？

答：Markov 矩阵具有一个特征值 1，其它特征值小于 1。此矩阵具有特征值 2，因此不可能是 Markov 矩阵。

e) 矩阵 $\mathbf{P} = (1/2)\mathbf{A}$ 可能是投影矩阵么？

答：投影矩阵的特征值为 1 或者 0，所以当 $c = 0$ 或者 2 时，矩阵 $\mathbf{P} = (1/2)\mathbf{A}$ 是投影矩阵。

3) 对于矩阵 \mathbf{A} 可做奇异值分解 $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ ，已知 $\mathbf{\Sigma} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ ，并且矩阵 \mathbf{U} 和 \mathbf{V}

均为 2 列。

a) 对于矩阵 \mathbf{A} 我们知道些什么？

答：矩阵 \mathbf{A} 为二阶可逆矩阵。

b) 如果 $\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & -5 \end{bmatrix}$ 呢？

答：不可能，因为奇异值分解要求奇异值大于 0。

c) 如果 $\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}$ 呢？

答：矩阵为奇异阵，秩为 1，零空间的维数为 1。四个子空间由 \mathbf{U} 和 \mathbf{V} 的列向量张成。例如 \mathbf{v}_2 就在矩阵的零空间中。

4) 矩阵 \mathbf{A} 对称并且正交。

a) 我们对它的特征值知道什么？

答：对称矩阵的特征值为实数，正交矩阵的特征值的绝对值为 1。因此矩阵的特征值为 +1 或者 -1。

b) 矩阵 \mathbf{A} 是否为正定矩阵？

答：特征值可以是 -1，非正定。

c) 矩阵 \mathbf{A} 没有重特征值？

答：只能是 +1 或者 -1，因此 3 阶以上肯定有重特征值。

d) 矩阵 \mathbf{A} 可对角化么？

答：所有的对称矩阵和所有的正交矩阵都可对角化。

e) 矩阵 \mathbf{A} 可逆么？

答：正交矩阵都可逆。

f) 证明 $\mathbf{P} = (1/2)(\mathbf{A} + \mathbf{I})$ 是投影矩阵。

答： $\mathbf{P}^2 = (\frac{1}{2}(\mathbf{A} + \mathbf{I}))^2 = \frac{1}{4}(\mathbf{A}^2 + 2\mathbf{A} + \mathbf{I})$ 。其中 $\mathbf{A}^2 = \mathbf{A}^T \mathbf{A} = \mathbf{I}$ ，所以 $\mathbf{P}^2 = \frac{1}{2}(\mathbf{A} + \mathbf{I}) = \mathbf{P}$ 。因此它可以是投影矩阵。此外 $(1/2)(\mathbf{A} + \mathbf{I})$ 的特征值为 1 或者 0，并且是对称矩阵，这亦可证明它是投影矩阵。

第 33 讲 左右逆和伪逆

Left and right inverses; pseudoinverse

本节主要介绍左右逆矩阵和伪逆矩阵。

两侧逆矩阵 2-sided inverse

矩阵 A 的两侧逆矩阵 A^{-1} 满足 $AA^{-1}=I=A^{-1}A$ 。这就是我们通常说矩阵 A 的逆矩阵。此时 $r=m=n$ ， A 为满秩方阵。

左逆矩阵 Left inverse

矩阵 A 的列满秩 $r=n$ ，列向量线性无关。则矩阵的零空间 $N(A)$ 只有零向量，方程 $Ax=b$ 无解或者有唯一解。当矩阵 A 列满秩的时候，矩阵 $A^T A$ 为可逆矩阵，这是讨论最小二乘问题的核心。

矩阵 $A^T A$ 是 $n \times n$ 的对称矩阵且满秩，因此 $A^T A$ 是可逆矩阵，即 $(A^T A)^{-1} A^T A = I$ 。我们称 $A_{\text{left}}^{-1} = (A^T A)^{-1} A^T$ 为 A 的左逆矩阵。 A_{left}^{-1} 为 $n \times m$ 矩阵，而 A 为 $m \times n$ 矩阵， $A_{\text{left}}^{-1} A = I$ 是 $n \times n$ 矩阵。

右逆矩阵 Right inverse

矩阵 A 的行满秩 $r=m$ ，行向量线性无关。则矩阵的左零空间 $N(A^T)$ 只有零向量。 A 的零空间维数为 $n-m$ ，因此有 $n-m$ 个自由变量， n 大于 m 时，方程 $Ax=b$ 有无穷多解。

矩阵 A 的右逆矩阵为 $A_{\text{right}}^{-1} = A^T (AA^T)^{-1}$ 。

通常情况下右乘左逆矩阵得不到单位阵， $AA_{\text{left}}^{-1} = A(A^T A)^{-1} A^T = P$ ，这是列空间的投影矩阵。仅在 $m=n$ 的条件下， $AA_{\text{left}}^{-1} = I$ 。一个长方形矩阵 A 不可能有两侧逆，因为 A 和 A^T 总有一个零空间的维数不是 0。

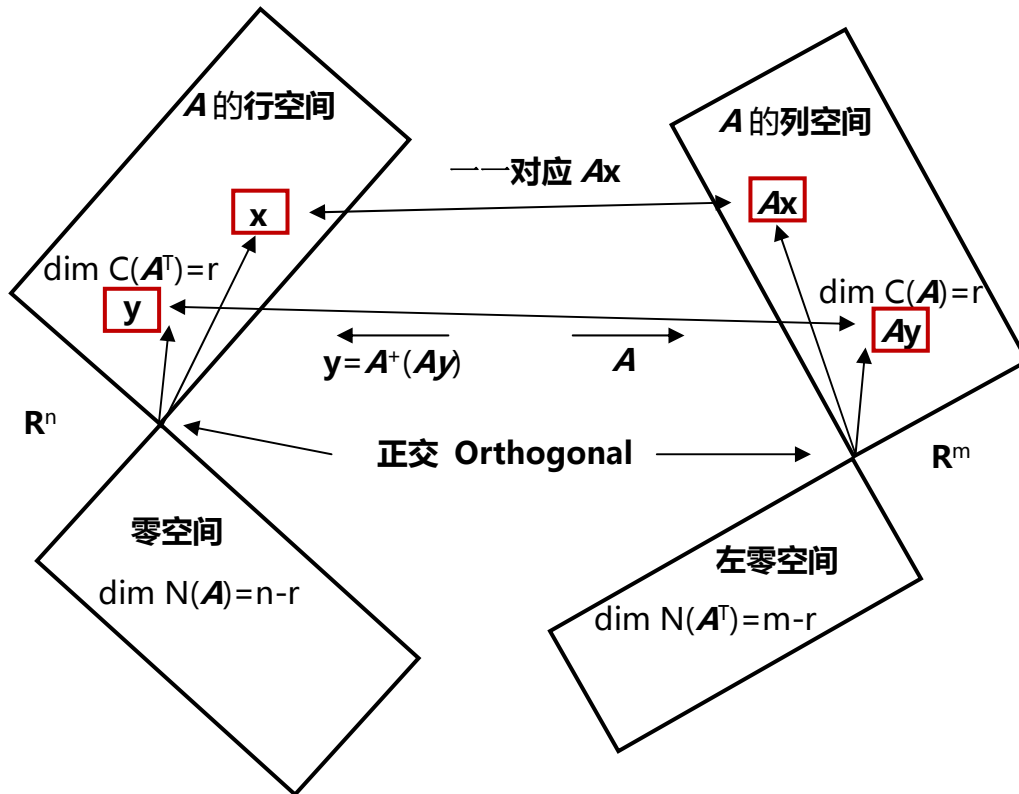
同样的，左乘右逆矩阵得到的是 $A_{\text{right}}^{-1} A = A^T (AA^T)^{-1} A$ ，这是向行空间投影的矩阵，投影矩阵在投影空间内表现的就像单位阵 I 一样。

伪逆矩阵 Pseudoinverse

可逆矩阵的零空间和左零空间只有零向量。列满秩的矩阵的零空间只有零向量，行满秩的矩阵的左零空间只有零向量。但对于不满秩的矩阵 ($r < n, r < m$)，其两个零空间均存在，这就使得它取不到逆矩阵。因为逆矩阵的作用可以看作是原矩阵的逆操作，但是矩阵 A 对其零空间中向量操作后变为 0 ，这时没有逆操作能够恢复这一过程，所以带有零空间就不能取逆矩阵。

观察不满秩矩阵 A 的四个子空间，其行空间和列空间的维数相等，均为 r 。在其行空间中的向量 x 经过矩阵 A 操作后，变为列空间中的向量 Ax 。而 x 与 Ax 为

一一对应的关系。如果将矩阵 A 限制在行空间和列空间上，它是个可逆矩阵，此时 A 的逆矩阵就是所谓的“伪逆矩阵 A^+ ”。



问题的关键就是一一对应，即对于行空间中的向量 $x \neq y$ ，其通过矩阵 A 映射到列空间得到的向量 $Ax \neq Ay$ 。

证明：若行空间中存在向量 y 与向量 x 在列空间中对应的向量相同，则有 $Ax - Ay = 0$ ，即 $A(x - y) = 0$ ，则 $x - y$ 为矩阵 A 零空间中的向量，但矩阵的行空间对线性运算封闭，因此 $x - y$ 为行空间中的向量，因两个子空间正交，所以有 $(x - y) = 0$ ，即 x 与 y 相同。因此行空间与列空间中的向量为一一对应。

统计学家非常需要伪逆矩阵，因为他们要做很多最小二乘法进行线性回归的问题。如果矩阵不满秩，则 $A^T A$ 不可逆，所以无法用之前的办法解决，此时要用到伪逆矩阵。

求伪逆矩阵 A^+ 的一个方法是利用奇异值分解 $A = U \Sigma V^T$ ，其中对角阵 Σ 两是由矩阵奇异值排列在对角线上构成的 $m \times n$ 矩阵，其秩为 r 。则其伪逆矩阵 Σ^+ 为 $n \times m$ 矩阵，矩阵的秩也为 r 。

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix}, \Sigma^+ = \begin{bmatrix} 1/\sigma_1 & & & \\ & \ddots & & \\ & & 1/\sigma_r & \\ & & & \end{bmatrix}$$

$$\Sigma \Sigma^+ = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}, \Sigma^+ \Sigma = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}$$

矩阵 Σ 右乘伪逆矩阵得到 $m \times m$ 矩阵，矩阵对角线前 r 个元素为 1，其它元素均为 0，而左乘伪逆矩阵得到 $n \times n$ ，矩阵对角线前 r 个元素为 1，其它元素均为 0，这两个都是 Σ 的投影矩阵，一个投影到行空间，一个投影到列空间。

而矩阵 A 的伪逆矩阵为 $A^+ = V \Sigma^+ U^T$ 。

逆矩阵满足四个性质：

$$AA^+A = A$$

$$A^+AA^+ = A^+$$

$$AA^+ = (AA^+)^T$$

$$A^+A = (A^+A)^T$$

注意： $AA^+ = U \Sigma V^T V \Sigma^+ U^T = U \Sigma \Sigma^+ U^T$ 得到的并不是形如 $\Sigma \Sigma^+$ 这种对角线上只有 1 和 0 的对角阵，所得结果是 A 行空间的投影矩阵。

例如 $A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ，通过奇异值分解计算可以得到 $A^+ = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}$ 。而 $AA^+ = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$ 。

第 34 讲 总复习

Course Review

本讲为线性代数课程总复习，复习的方法就是做往年试题。

1) 已知 $Ax = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ 无解， $Ax = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ 仅有一解。

a) 这个矩阵为 $m \times n$ 矩阵，秩为 r ，那么 m 是多少？ r ？

答： $m=3$ 。 $Ax = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ 无解说明 r 小于 m ， $Ax = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ 仅有一解说明矩阵零空间只

有零向量，所以 $r=n$ 。例如 $A = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ 或者 $A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$ 。

b) 判断 $\det(A^T A) = \det(AA^T)$ 是否成立？

答：不成立。可以看下面两题的答案，分别是可逆和不可逆的矩阵。需要注意的是：只有当矩阵是方阵时，才有行列式的乘积等于乘积的行列式。

c) $A^T A$ 是否可逆？

答：可逆，矩阵 A 列满秩 $r=n$ ，因此 $A^T A$ 可逆。

d) AA^T 是否正定？

答：否，不满秩，不可能是正定矩阵。

e) $A^T y = c$ ，证明对于任意右侧向量至少有一个解。

答：矩阵 A^T 为 $n \times 3$ 矩阵，行满秩，矩阵 A^T 的零空间维数为 $3-n$ ，方程有无穷多解。

2) 已知矩阵 $A = [v_1 \ v_2 \ v_3]$

a) 求解 $Ax = v_1 - v_2 + v_3$

答： $x = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$

b) 若 $v_1 - v_2 + v_3 = 0$ ，则上一题的解是否不唯一？

答：列向量线性相关，零空间有非零向量， $Ax = 0$ 解不唯一。

c) 若 v_1, v_2 和 v_3 为标准正交，求 v_1, v_2 线性组合距离 v_3 最近？

答：原点。

3) 已知 Markov 矩阵 $A = \begin{bmatrix} 0.2 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.3 \\ 0.4 & 0.4 & 0.4 \end{bmatrix}$ 。

a) 特征值？

答：列 1+列 2=2×列 3，所以矩阵为奇异阵，有一个特征值 $\lambda_1=0$ 。Markov 矩阵还有一个特征值 $\lambda_2=1$ ，从迹可知最后一个特征值为 $\lambda_3=-0.2$ 。

b) $u_k = A^k \begin{bmatrix} 0 \\ 10 \\ 0 \end{bmatrix}$ ，求 k 步之后的状态？当 k 趋近于无穷时的状态？

答： $u_k = c_1 \lambda_1^k x_1 + c_2 \lambda_2^k x_2 + c_3 \lambda_3^k x_3$ 。代入特征值可知，当 k 趋近于无穷时，只剩下 $c_2 x_2$ 。求解矩阵的特征向量 x_2 ，得到 $x_2 = \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix}$ 。 u_0 引入 10 个人 (Markov 矩阵可以当作人口流动模型来看)，并且最终的分配比例为 3:3:4，所以不用逐个求解特征向量和 c，可得 $u_k = \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix}$ 。

4) 求符合题目要求的 2×2 矩阵。

a) 投影到 $a = \begin{bmatrix} 4 \\ -3 \end{bmatrix}$ 所在直线的投影矩阵。

答： $P = \frac{aa^T}{a^T a} = \dots$

b) 矩阵具有特征值 0 和 3，对应的特征向量分别为 $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 和 $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ 。

答： $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}^{-1} = \dots$

c) 矩阵 A 不能分解为 $B^T B$ 。

答：给出一个不对称矩阵就可以了。

d) 矩阵有正交的特征向量，但是非对称。

答：反对称矩阵或者在复数域上有正交特征向量的其它矩阵。

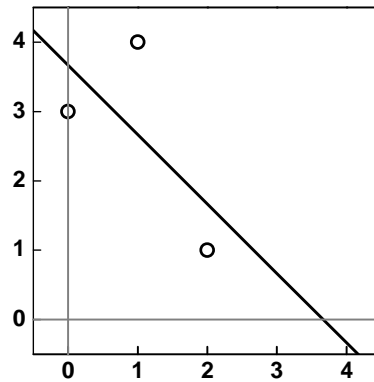
5) 已知 $\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}$ ， $\begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 11/3 \\ -1 \end{bmatrix}$ 。

a) 求投影 p 。

答：把 $\begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 11/3 \\ -1 \end{bmatrix}$ 代入得到 $\mathbf{p} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 11/3 \\ -1 \end{bmatrix} = \begin{bmatrix} 11/3 \\ 8/3 \\ 5/3 \end{bmatrix}$

b) 画图，三个点和拟合曲线。

答：



c) 找出另一个非零向量 \mathbf{b} ，使得 $\begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ 。

答 求解的 \mathbf{b} 要与系数矩阵的列空间正交，才会使得最优解为 $\mathbf{0}$ 。解得 $\mathbf{b} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ 。